Werk

Titel: Metrika
Verlag: Physica-Verlag
Jahr: 1993
Kollektion: Mathematica
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN358794056_0040
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN358794056_0040

Übergeordnetes Werk

Werk Id: PPN358794056
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN358794056

Contact

Niedersächsische Staats- und Universitätsbibliothek Göttingen
Georg-August-Universität Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen
Germany
Email: gdz@sub.uni-goettingen.de

Metrika (1993) 40:1–23

Metrika

Construction of Minimax-Tests for Bounded Families of Probability-Densities

ROBERT HAFNER

Institut für Angewandte Statistik, Johannes-Kepler- Universität Linz, 4045 Linz/Auhof, Austria

Summary: A new method for the construction of least favourable pairs of densities and of minimax-tests is given for the compound test-problem $H_0: \underline{g} \leq f \leq \bar{g}$ against $H_1: \underline{f} \leq f \leq \bar{f}$, the bounds $\underline{g}, \bar{g}, \underline{f}, \bar{f}$ being fixed. The main tool of the method is the risk-function for simple test-problems.

Key words: Minimax-tests, robust tests, least favourable pairs of distributions, risk-function.

1 Introduction

Suppose $(X, \mathscr{A})$ is the sample-space and the σ-algebra of measurable events of some experiment, μ denotes a σ-finite measure on $(X, \mathscr{A})$ and $\mathscr{M}$ the set of μ-densities of μ-dominated probability-measures on $(X, \mathscr{A})$. By $f, g \ldots$ we denote the μ-densities of probability-measures on $(X, \mathscr{A})$. Now two subsets of $\mathscr{M}$ are defined:

$$\mathscr{M}_0 := \{f: \underline{g} \leq f \leq \bar{g} \ldots \mu - \text{a.e.}\} ,$$

$$\mathscr{M}_1 := \{f: \underline{f} \leq f \leq \bar{f} \ldots \mu - \text{a.e.}\} ,$$

where the nonnegative, measurable bounds $\underline{g}, \bar{g}, \underline{f}, \bar{f}$ are subject to the conditions:

$$\int_X \underline{g}\, d\mu \leq 1 \leq \int_X \bar{g}\, d\mu , \quad \int_X \underline{f}\, d\mu \leq 1 \leq \int_X \bar{f}\, d\mu .$$

We consider the compound test-problem $(\mathscr{M}_0, \mathscr{M}_1)$:

0026–1335/93/1/1–23 $2.50

$$H_0: f \in \mathcal{M}_0 \quad H_1: f \in \mathcal{M}_1 \ .$$

In [1] Kuznetsov describes a method to construct minimax-tests for it. In the present paper a different method for the construction of these tests is developed. The main tool used is the risk-function for simple test-problems $(f_0, f_1): H_0: f = f_0$ versus $H_1: f = f_1$. This tool has already proved its power for the construction of least favourable pairs of distributions and minimax-tests for compound test-problems $(\mathcal{M}_0, \mathcal{M}_1)$, where the sets $\mathcal{M}_0$, $\mathcal{M}_1$ are either defined by:

$$\mathcal{M}_i := \{P \in \mathcal{M}: P(A) \leq (1-\varepsilon_i) \cdot P_i(A) + \delta_i, \ldots \forall A \in \mathcal{A}\} \quad i = 0, 1 \ ,$$

so-called contamination-neighbourhoods of the central probability-measures P_0, P_1, see [2], [3], or by:

$$\mathcal{M}_i := \{P \in \mathcal{M}: P(A) \leq P_i(A^{\varepsilon_i}) + \delta_i, \ A \in \mathcal{A}\} \quad i = 0, 1 \ ,$$

with $A^{\varepsilon} := \{x: d(x, A) < \varepsilon\}$ – so-called Prokhorov-neighbourhoods of the measures P_0, P_1, see [4], [5].

As is well known these compound problems are fundamental in the theory of robust testing. The solutions, the existence of which has been known since the paper of Huber and Strassen [6] are constructive, simple and apt for practical application.

By the risk-function-method we also obtain a remarkably simple method for the construction of both, a pair of least favourable distributions as well as the family of minimax-tests for the problem treated by Kuznetsov.

2 The Risk-Function

We mention some well known facts about Neyman-Pearson-tests (NP-tests) for simple test-problems $H_0: f = f_0 \ H_1: f = f_1$. The family of these NP-tests is given by:

$$\varphi_{c,\gamma}(x | f_0, f_1) = \begin{cases} 1 \\ \gamma \ldots f_1(x)/f_0(x) \gtreqless c \ldots \\ 0 \end{cases} \quad \text{for} \quad 0 \leq c \leq \infty \ , \quad 0 \leq \gamma \leq 1 \ .$$

the quantities $\alpha(c, \gamma) = E_{f_0}(\varphi_{c,\gamma})$ and $\beta(c, \gamma) = E_{f_1}(1 - \varphi_{c,\gamma})$ are respectively the risk of the first and the risk of the second kind under the test $\varphi_{c,\gamma}$. The pair $(\alpha, \beta) = (\alpha(c, \gamma), \beta(c, \gamma))$ is called the risk of the test $\varphi_{c,\gamma}$.

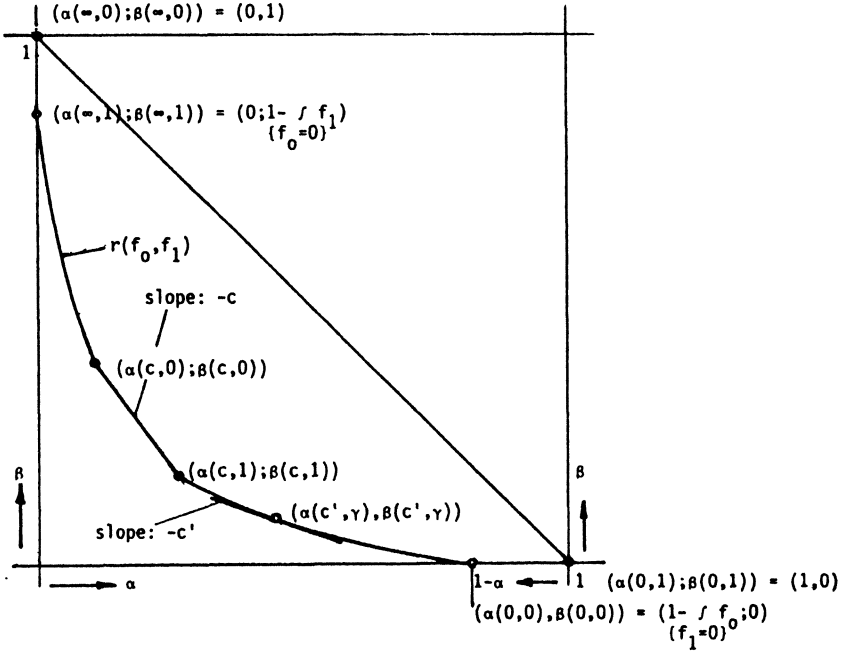

Fig. 1

In the (α,β)-plane the set $\{(\alpha(c,\gamma),\beta(c,\gamma)): 0\leq c\leq\infty,\ 0\leq\gamma\leq 1\}$ is a convex curve joining the points $(\alpha(\infty,0),\beta(\infty,0))=(0,1)$ and $(\alpha(0,1),\beta(0,1))=(1,0)$ (see Fig. 1). It is called the risk-function of the test problem and shall be denoted by $r(f_0,f_1)$. $(\alpha(c,\gamma),\beta(c,\gamma))$ is a parameter-represention of $r(f_0,f_1)$. Since

$$\begin{aligned}1-\alpha(c,\gamma)&=P_{f_0}(f_1/f_0<c)+(1-\gamma)\cdot P_{f_0}(f_1/f_0=c)\\ \beta(c,\gamma)&=P_{f_1}(f_1/f_0<c)+(1-\gamma)\cdot P_{f_1}(f_1/f_0=c)\end{aligned}\tag{1}$$

$(1-\alpha(c,\gamma),\beta(c,\gamma))$ and thus $r(f_0,f_1)$ can be interpreted as the map of the distribution functions of the density-quotient $q(x)=f_1(x)/f_0(x)$ under f_0 and f_1 plotted one against the other.

For c-values with $P_{f_i}(f_1/f_0=c)>0$ the risk-function $r(f_0,f_1)$ contains line-segments of slope $-c$ joining the points $(\alpha(c,0),\beta(c,0))$ and $(\alpha(c,1),\beta(c,1))$. In the general point $(\alpha(c,\gamma),\ \beta(c,\gamma))$ the straight line with slope $-c$ is a tangent of $r(f_0,f_1)$. This is also true if $r(f_0,f_1)$ has an edge in $(\alpha(c,\gamma),\beta(c,\gamma))$. Finally the optimality of NP-tests implies:

$$\beta(c,\gamma) = \inf\{E_{f_1}(1-\varphi) : \varphi \in \Phi,\ E_{f_0}(\varphi) = \alpha(c,\gamma)\}\ ,$$
$$\alpha(c,\gamma) = \inf\{E_{f_0}(\varphi) : \varphi \in \Phi,\ E_{f_1}(1-\varphi) = \beta(c,\gamma)\}\ , \tag{2}$$

i.e. $r(f_0, f_1)$ is the lower border of the convex set

$$R = \{(\alpha,\beta) : \alpha = E_{f_0}(\varphi), \beta = E_{f_1}(1-\varphi) \ldots \forall\varphi : 0 \le \varphi \le 1\}\ ,$$

the so-called risk-set of the test-problem (f_0, f_1).

For the purpose of the present paper it is useful to extend the notion of risk-function to the case, where f_0 and/or f_1 are the densities of general positiv μ-dominated measured P_0 and P_1. $(\alpha(c,\gamma), \beta(c,\gamma))$ and $r(f_0, f_1)$ shall be defined in the same way as for probability-measures. The graph of $r(f_0, f_1)$ is also monotonically decreasing, convex and joins the points $(\alpha,\beta) = (0, \int_X f_1 d\mu)$ and $(\alpha,\beta) = (\int_X f_0 d\mu, 0)$. As regards tangents, all statements made earlier remain valid.

3 Results

We consider the generalized simple test-problems $(\underline{g},\underline{f})$, $(\underline{g},\bar{f})$, $(\bar{g},\underline{f})$, $(\bar{g},\bar{f})$. The corresponding risk-functions $r(\underline{g},\underline{f})$, $r(\underline{g},\bar{f})$, $r(\bar{g},\underline{f})$, $r(\bar{g},\bar{f})$ respectively connect the points: $(0, \int_X \underline{f})$, $(\int_X \underline{g}, 0)$; $(0, \int_X \bar{f})$, $(\int_X \underline{g}, 0)$; $(0, \int_X \underline{f})$, $(\int_X \bar{g}, 0)$; $(0, \int_X \bar{f})$, $(\int_X \bar{g}, 0)$ in the (α,β)-plane (compare Fig. 2).

We shift these risk-functions in the way shown in Fig. 3. The shifted curves are denoted by $r^*(\underline{g},\underline{f})$, $r^*(\underline{g},\bar{f})$, $r^*(\bar{g},\underline{f})$, and $r^*(\bar{g},\bar{f})$. $r^*(\bar{g},\bar{f})$ is identical with $r(\bar{g},\bar{f})$. The curves $r^*(\cdot,\cdot)$ now connect the following endpoints:

$$\begin{aligned}
&r^*(\underline{g},\underline{f}) : P_1 = (1 - \textstyle\int_X \underline{g}, 1) && \text{and} \quad Q_1 = (1, 1 - \textstyle\int_X \underline{f})\ ,\\
&r^*(\underline{g},\bar{f}) : P_2 = (1 - \textstyle\int_X \underline{g}, \textstyle\int_X \bar{f}) && \text{and} \quad Q_2 = (1, 0)\ ,\\
&r^*(\bar{g},\underline{f}) : P_3 = (0, 1) && \text{and} \quad Q_3 = (\textstyle\int_X \bar{g}, 1 - \textstyle\int_X \underline{f})\ ,\\
&r^*(\bar{g},\bar{f}) : P_4 = (0, \textstyle\int_X \bar{f}) && \text{and} \quad Q_4 = (\textstyle\int_X \bar{g}, 0)\ .
\end{aligned} \tag{3}$$

Finally, denote by r^* the convex-hull of the four curves $r^*(\cdot,\cdot)$; r^* is the fat line shown in Fig. 3 connecting the endpoints $(0,1)$ and $(1,0)$. Then the main result of this paper is the following:

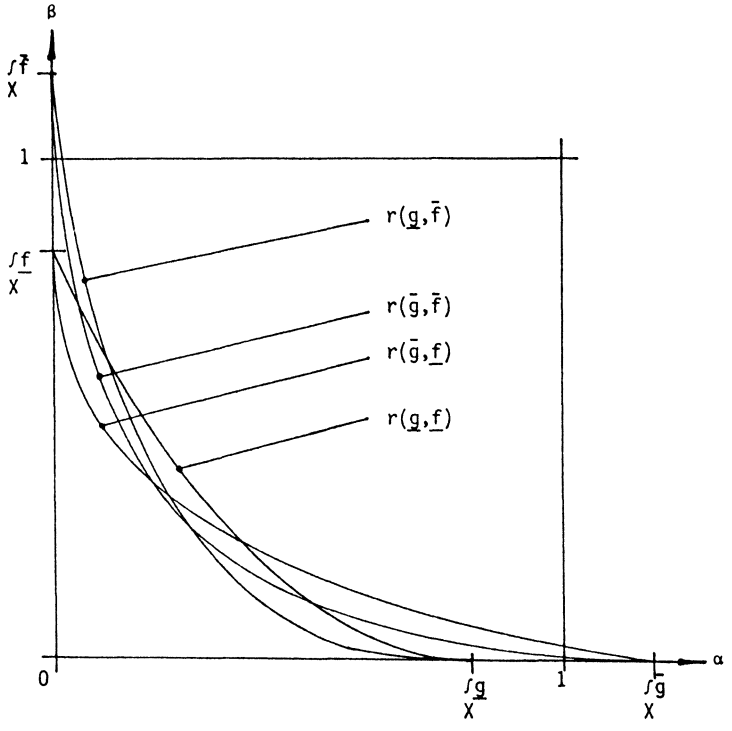

Fig. 2

Theorem: There exists a pair of admissible densities $g^* \in \mathcal{M}_0$ *and* $f^* \in \mathcal{M}_1$ *with* $r(g^*,f^*) = r^*$ *and the family of NP-tests* $\varphi^*_{c,\gamma}$ *for the simple test-problem* (g^*,f^*) *is minimax for the compound test-problem* $H_0: f \in \mathcal{M}_0$ *against* $H_1: f \in \mathcal{M}_1$, *i.e.:*

$$\alpha^*(c,\gamma) = E_{g^*}(\varphi^*_{c,\gamma}) \geq E_g(\varphi^*_{c,\gamma}) \quad \forall g \in \mathcal{M}_0 \ ,$$

$$1-\beta^*(c,\gamma) = E_{f^*}(\varphi^*_{c,\gamma}) \leq E_f(\varphi^*_{c,\gamma}) \quad \forall f \in \mathcal{M}_1 \ .$$

In fact three different situations are possible.

Case 1: r^* is the convex-hull of $\mathring{r}(\underline{g},\bar{f})$ and $\mathring{r}(\bar{g},\underline{f})$ alone, whereas $\mathring{r}(\underline{g},\underline{f})$ and $\mathring{r}(\bar{g},\bar{f})$ are situated above r^*. Figure 4 shows this situation and we refer to this figure in the following. r^* connects the points $(0,1) = P_3, R, S$ and $Q_2 = (1,0)$. The slope of the line-segment $\overline{RS}$ shall be denoted by $-c^*$.

In this case the density-ratio f^*/g^*, the knowledge of which is sufficient for the construction of the NP-tests $\varphi^*_{c,\gamma}$, is given by:

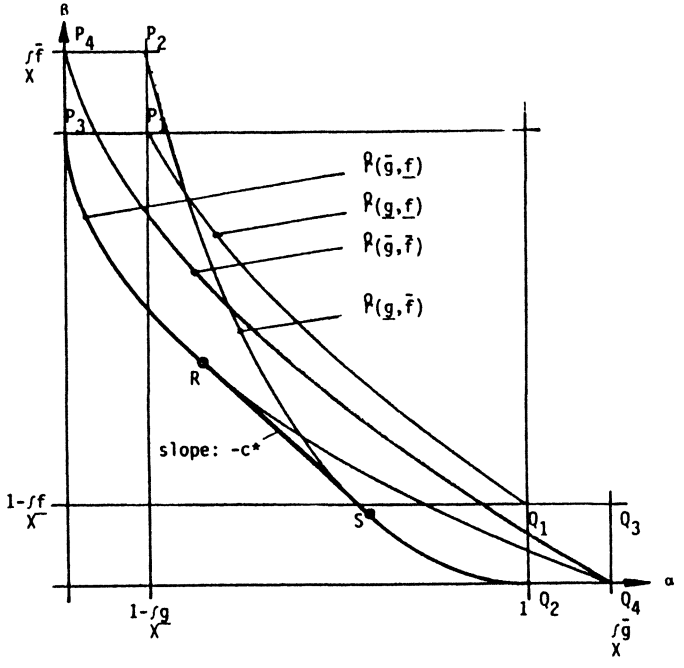

Fig. 3

$$f^*/g^* = \begin{cases} f/\bar{g} & \text{for} \quad x \in A = \{f/\bar{g} > c^*\} \,, \\ c^* & \text{for} \quad x \in B = \{f/\bar{g} \leq c^* \leq \bar{f}/g\} \,, \\ \bar{f}/g & \text{for} \quad x \in C = \{\bar{f}/g < c^*\} \,, \end{cases} \tag{4}$$

which can be written in the form:

$$f^*/g^* = \min\{\max\{f/\bar{g}, c^*\}, \bar{f}/g\} \,.$$

If (α, β) is a point of r^*, the corresponding test $\varphi^*_{c,\gamma}$ is determined by the slope $-c$ of r^* in this point. For $c \neq c^*$ it is:

$$\varphi^*_{c,\gamma} \equiv \varphi_{c,\gamma}(\bar{g}, f) \quad \text{for} \quad c > c^* \,,$$

$$\varphi^*_{c,\gamma} \equiv \varphi_{c,\gamma}(g, \bar{f}) \quad \text{for} \quad c < c^* \,.$$

The densities g^*, f^* are:

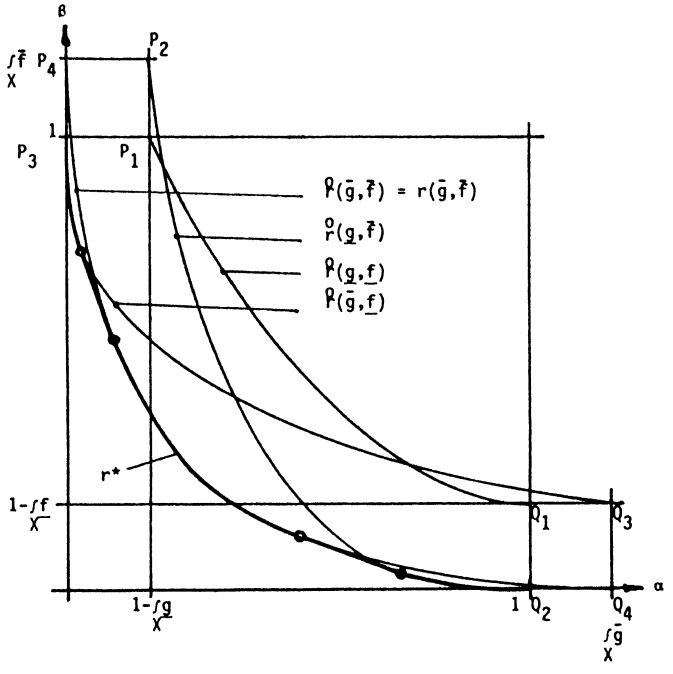

Fig. 4

$g^* =$	$f^* =$	for $x \in \ldots$
$\bar{g}$	$\underline{f}$	$A = \{c^* < \underline{f}/\bar{g}\}$
g	$c^* \cdot g$	$B = \{\underline{f}/\bar{g} \leq c^* \leq \bar{f}/\underline{g}\}$
$\underline{g}$	$\bar{f}$	$C = \{\bar{f}/\underline{g} < c^*\}$

(6)

where g has to be chosen according to the conditions:

$$\underline{p} := \max\{\underline{g}, \underline{f}/c^*\} \leq g \leq \min\{\bar{g}, \bar{f}/c^*\} =: \bar{p} \ , \tag{7}$$

$$\int_B g = 1 - \int_A \bar{g} - \int_C \underline{g} =: \Delta \ . \tag{8}$$

We will show, that these conditions can be fulfilled, that $(g^*, f^*) \in \mathcal{M}_0 \times \mathcal{M}_1$, i.e. (g^*, f^*) is admissible, and that $r^* = r(g^*, f^*)$.

Case 2: r^* is the convex hull of $r(\underline{g}, \bar{f})$, $r(\bar{g}, \underline{f})$ and $r(g, \underline{f})$ whereas $r(\bar{g}, \bar{f})$ is situated strictly above r^*. Figure 5 shows this situation and we refer to this figure in the following. r^* connects the points P_3, R, S, T, U, Q_2.

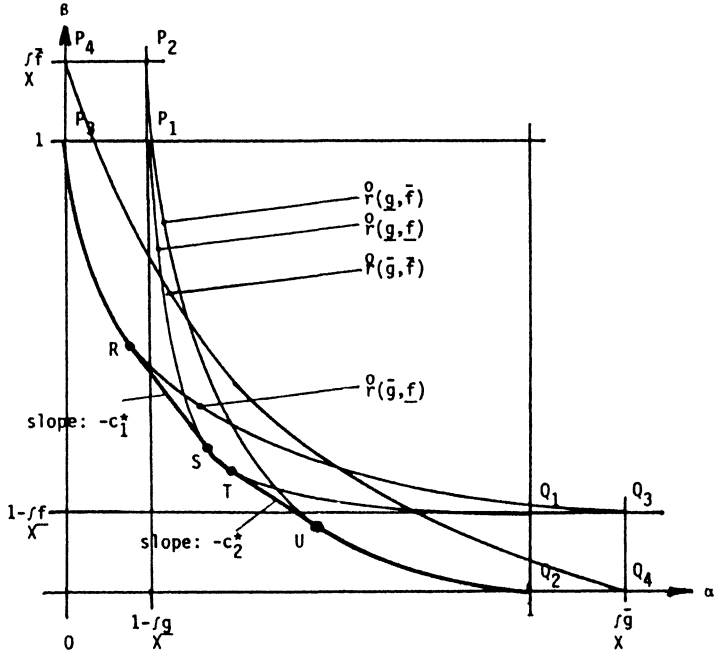

Fig. 5

The line-segments $\overline{RS}$ and $\overline{TU}$ may have slopes $-c_1^*$ and $-c_2^*$ respectively. In this case the densities g^*, f^* and the density-ratio f^*/g^* are:

$g^* =$	$f^* =$	$f^*/g^* =$	for $x \in \ldots$
$\bar{g}$	$\underline{f}$	$\underline{f}/\bar{g}$	$A = \{c_1^* < \underline{f}/\bar{g}\}$
$\underline{f}/c_1^*$	$\underline{f}$	c_1^*	$B = \{\underline{f}/\bar{g} \le c_1^* \le \underline{f}/\underline{g}\}$
$\underline{g}$	$\underline{f}$	$\underline{f}/\underline{g}$	$C = \{c_2^* < \underline{f}/\underline{g} < c_1^*\}$
$\underline{g}$	$c_2^* \cdot \underline{g}$	c_2^*	$D = \{\underline{f}/\underline{g} \le c_2^* \le \bar{f}/\underline{g}\}$
$\underline{g}$	$\bar{f}$	$\bar{f}/\underline{g}$	$E = \{\bar{f}/\underline{g} < c_2^*\}$

(9)

The ratio f^*/g^* can thus be written in the form:

$$f^*/g^* = \min\{\max\{\min\{\max(\underline{f}/\bar{g}, c_1^*), \underline{f}/\underline{g}\}, c_2^*\}, \bar{f}/\underline{g}\} \; . \tag{10}$$

or even shorter:

$$f^*/g^* = \operatorname{med}\{\underline{f}/\bar{g}, \underline{f}/\underline{g}, \bar{f}/\underline{g}, c_1^*, c_2^*\} \; ,$$

if med $\{a_1, \ldots, a_{2k+1}\}$ denotes the median of the $(2k+1)$-values $a_1, \ldots, a_{2k+1}$.

The pair (g^*, f^*) is admissible, i.e. $(g^*, f^*) \in \mathcal{M}_0 \times \mathcal{M}_1$, and if (α, β) is a point on r^* where the slope of r^* is $-c$, then for the cases $c_1^* < c$, $c_2^* < c < c_1^*$, $c < c_2^*$ the test $\varphi_{c,\gamma}^*$ is identical with $\varphi_{c,\gamma}(\bar{g}, \underline{f})$, $\varphi_{c,\gamma}(\underline{g}, \underline{f})$, $\varphi_{c,\gamma}(\underline{g}, \bar{f})$ respectively.

Case 3: r^* is the convex hull of $\mathring{r}(\underline{g}, \bar{f})$, $\mathring{r}(\bar{g}, \underline{f})$ and $\mathring{r}(\bar{g}, \bar{f})$, whereas $\mathring{r}(\underline{g}, \underline{f})$ is situated above r^*. This situation is shown in Fig. 6 to which we refer in the sequel. r^* connects the points P_3, R, S, T, U, Q_2 and the slopes of the line-segments $\overline{RS}$ and $\overline{TU}$ shall be $-c_1^*$ and $-c_2^*$ respectively. The densities g^*, f^* and the ratio f^*/g^* are:

$g^* =$	$f^* =$	$f^*/g^* =$	for $x \in \ldots$
$\bar{g}$	$\underline{f}$	$\underline{f}/\bar{g}$	$A = \{c_1^* < \underline{f}/\bar{g}\}$
$\bar{g}$	$c_1^* \cdot \bar{g}$	c_1^*	$B = \{\underline{f}/\bar{g} \le c_1^* \le \bar{f}/\bar{g}\}$
$\bar{g}$	$\bar{f}$	$\bar{f}/\bar{g}$	$C = \{c_2^* < \bar{f}/\bar{g} < c_1^*\}$
$\bar{f}/c_2^*$	$\bar{f}$	c_2^*	$D = \{\bar{f}/\bar{g} \le c_2^* \le \bar{f}/\underline{g}\}$
$\underline{g}$	$\bar{f}$	$\bar{f}/\underline{g}$	$E = \{\bar{f}/\underline{g} < c_2^*\}$

(11)

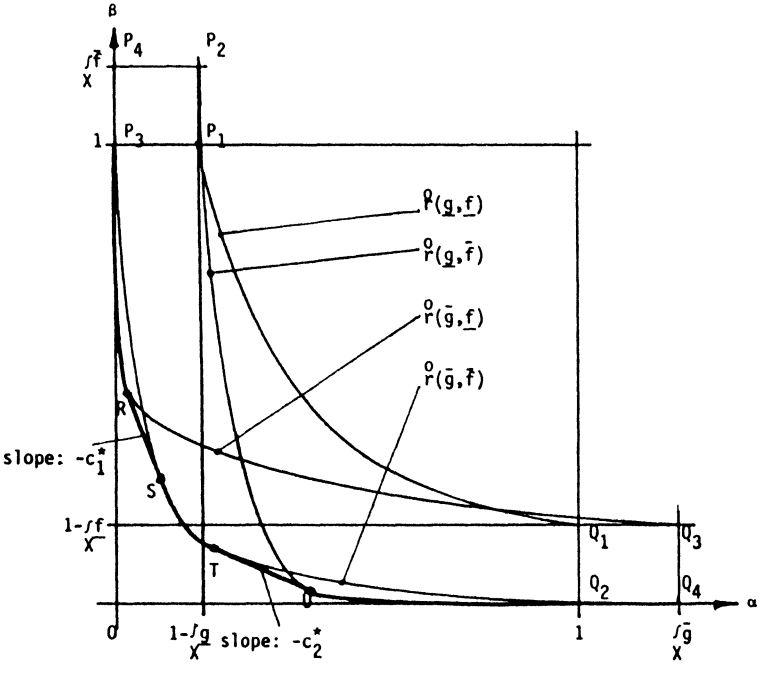

Fig. 6

The ratio f^*/g^* can here be written in the form:

$$f^*/g^* = \text{med}\{\underline{f}/\bar{g},\ \bar{f}/\bar{g},\bar{f}/\underline{g},c_1^*,c_2^*\}\ . \tag{12}$$

The pair (g^*,f^*) is admissible and if $(\alpha,\beta)\in r^*$ with the slope $-c$ in this point, then the corresponding test $\varphi_{c,\gamma}^*$ for the cases $c_1^*<c$, $c_2^*<c<c_1^*$, $c<c_2^*$ is identical with $\varphi_{c,\gamma}(\bar{g},\underline{f})$, $\varphi_{c,\gamma}(\bar{g},\bar{f})$ and $\varphi_{c,\gamma}(\underline{g},\bar{f})$ respectively.

The above Cases 1, 2 and 3 exhaust all possibilities, for we will show, that a situation, where all four curves $\dot{r}(\bar{g},\underline{f})$, $\dot{r}(\underline{g},\underline{f})$, $\dot{r}(\bar{g},\bar{f})$ and $\dot{r}(\underline{g},\bar{f})$ contribute to the hull r^* does not occur.

4 Proofs

We refer to Fig. 7 and denote the convex-hull of $\dot{r}(\underline{g},\bar{f})$ and $\dot{r}(\bar{g},\underline{f})$ by R^*. R^* connects the points $(0,1)=P_3,R,S,Q_2=(1,0)$, the slope of $\overline{RS}$ being $-c^*$. As in Case 1 of Section 3 we divide the sample-space X into three parts:

$$A=\{c^*<\underline{f}/\bar{g}\}\ ,\quad B=\{\underline{f}/\bar{g}\le c^*\le\bar{f}/\underline{g}\}\ ,\quad C=\{\bar{f}/\underline{g}<c^*\}\ , \tag{13}$$

and define two functions $\underline{p}(x)$ and $\bar{p}(x)$ on B:

$$\underline{p}=\max\{\underline{g},\underline{f}/c^*\}\quad\text{and}\quad\bar{p}(x)=\min\{\bar{g},\bar{f}/c^*\}\ . \tag{14}$$

Since by definition of B in (13) we have $\underline{f}/c^*\le\bar{g}$ and $\underline{g}\le\bar{f}/c^*$ on B and thus:

$$\underline{p}\le\bar{p}\quad\text{on}\quad B\ . \tag{15}$$

Finally we define (see Fig. 7):

$$\Delta:=1-\int_A\bar{g}-\int_C\underline{g}\ . \tag{16}$$

If $(\alpha(R),\beta(R))$ and $(\alpha(S),\beta(S))$ are the (α,β)-coordinates of R and S we have from the definition of the risk-function:

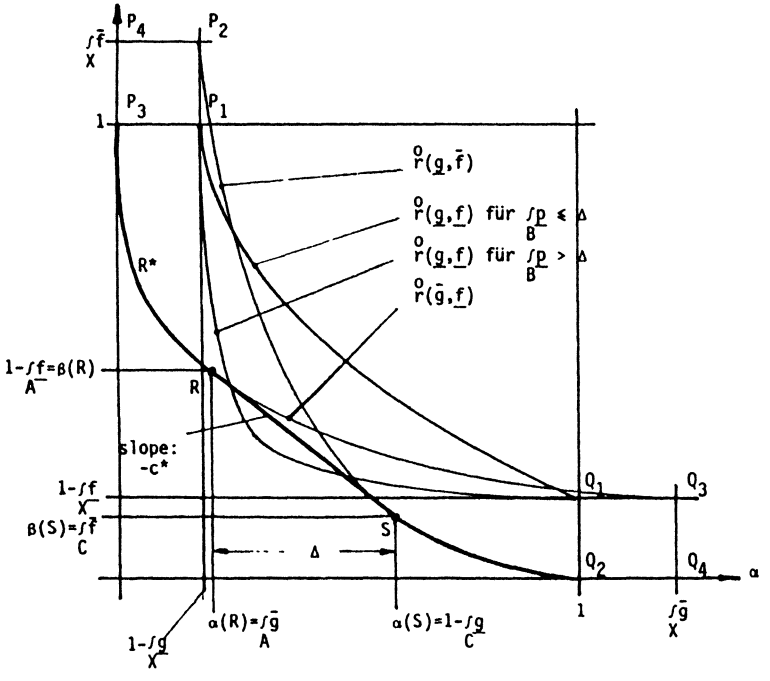

Fig. 7

$$\alpha(R) = \int_A \bar{g}\ , \quad \alpha(S) = 1 - \int_C \underline{g}\ , \quad \alpha(S) - \alpha(R) = \Delta$$

$$\beta(R) = 1 - \int_A \underline{f}\ , \quad \beta(S) = \int_C \bar{f}\ , \quad \beta(R) - \beta(S) = \Delta \cdot c^*\ . \tag{17}$$

We show:

Lemma 1: Depending upon $\int_B p \leq \Delta$ *or* $> \Delta$ *the curve* $r^*(\underline{g},\underline{f})$ *is situated totally above or partly below* R^* (*both cases are shown in Fig. 7*).

Proof: The case $\int_B p \leq \Delta$:

R^* can be interpreted as risk-function of the simple test-problem (g^*, f^*), the densities g^*, f^* being defined by:

$g^* =$	$f^* =$	for $x \in \ldots$
$\bar{g}$	$\underline{f}$	$A = \{c^* < \underline{f}/\bar{g}\}$
g	$c^* \cdot g$	$B = \{\underline{f}/\bar{g} \leq c^* \leq \bar{f}/\underline{g}\}$
$\underline{g}$	$\bar{f}$	$C = \{\bar{f}/\underline{g} < c^*\}$

(18)

where we choose g such, that $g \geq \underline{p}$ and $\int_B g = \Delta$.

Note that (g^*, f^*) in general is not admissible, i.e. not element of $\mathcal{M}_0 \times \mathcal{M}_1$, for this is only the case if $g \leq \bar{p}$ and this in turn can only be achieved, if $\int_B \bar{p} \geq \Delta$, which however has not been presupposed. The condition $\int_B g = \Delta$ however guarantees, that g^* and f^* are probability densities (see also Fig. 7).

We now consider the risk-function $r(\underline{g}, f^*)$ and shift it to connect the points $P_1 = (1 - \int_X \underline{g}, 1)$ and $Q_2 = (1, 0)$. The shifted curve is denoted by $\mathring{r}(\underline{g}, f^*)$. Figure 8 shows $R^* = r(g^*, f^*)$, $\mathring{r}(\underline{g}, f^*)$ and $\mathring{r}(\underline{g}, \underline{f})$.

We intend to show, that the relative position of the three curves R^*, $\mathring{r}(\underline{g}, f^*)$ and $\mathring{r}(\underline{g}, \underline{f})$ is as shown in Fig. 8, i.e. $\mathring{r}(\underline{g}, \underline{f})$ is situated above $\mathring{r}(\underline{g}, f^*)$ and this in turn is above R^* so that the first part of Lemma 1 will be proved.

For this purpose we choose $\mathring{\beta} \in (0, 1)$ arbitrarily and denote by $T = (\mathring{\alpha}, \mathring{\beta})$ and $T' = (\mathring{\alpha}', \mathring{\beta}' = \mathring{\beta})$ the points on R^* and $\mathring{r}(\underline{g}, f^*)$ with second coordinate $\mathring{\beta}$ (see Fig. 8). From (2) we have

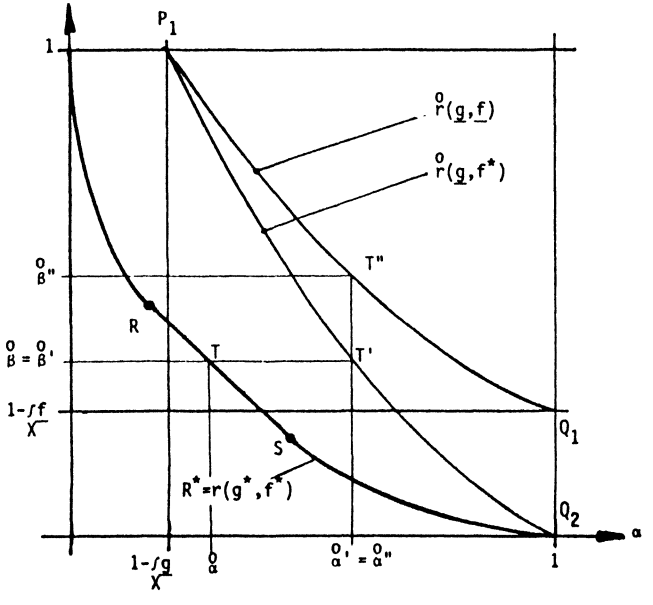

Fig. 8

$$\begin{aligned} \mathring{\alpha} &= \inf_{\Phi}\{E_{g^*}(\varphi): E_{f^*}(1-\varphi) = \mathring{\beta}\}\ , \\ \mathring{\alpha}' &= \inf_{\Phi}\{E_{\underline{g}}(\varphi): E_{f^*}(1-\varphi) = \mathring{\beta}\} + (1 - \int_X \underline{g})\ . \end{aligned} \tag{19}$$

Because of $g^* \geq \underline{g}$ (see (14) and (18)) we have:

$$E_{g^*}(\varphi) - E_{\underline{g}}(\varphi) \leq E_{g^*}(1) - E_{\underline{g}}(1) = 1 - \int_X \underline{g} \tag{20}$$

and thus $\mathring{\alpha} \leq \mathring{\alpha}'$. Consequently $\mathring{r}(\underline{g}, f^*)$ is situated to the right and as such above R^*.

Choose now $\mathring{\alpha}' \in [1 - \int_X \underline{g}, 1]$ and denote by $T' = (\mathring{\alpha}', \mathring{\beta}')$ and $T'' = (\mathring{\alpha}'' = \mathring{\alpha}', \mathring{\beta}'')$ the points on $\mathring{r}(\underline{g}, f^*)$ and $\mathring{r}(\underline{g}, \underline{f})$ with abscissa $\mathring{\alpha}'$ (see Fig. 8), then again from (2) we have:

$$\begin{aligned} \mathring{\beta}' &= \inf_{\Phi}\{E_{f^*}(1-\varphi): E_{\underline{g}}(1-\varphi) = 1 - \mathring{\alpha}'\}\ , \\ \mathring{\beta}'' &= \inf_{\Phi}\{E_{\underline{f}}(1-\varphi): E_{\underline{g}}(1-\varphi) = 1 - \mathring{\alpha}'\} + 1 - \int_X \underline{f}\ . \end{aligned} \tag{21}$$

From $f^* \geq \underline{f}$ (see (18) and (14)) we conclude:

$$E_{f^*}(1-\varphi) - E_{\underline{f}}(1-\varphi) \leq E_{f^*}(1) - E_{\underline{f}}(1) = 1 - \int_X \underline{f} \tag{22}$$

and from this $\mathring{\beta}' \leq \mathring{\beta}''$. As a consequence $\mathring{r}(\underline{g}, \underline{f})$ is seen to lie above $r(g^*, f^*)$ and thus also above R^*. This concludes the first part of the proof.

The case $\int_B \underline{p} > \Delta$:
We choose (g^*, f^*) according to:

$g^* =$	$f^* =$	for $x \in \ldots$
$\bar{g}$	$\underline{f}$	$A = \{c^* < \underline{f}/\bar{g}\}$
$\delta \cdot \underline{p}$	$c^* \cdot \delta \underline{p}$	$B = \{\underline{f}/\bar{g} \leq c^* \leq \bar{f}/\underline{g}\}$
$\underline{g}$	$\bar{f}$	$C = \{\bar{f}/\underline{g} < c^*\}$

(23)

with $\delta = \Delta / \int_B \underline{p} < 1$. Then g^*, f^* are probability-densities and $R^* = r(g^*, f^*)$. Because of $\delta < 1$ the pair (g^*, f^*) however is not admissible.

Consider the set

$$D = \{\underline{f}/\underline{g} > c^*\} \ . \tag{24}$$

Because of $\underline{f}/\bar{g} \leq \underline{f}/\underline{g} \leq \bar{f}/\underline{g}$ we have:

$$A \subset D \subset A \cup B \Leftrightarrow B \cup C \supset D^c \supset C \ . \tag{25}$$

Thus from the definition of $\underline{p}$ in (14) and from (23) we see:

$$f^* = \delta \cdot \underline{f} \quad \text{on} \quad D \cap B \ , \quad g^* = \delta \cdot \underline{g} \quad \text{on} \quad D^c \cap B \ . \tag{26}$$

The test $\varphi = I_D$ is NP-test for the generalized test-problem $(\underline{g},\underline{f})$ and, considered as a test for the problem (g^*,f^*), is equivalent to $\varphi' = I_A + \gamma I_B$ if $\gamma = P_{g^*}(D \cap B)/P_{g^*}(B)$. Therefore to φ there corresponds a point $U = (\alpha(U),\beta(U))$ on $R^* = r(g^*,f^*)$ and a point $U' = (\alpha(U'),\beta(U'))$ on $\mathring{r}(\underline{g},\underline{f})$. In U and U' the slopes of R^* and $\mathring{r}(\underline{g},\underline{f})$ are both $-c^*$ (see Fig. 9). It is:

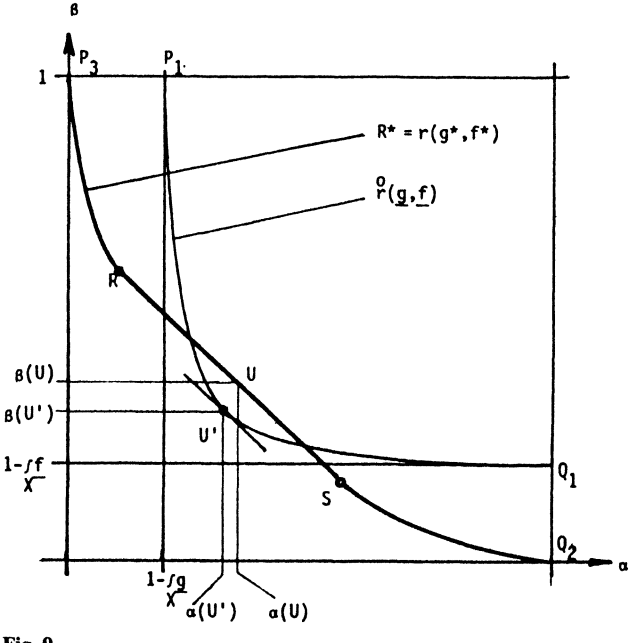

Fig. 9

$$
\begin{aligned}
(\alpha(U),\beta(U)) &= (P_{g^*}(D),P_{f^*}(D^c)) \ , \\
(\alpha(U'),\beta(U')) &= (P_{\underline{g}}(D)+1-\int_X \underline{g},P_{\underline{f}}(D^c)+1-\int_X \underline{f}) \ .
\end{aligned} \tag{27}
$$

From (23), (25), (26) and (27) we conclude:

$$
\begin{aligned}
\alpha(U)-\alpha(U') &= P_{g^*}(D)-P_{\underline{g}}(D)-1+\int_X \underline{g} = P_{\underline{g}}(D^c)-P_{g^*}(D^c) \\
&= (1-\delta)\cdot P_{\underline{g}}(D^c\cap B)\geq 0 \ , \\
\beta(U)-\beta(U') &= P_{f^*}(D^c)-P_{\underline{f}}(D^c)-1+\int_X \underline{f} = P_{\underline{f}}(D)-P_{f^*}(D) \\
&= (1-\delta)\cdot P_{\underline{f}}(D\cap B)\geq 0 \ ,
\end{aligned} \tag{28}
$$

where, because of $0\leq\Delta<\int_B \underline{p} = P_{\underline{g}}(D^c\cap B)+P_{\underline{f}}(D\cap B)/c^*$ at least one of the inequalities (28) is strict. From this we see that $\bar{U}$ is situated below R^* and thus $\dot{r}(\underline{g},\underline{f})$ partly reaches below R^* as shown in Fig. 9. This concludes the proof of Lemma 1. ♠

Along the same lines of argumentation one can show

Lemma 2: Depending upon $\int_B \bar{p}\geq\Delta$ or $<\Delta$ the curve $\dot{r}(\bar{g},\bar{f})$ is situated totally above or partly below R^*.

The Lemmata 1 and 2 show, that three possible cases have to be distinguished:
1. $\int_B \underline{p}\leq\Delta\leq\int_B \bar{p}$: $\dot{r}(\underline{g},\underline{f})$ and $\dot{r}(\bar{g},\bar{f})$ are both situated above R^*. Figure 4 shows this case.
2. $\Delta<\int_B \underline{p}\leq\int_B \bar{p}$: $\dot{r}(\bar{g},\bar{f})$ is above and $\dot{r}(\underline{g},\underline{f})$ partly below R^*. Figure 5 shows this case.
3. $\int_B \underline{p}\leq\int_B \bar{p}<\Delta$: $\dot{r}(\underline{g},\underline{f})$ is above and $\dot{r}(\bar{g},\bar{f})$ partly below R^*. This case is shown in Fig. 6.

We discuss the three cases in detail:

1. $\int_B \underline{p}\leq\Delta\leq\int_B \bar{p}$:

We refer to Section 3. Suppose g^* and f^* are defined according to (6), (7) and (8). From $\int_B \underline{p}\leq\Delta\leq\int_B \bar{p}$ it is clear that conditions (7) and (8) can be fulfilled. Furthermore we have $R^* = r(g^*,f^*)$ and $(g^*,f^*)\in\mathcal{M}_0\times\mathcal{M}_1$, i.e. the pair (g^*,f^*) is admissible. We show that it is also least favourable.

For this purpose consider the family $(\varphi^*_{c,\gamma})$ of NP-tests for (g^*,f^*). For $c>c^*$ we have (see (6)):

$$\{\varphi^*_{c,\gamma}>0\} \subset A \ , \quad g^* = \bar{g} \quad \text{and} \quad f^* = \underline{f} \quad \text{on} \quad A \ .$$

From this follows:

$$E_g(\varphi^*_{c,\gamma}) \leq E_{g^*}(\varphi^*_{c,\gamma}) \ , \quad E_f(\varphi^*_{c,\gamma}) \geq E_{f^*}(\varphi^*_{c,\gamma}) \quad \forall (g,f) \in \mathcal{M}_0 \times \mathcal{M}_1 \ . \tag{29}$$

For $c<c^*$ we have (see (6)):

$$\{1-\varphi^*_{c,\gamma}>0\} \subset C \ , \quad g^* = \underline{g} \quad \text{and} \quad f^* = \bar{f} \quad \text{on} \quad C \ .$$

From this:

$$E_g(1-\varphi^*_{c,\gamma}) \geq E_{g^*}(1-\varphi^*_{c,\gamma}) \ , \quad E_f(1-\varphi^*_{c,\gamma}) \leq E_{f^*}(1-\varphi^*_{c,\gamma}) \quad \forall (f,g) \in \mathcal{M}_0 \times \mathcal{M}_1 \ . \tag{30}$$

Finally for $c = c^*$ we have (see (6)): $\varphi^*_{c^*,\gamma} = I_A + \gamma I_B$. Thus:

$$\begin{aligned} E_g(\varphi^*_{c^*,\gamma}) &= P_g(A) + \gamma P_g(B) = (1-\gamma) P_g(A) + \gamma \cdot (1 - P_g(C)) \\ &\leq (1-\gamma) P_{g^*}(A) + \gamma \cdot (1 - P_{g^*}(C)) = E_{g^*}(\varphi^*_{c^*,\gamma}) \quad \forall g \in \mathcal{M}_0 \ , \end{aligned} \tag{31}$$

and analogously:

$$E_f(\varphi^*_{c^*,\gamma}) \geq E_{f^*}(\varphi^*_{c^*,\gamma}) \quad \forall f \in \mathcal{M}_1 \ . \tag{32}$$

(29), (30), (31) and (32) show that (g^*,f^*) is a pair of least favourable distributions in $\mathcal{M}_0 \times \mathcal{M}_1$ and that the test-family $(\varphi^*_{c,\gamma} = \varphi_{c,\gamma}(g^*,f^*))$ is minimax for the problem $(\mathcal{M}_0, \mathcal{M}_1)$.

2. $\Delta < \int_B \underline{p} \leq \int_B \bar{p}$:

This case is shown in Fig. 5. Here r^* is the convex-hull of $r(\bar{g},\underline{f})$, $r(\underline{g},\bar{f})$ and $r(\underline{g},\underline{f})$. Let g^* and f^* be defined according to (9), so that $\underline{g} \leq g^* \leq \bar{g}$ and $\underline{f} \leq f^* \leq \bar{f}$. We show $r(g^*,f^*) = r^*$.

From the definition of (g^*,f^*) in (9) it is evident, that $r(g^*,f^*)$ contains the arcs $\overline{P_3 R}$, $\overline{ST}$ and $\overline{UQ_2}$ in Fig. 5. It is also evident that between these arcs $r(g^*,f^*)$ contains line-segments of slopes $-c_1^*$ and $-c_2^*$. What is not evident is the fact that these line-segments are identical with $\overline{RS}$ and $\overline{TU}$, i.e. that they have the same length.

The β-projection of the line-segment with slope $-c_1^*$ of $r(g^*,f^*)$ is, with the notation of (9), obviously $P_{f^*}(B) = P_{\underline{f}}(B)$. On the other hand the β-projection of $\overline{RS}$ is, since $S \in \dot{r}(\underline{g},\underline{f})$ and $R \in \dot{r}(\bar{g},\underline{f})$ and $\overline{RS}$ is tangent to both these risk-functions, given by (see Fig. 5):

$$\beta(R)-\beta(S) = \left[\left(1-\int_X \underline{f}\right)+P_{\underline{f}}(A^c)\right] - \left[\left(1-\int_X \underline{f}\right)+P_{\underline{f}}((A\cup B)^c)\right] = P_{\underline{f}}(B) \ .$$

Thus the line-segment with slope $-c_1^*$ of $r(g^*,f^*)$ and $\overline{RS}$ have equal length. In the same manner we see that the line segment with slope $-c_2^*$ of $r(g^*,f^*)$ and $\overline{TU}$ have equal length. From this finally follows $r(g^*,f^*) = r^*$ and especially $(g^*,f^*) \in \mathcal{M}_0 \times \mathcal{M}_1$.

In order to see that (g^*,f^*) is least favourable consider the family of NP-tests $(\varphi^*_{c,\gamma} = \varphi_{c,\gamma}(g^*,f^*))$.

A. $c_1^* < c$: Since $\{\varphi^*_{c,\gamma} > 0\} \subset A$; $g^* = \bar{g}$ and $f^* = \underline{f}$ on A (see (9)). We have:

$$E_g(\varphi^*_{c,\gamma}) \le E_{g^*}(\varphi^*_{c,\gamma}) \quad \text{and} \quad E_f(\varphi^*_{c,\gamma}) \ge E_{f^*}(\varphi^*_{c,\gamma}) \quad \forall (g,f) \in \mathcal{M}_0 \times \mathcal{M}_1 \ . \tag{33}$$

B. $c_2^* < c < c_1^*$: (9) shows $\{\varphi^*_{c,\gamma} > 0\} \subset A \cup B \cup C$, $f^* = \underline{f}$ on $A \cup B \cup C$ and $\{1-\varphi^*_{c,\gamma} > 0\} \subset C \cup D \cup E$, $g^* = \underline{g}$ on $C \cup D \cup E$. From this we conclude:

$$E_g(1-\varphi^*_{c,\gamma}) \ge E_{g^*}(1-\varphi^*_{c,\gamma}) \quad \text{and} \quad E_f(\varphi^*_{c,\gamma}) \ge E_{f^*}(\varphi^*_{c,\gamma}) \quad \forall (g,f) \in \mathcal{M}_0 \times \mathcal{M}_1 \ . \tag{34}$$

C. $c < c_2^*$: Here $\{1-\varphi^*_{c,\gamma} > 0\} \subset E$; $g^* = \underline{g}$ and $f^* = \bar{f}$ on E (see 9)) and:

$$\begin{aligned} &E_g(1-\varphi^*_{c,\gamma}) \ge E_{g^*}(1-\varphi^*_{c,\gamma}) \ , \\ &E_f(1-\varphi^*_{c,\gamma}) \le E_{f^*}(1-\varphi^*_{c,\gamma}) \quad \forall (g,f) \in \mathcal{M}_0 \times \mathcal{M}_1 \ . \end{aligned} \tag{35}$$

D. $c = c_1^*$: In this case $\varphi^*_{c,\gamma} = I_A + \gamma I_B$, with $f^* = \underline{f}$ on $A \cup B$, $g^* = \bar{g}$ on A and $g^* = \underline{g}$ on $C \cup D \cup E = (A \cup B)^c$ (see (9)). Thus we have:

$$E_f(\varphi^*_{c,\gamma}) \ge E_{f^*}(\varphi^*_{c,\gamma}) \quad \forall f \in \mathcal{M}_1 \ ,$$

$$E_g(\varphi^*_{c,\gamma}) = P_g(A)+\gamma P_g(B) = (1-\gamma)P_g(A)+\gamma(1-P_g(C\cup D\cup E))$$

$$\leq (1-\gamma)P_{g^*}(A)+\gamma(1-P_{g^*}(C\cup D\cup E)) = E_{g^*}(\varphi^*_{c,\gamma}) \quad \forall g\in\mathcal{M}_0 \ . \tag{36}$$

E. $c=c_2^*$: Here $1-\varphi^*_{c,\gamma} = I_E+(1-\gamma)I_D$, $f^* = \underline{f}$ on $A\cup B\cup C = (D\cup E)^c$, $f^* = \bar{f}$ on E and $g^* = \underline{g}$ on $D\cup E$. From this:

$$E_g(1-\varphi^*_{c,\gamma}) \geq E_{g^*}(1-\varphi^*_{c,\gamma}) \quad \forall g\in\mathcal{M}_0 \ ,$$

$$E_f(1-\varphi^*_{c,\gamma}) = P_f(E)+(1-\gamma)P_f(D) = \gamma P_f(E)+(1-\gamma)(1-P_f(A\cup B\cup C))$$

$$\leq \gamma P_{f^*}(E)+(1-\gamma)(1-P_{f^*}(A\cup B\cup C))$$

$$= E_{f^*}(1-\varphi^*_{c,\gamma}) \quad \forall f\in\mathcal{M}_1 \ . \tag{37}$$

From (33)–(37) we conclude, that (g^*,f^*) is least favourable and the family of NP-tests $(\varphi^*_{c,\gamma} = \varphi_{c,\gamma}(g^*,f^*))$ is minimax for the test-problem $(\mathcal{M}_0, \mathcal{M}_1)$.

3. $\int_B \underline{p} \leq \int \bar{p} < \Delta$:

This case is shown in Fig. 6. The argumentation is similar to that of the previous Case 2. After proving that r^*, the convex hull of $\dot{r}(\underline{g},\bar{f})$, $\dot{r}(\bar{g},\underline{f})$ and $\dot{r}(\bar{g},\bar{f})$, is the risk-function of the pair (g^*,f^*) defined by (11), one shows that (g^*,f^*) is least favourable and the corresponding family of NP-tests $(\varphi^*_{c,\gamma} = \varphi_{c,\gamma}(g^*,f^*))$ is minimax for $(\mathcal{M}_0,\mathcal{M}_1)$.

5 Calculation of c^*, c_1^*, c_1^*

Although we have made extensive use of the graphical representation of risk-functions the decision which of the three possible cases is given and especially the determination of the slopes c^*, c_1^*, c_2^*, the knowledge of which is necessary and sufficient for the calculation of the density-quotient f^*/g^* and the test-family $(\varphi_{c,\gamma}(f^*,g^*))$, has to be carried out by numerical calculation. We list the necessary steps:

1. Calculation of c^*, the negative slope of the common tangent of $\dot{r}(\bar{g},\underline{f})$ and $\dot{r}(\underline{g},\bar{f})$. c^* is determined by the condition, that the line connecting the points $(\dot{\alpha}(c^*,0\,|\,\bar{g},\underline{f}), \dot{\beta}(c^*,0\,|\,\bar{g},\underline{f}))$ and $(\dot{\alpha}(c^*,0\,|\,\underline{g},\bar{f}), \dot{\beta}(c^*,0\,|\,\underline{g},\bar{f}))$ of $\dot{r}(\bar{g},\underline{f})$ and $\dot{r}(\underline{g},\bar{f})$ respectively must have slope $-c^*$:

$$\frac{\dot{\beta}(c^*,0|\bar{g},\underline{f})-\dot{\beta}(c^*,0|\underline{g},\bar{f})}{\dot{\alpha}(c^*,0|\underline{g},\bar{f})-\dot{\alpha}(c^*,0|\bar{g},\underline{f})}=c^* \ , \tag{38}$$

with:

$$\dot{\alpha}(c,0|\bar{g},\underline{f})=P_{\bar{g}}(\underline{f}/\bar{g}>c) \ , \qquad \dot{\beta}(c,0|\bar{g},\underline{f})=P_{\underline{f}}(\underline{f}/\bar{g}\leq c)+1-\int_X \underline{f} \ ,$$

$$\dot{\alpha}(c,0|\underline{g},\bar{f})=P_{\underline{g}}(\bar{f}/\underline{g}>c)+1-\int_X \underline{g} \ , \quad \dot{\beta}(c,0|\underline{g},\bar{f})=P_{\bar{f}}(\bar{f}/\underline{g}\leq c) \ .$$

Equation (38) has to be solved numerically.

2. In order to see, which of the three possible cases is given, calculate the coordinates:

$$\dot{\alpha}(c^*,0|\underline{g},\underline{f})=P_{\underline{g}}(\underline{f}/\underline{g}>c^*)+1-\int_X \underline{g} \ , \quad \dot{\alpha}(c^*,0|\bar{g},\bar{f})=P_{\bar{g}}(\bar{f}/\bar{g}>c^*) \ ,$$

$$\dot{\beta}(c^*,0|\underline{g},\underline{f})=P_{\underline{f}}(\underline{f}/\underline{g}\leq c^*)+1-\int_X \underline{f} \ , \quad \dot{\beta}(c^*,0|\bar{g},\bar{f})=P_{\bar{f}}(\bar{f}/\bar{g}\leq c^*) \ ,$$

and from these the slopes c' and c'' given by:

$$c'=\frac{\dot{\beta}(c^*,0|\bar{g},\underline{f})-\dot{\beta}(c^*,0|\underline{g},\underline{f})}{\dot{\alpha}(c^*,0|\underline{g},\underline{f})-\dot{\alpha}(c^*,0|\bar{g},\underline{f})} \ , \qquad c''=\frac{\dot{\beta}(c^*,0|\bar{g},\underline{f})-\dot{\beta}(c^*,0|\bar{g},\bar{f})}{\dot{\alpha}(c^*,0|\bar{g},\bar{f})-\dot{\alpha}(c^*,0|\bar{g},\underline{f})} \ .$$

Then we have the following implications:

A. $c'\leq c^*$ and $c''\leq c^*$: Case 1 of Section 3 is given.
B. $c''\leq c^*<c'$: Case 2 of Section 3 is given.
C. $c'\leq c^*<c''$: Case 3 of Section 3 is given.

The case $c^*<c',c''$ is impossible. If A holds no further calculations are necessary. For B we continue to 3 and for C to 4.

3. Calculate the slope c_1^* of the common tangent to $\dot{r}(\bar{g},\underline{f})$ and $\dot{r}(\underline{g},\underline{f})$ and the slope c_2^* of the common tangent to $\dot{r}(\underline{g},\underline{f})$ and $\dot{r}(\underline{g},\bar{f})$ by solving the appropriate equations (38).

4. Calculate the slope c_1^* of the common tangent to $\dot{r}(\bar{g},\underline{f})$ and $\dot{r}(\bar{g},\bar{f})$ and the slope c_2^* of the common tangent to $\dot{r}(\bar{g},\bar{f})$ and $\dot{r}(\underline{g},\bar{f})$ in the same way.

With these calculations carried out the density quotient $q^* = f^*/g^*$ and the corresponding family of NP-tests is known in all possible cases.

We conclude this section by mentioning the well known fact, that for a sample of size n the pair of n-dimensional densities $\left(\prod_{j=1}^{n} f^*(x_j), \prod_{j=1}^{n} g^*(x_j)\right)$ is least favourable. The minimax-tests can thus be determined from the density-quotient $q^*(x_1, \ldots, x_n) = \prod_{j=1}^{n} f^*(x_j)/g^*(x_j) = \prod_{j=1}^{n} q^*(x_j)$.

6 An Example

Suppose $\psi(x)$ is a probability-density on $\boldsymbol{R}$ with $\psi(x) = \psi(-x)$ and

$$\frac{\psi'(x)}{\psi(x)} = \frac{d \ln(\psi(x))}{dx} \ldots \quad \text{strictly nonincreasing} . \tag{39}$$

The corresponding distribution-function is denoted by $\Psi(x)$.

Consider the test-problem $H_0: f(x) = \psi(x+\mu)$ $H_1: f(x) = \psi(x-\mu)$. Its risk-function $r = r(\psi(x+\mu), \psi(x-\mu))$ is symmetric with respect to the line $\alpha = \beta$, i.e. if $(\alpha, \beta) \in r$ also $(\beta, \alpha) \in r$.

As a consequence of (39) the density-quotient $q(x) = \psi(x-\mu)/\psi(x+\mu)$ is strictly increasing and thus the NP-tests are of the form $\varphi_{c,\gamma}(x) = I_{[c', \infty]}(x)$.

Now for two constants $\underline{\Delta}$ and $\bar{\Delta}$ with $0 \leq \underline{\Delta} \leq 1 \leq \bar{\Delta}$ we introduce the bounds $\underline{g}, \bar{g}; \underline{f}, \bar{f}$:

$$\begin{aligned} \underline{g}(x) &= \underline{\Delta} \cdot \psi(x+\mu) , \quad \bar{g}(x) = \bar{\Delta} \cdot \psi(x+\mu) ; \\ \underline{f}(x) &= \underline{\Delta} \cdot \psi(x-\mu) , \quad \bar{f}(x) = \bar{\Delta} \cdot \psi(x-\mu) . \end{aligned} \tag{40}$$

We intend to show, that the subset $[0, 1] \times [1, \infty)$ of the $(\underline{\Delta}, \bar{\Delta})$-plane is divided into three parts corresponding to the three cases described in Sections 3 and 4.

In Fig. 10 the risk-functions $r = r(\psi(x+\mu), \psi(x-\mu))$, $r^*(\underline{g}, \bar{f})$, $r^*(\bar{g}, \underline{f})$, $r^*(\underline{g}, \underline{f})$ and $r^*(\bar{g}, \bar{f})$ are shown. The latter four functions obviously are generated by suitably stretching and shifting r. Because of symmetry the common tangent to $r^*(\underline{g}, \bar{f})$ and $r^*(\bar{g}, \underline{f})$, the line $\overline{RS}$, has slope $c^* = -1$.

If A, B, C are the points where r, $r^*(\underline{g}, \underline{f})$ and $r^*(\bar{g}, \bar{f})$ have slope $c^* = -1$, then we have the following possibilities:

Case 1: B and C are above $\overline{RS}$,

Case 2: B is below $\overline{RS}$ (and automatically C above $\overline{RS}$), (41)

Case 3: C is below $\overline{RS}$ (and automatically B above $\overline{RS}$) .

The conditions (41) therefore are equivalent to:

Case 1: $\alpha(R)+\alpha(S)\leq 2\alpha(B)$ and $\leq 2\alpha(C)$,

Case 2: $\alpha(R)+\alpha(S)>2\alpha(B)$,

Case 3: $\alpha(R)+\alpha(S)>2\alpha(C)$.

The coordinates of A, B, C, R and S are:

$$A=(\alpha(A),\beta(A))=(1-\Psi(\mu),1-\Psi(\mu)) ,$$

$$B=(\alpha(B),\beta(B))=(1-\underline{\Delta}\cdot\Psi(\mu),1-\underline{\Delta}\cdot\Psi(\mu)) ,$$

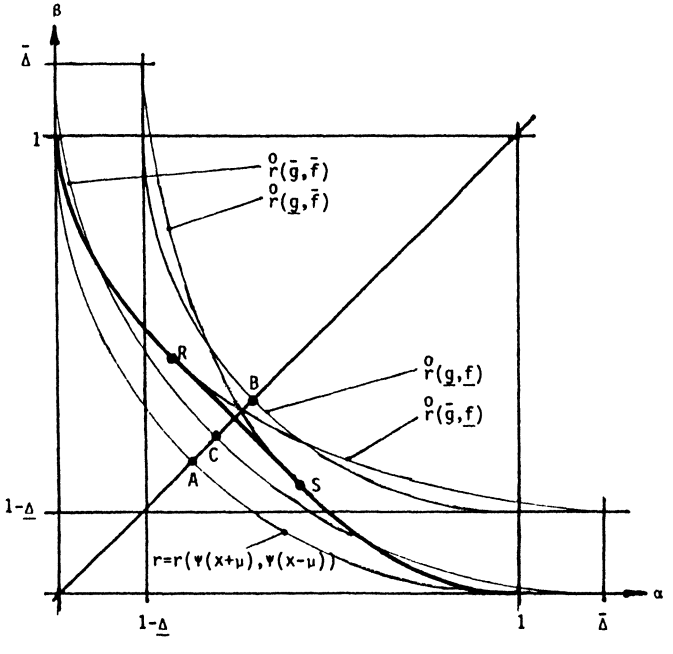

Fig. 10

$$C = (\alpha(C), \beta(C)) = (\bar{\Delta} \cdot (1 - \Psi(\mu)), \bar{\Delta} \cdot (1 - \Psi(\mu))) \ ,$$

$$R = (\alpha(R), \beta(R)) = (P_g(\underline{\Delta}/\bar{\Delta} \cdot q(x) > 1), P_f(\underline{\Delta}/\bar{\Delta} \cdot q(x) \leq 1) + 1 - \underline{\Delta})$$
$$= (\bar{\Delta} \cdot (1 - \Psi(\mu + q^{-1}(\bar{\Delta}/\underline{\Delta}))), 1 - \underline{\Delta} \cdot \Psi(\mu - q^{-1}(\bar{\Delta}/\underline{\Delta}))) \ ,$$

$$S = (\alpha(S), \beta(S)) = (\beta(R), \alpha(R)) \ ,$$

so that with the abbreviations:

$$z = \bar{\Delta}/\underline{\Delta} \ , \quad H(z) = \Psi(\mu - q^{-1}(z)) - z(1 - \Psi(\mu + q^{-1}(z))) \ ,$$

$$\delta_1(z) = 1/(H(z) + 2z(1 - \Psi(\mu))), \delta_2(z) = 1/(-H(z) + 2\Psi(\mu))$$

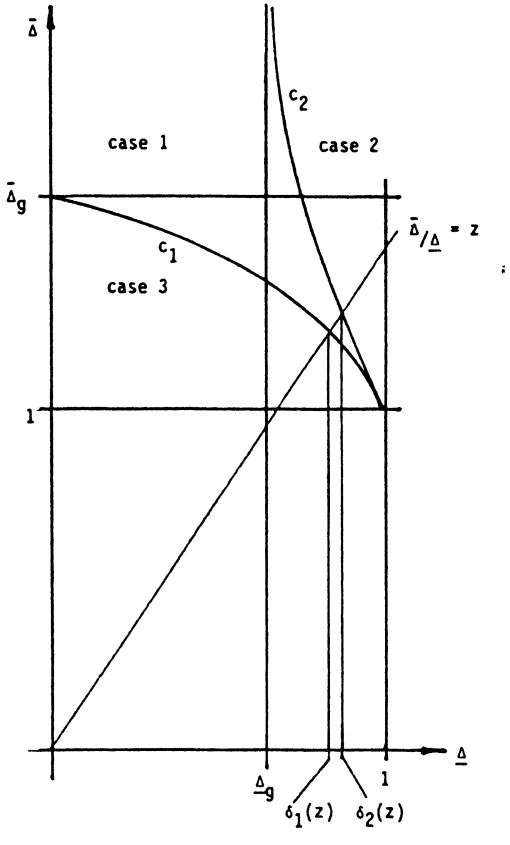

Fig. 11

we obtain:

Case 1: if $\delta_1(z) \leq \underline{\Delta} \leq \delta_2(z)$,

Case 2: if $\delta_2(z) < \underline{\Delta}$,

Case 3: if $\underline{\Delta} < \delta_1(z)$.

Figure 11 shows the situation. The curves C_1 and C_2 are given by the parameter-representations:

$$C_1 : (\underline{\Delta}(z), \bar{\Delta}(z)) = (\delta_1(z), z\delta_1(z)) ,$$

$$C_2 : (\underline{\Delta}(z), \bar{\Delta}(z)) = (\delta_2(z), z\delta_2(z)) .$$

The horizontal asymptote to C_1 and the vertical asymptote to C_2 are situated at $\bar{\Delta}_g = 1/2(1 - \Psi(\mu))$ and $\underline{\Delta}_g = 1/2\,\Psi(\mu)$ respectively.

The special case shown in Fig. 11 results if we choose the density of the $N(0,1)$-distribution for ψ and $\mu = 0,55$. In this case $r = r(\psi(x+\mu), \psi(x-\mu))$ is almost a circle.

References

[1] Kutznetsov VP (1982) Minimax Tests for Bounded Families of Distribution Densities. Theory Prob App 27:299–309

[2] Österreicher F (1978) On the Construction of Least Favourable Pairs of Distributions. Z Wahrscheindlichkeitstheorie verw Gebiete 43:49–55

[3] Hafner R (1982) Construction of Least Favourable Pairs of Distributions and of Robust Tests for Contamination Neighbourhoods. Math Operationsforsch Statist, Ser Statistics 13:47–56

[4] Hafner R (1982) Simple Construction of Least Favourable Pairs of Distributions and of Robust Tests for Prokhorov-Neighbourhoods. Math Operationsforsch Statist, Ser Statistics 13:33–46

[5] Hafner R (1987) Construction of Minimax-Tests for Bounded Families of Distribution-Functions. In: Sendler W (ed) Contributions to Stochastics. Physica-Verlag, Heidelberg

[6] Huber PJ, Strassen V (1973) Minimax Tests and the Neyman-Pearson Lemma for Capacities. Ann Statistics 1:251–263

[7] Sendler W (1971) Einige maßtheoretische Sätze zur Behandlung trennscharfer Tests. Z Wahrscheinlichkeitstheorie verw Gebiete 18:183–196

[8] Österreicher F, Thaler M (1978) The Fundamental Neyman-Pearson Lemma and the Radon-Nikodym Theorem from a Common Statistical Point of View. Int J Math Educ Sci Technol 9:163–176

Received 8 August 1991
Revised version 28 November 1991

Book Review

Pfanzagl J: *Estimation in semiparametric models.* Springer-Verlag 1990, 112 S., DM 30,–

This monography is an update of Pfanzagl and Wefelmeyer's important (1992) "Contributions to a general asymptotic statistical theory". Its subject is the same as that of the 1982 monograph, the estimation of Euclidean parameters in a nonparametric context. The emphasis is however now on construction of estimates rather than calculation of information bounds. In these notes Pfanzagl discusses a number of approaches to estimation, many of which have appeared in the literature in connection with special cases during the last 30–40 years but whose general applicability in a nonparametric context began to be recognized in 1982–89. In particular,

Section 3,4:

The sample splitting and discretization tricks applied to one step improvements introduced by LeCam and Hajek and developed in this context by Hasminskii and Ibragimov, Klaassen, Schick and the author of this review among others are described with care.

Section 5:

The main focus of this section is the construction of improvements (using one step estimates) with specified not necessarily efficient influence functions. The emphasis is on selecting the best within a family of influence functions smoothly indexed by an estimable Euclidean parameter. This approach has, for instance, been considered by Jaeckel (1971).

Section 6:

This is a critique of Godambe's and others' approach to efficiency via estimating equations.

Finally, Sections 7–9 contain a discussion of recent work of Pfanzagl's and van der Vaart's on estimation in mixture models. This includes some numerical results on the two sample scale mixture of exponentials model and a much studied model of Neyman and Scott which can be viewed as the simplest example of estimation of fixed effects in a mixed model (in the sense of Scheffé). Pfanzagl proposes what can be viewed as one step estimation using the method of sieves to estimate the ubiquitous (in these models) derivative of a log density. The numerical results are encouraging but rather limited.

The performance of method of sieves estimation is consistent with that found with a similar use of such procedures in theses by Faraway and Jin in Berkeley (1988 and 1990 respectively).

There are a number of interesting observations in these notes. Unfortunately the presentation is abstract and rather forbidding for nonspecialists in this realm of asymptotic methods.

Berkeley *P.J. Bickel*

Metrika (1993) 40:25–35

Metrika

A Fixed Sample Size Selection Procedure for Negative Binomial Populations

MADHURI S. MULEKAR

Dept. of Mathematics & Statistics, Faculty Court South # 3, University of South Alabama, Mobile, AL 36688, (205) 460-6264, USA

LINDA J. YOUNG

Department of Biometry, 103 Miller Hall, University of Nebraska, Lincoln NE 68583, (405) 744-5684, USA

Summary: A fixed sample size procedure for selecting the 'best' of k negative binomial populations is developed. Selection is made in such a way that the probability of correct selection is at least P^* whenever the distance between the probabilities of success is at least δ^*. The exponent r is assumed to be known and the same for all populations. Extensive computer calculations* were employed to obtain the exact least favorable configuration. The smallest sample sizes needed to meet specifications (P^*, δ^*) are tabulated for $r = 1(1)5$; $\delta^* = 0.05(0.05)0.55$ and $P^* = 0.75, 0.80, 0.90, 0.95, 0.98, 0.99$ involving $k = 3(1)6, 8, 10$ populations.

Key words and phrases: Fixed sample selection procedure, Negative binomial populations, Indifference zone approach, Least favorable configuration.

1 Introduction

The negative binomial distribution was used by Williamson and Brentherton (1964) to model industrial purchasing. If the purchasing occasions are distributed as a poisson distribution, same for all consumers and the amounts bought per occasion are distributed as a logarithmic series distribution, then the total amount purchased over all occasions follows a negative binomial distribution. A mathematical derivation of this model is given by Quenouille (1949). A criticism of this model is that it is inconsistent with general experience to suppose that different consumer's average purchasing patterns are the same. In the industrial or consumer purchasing experiments, the hypothesis that different groups, departments, or companies have the same purchase patterns is meaningless if the groups are actually different. In such a case, it seems obvious that the different groups

* All the computations were carried out on the Alabama Supercomputer.
Part of this work was completed when the authors were at the Department of Statistics, Oklahoma State University, Stillwater, OK 74078.

0026–1335/93/1/25–35 $2.50

will show differences in purchase amounts, however small in effect. Thus the hypothesis that there is no difference in purchase patterns of different groups is unrealistic, the real problem is to select the group with the largest average purchases.

Consider k ($k \geq 2$) negative binomial populations, such that each population π_i is characterized by a probability function $nb(x;p_i,r) = \binom{r+x-1}{r-1} p_i^r (1-p_i)^x$, ($i = 1,2,\ldots,k$). Hence each population π_i is associated with a fixed but unknown probability of success p_i ($0 < p_i < 1$, $i = 1,2,\ldots,k$). We assume that r is known or estimated from the past experiments and is the same for all populations. The estimation of r has been studied extensively by Shenton and Bowman (1967), Willson et al. (1984), and Clark and Perry (1989).

For the negative binomial, the probability of success is neither a location nor a rank parameter and the population variance depends on the unknown mean. With the standard error of sample mean unknown, this dependency makes it difficult to obtain the optimal sample size. To overcome this difficulty Chambers and Jarratt (1964) proposed large sample double-sampling procedure. In this procedure the unknown population variance is estimated using the first sample. Gupta and Nagel (1971) considered a selection rule in the framework of a subset selection problem and Bechhofer et al. (1968) developed a sequential selection procedure.

We propose a fixed sample procedure for selecting population with the largest probability of success using the indifference zone approach. Unlike the procedure of Chambers and Jarratt (1964), which uses large sample approximation, this procedure is based on the exact distribution of the statistics. In this procedure the Probability of Correct Selection (PCS) is minimized over all possible values of the largest probability of success to obtain the minimum sample size. The proposed procedure and the associated PCS are given in Section 2. The Least Favorable Configuration (LFC) and the monotonicity property of the PCS are derived in Section 3. The sufficient condition for the existence of the smallest sample size and the tables of sample sizes required to apply this procedure are given in Section 4.

2 Proposed Procedure and Probability of Correct Selection

Let $p_{[1]} \leq p_{[2]} \leq \ldots \leq p_{[k-1]} \leq p_{[k]}$ be the ranked values of p_i ($i = 1,2,\ldots,k$). Before the experimentation begins, the investigator is not aware of the association between $p_{[i]}$ and $\pi_1, \pi_2, \ldots,$ or π_k. The problem of selecting the 'best' population can be stated as the problem of selecting a population associated with $p_{[k]}$. The investigator is willing to accept any one of the t populations as the 'best' if

there are t ties for first place. The simple distance measure $\delta = p_{[k]} - p_{[k-1]}$ is used as an indicator of the true difference between the best and the second best population. For P^* and δ^* specified by the user, the goal is to select the population associated with $p_{[k]}$ such that the PCS $\geq P^*$ whenever $\delta \geq \delta^*$, $(0 < p_i < 1$, $k^{-1} < P^* \leq 1$ and $0 \leq \delta^* \leq 1)$.

Let X_{ij} $(j = 1, 2, \ldots, n)$ be independent observations from the population π_i $(i = 1, 2, \ldots, k)$, having probability mass functions $nb(x_i; p_i, r)$, respectively. Define $T_i = \sum X_{ij}$. Then $(T_1, T_2, \ldots, T_k)$ is an independent set of sufficient statistics. The selection rule is given as follows:

Selection Rule: Select the population π_i associated with the min $(T_1, T_2, \ldots, T_k)$ and randomize with equal probability if there are any ties.

After selecting the best population, the confidence statement can be made about the correctness of the selection. Suppose p_s is the true p-value of the population selected and p_u is the maximum p-value over all unselected populations. Then with confidence P^*, p_s and p_u satisfy $p_{[k]} - \delta^* \leq p_s \leq p_{[k]}$, i.e., $0 \leq p_{[k]} - p_s \leq \delta^*$ or equivalently, $p_{[k-1]} \leq p_u \leq p_{[k-1]} + \delta^*$, i.e., $0 \leq p_u - p_{[k-1]} \leq \delta^*$. The statement that the interval $[p_s, p_s + \delta^*]$ covers the true best p value with confidence P^* is equivalent to the previous statement. Since the procedure is defined by the common sample size, the confidence statement can also be made after the determination of n for specified (P^*, δ^*). For given confidence level P^*, the δ^* can be interpreted as the maximum error likely to be committed in the selection procedure.

Suppose $T_{(i)}$ corresponds to the population associated with $p_{[i]}$ $(i = 1, 2, \ldots, n)$. The experimenter is unaware of this association between $T_{(i)}$ and $p_{[i]}$. Since samples are drawn independently, for a fixed number of populations, under the LFC, $p_{[k]} - \delta^* = p_{[k-1]} = \ldots = p_{[2]} = p_{[1]}$,

$$\mathrm{PCS} = \sum_{x=0}^{\infty} nb(x; p_{[k]}, nr) \sum_{i=0}^{k-1} \frac{\binom{k-1}{i}}{1+i}$$

$$\times nb^i(x; p_{[k-1]}, nr)(1 - NB(x; p_{[k-1]}, nr))^{k-i-1}$$

where $NB(x; p, r) = \sum_{j=0}^{x} nb(j; p, r)$.

3 The Least Favorable Configuration

Since the conditions under which Mahamunulu (1967) has obtained the LFC for the selection procedures are not satisfied by the negative binomial distribution parametrized using p, we must proceed to obtain the LFC. Consider configurations $\Omega_A = \{p_{[k]} - \delta \geq p_{[k-1]} \geq \ldots \geq p_{[1]}\}$ and $\Omega_B = \{p_{[k]} - \delta = p_{[k-1]} = \ldots = p_{[1]}\}$. For any specified $p_{[k]} = p^0_{[k]}$ and $\delta = \delta^0$, if the selection procedure satisfies the specifications for configuration Ω_B then it also satisfies the specifications for configuration Ω_A. To prove that the LFC can be found using Ω_B, the following three lemmas are required.

Lemma 1: For any non-negative integer x, positive r and any θ $(0 \leq \theta \leq 1)$ independent of p, the function $H(x;p,\theta) = NB(x-1;p,r) + \theta nb(x;p,r)$ is a non-decreasing function of p $(0<p<1)$. It is strictly increasing on $(0,1)$ except for $\theta = 0$ and $x = 0$.

Proof: The proof is accomplished using the derivative of the cumulative distribution function of the binomial distribution and the relation between the binomial and negative binomial mass functions and the corresponding distribution functions.

Define the 'continuous negative binomial' (CNB) random variable Y to be uniformly distributed in the interval $(x-0.5, x+0.5)$ with the total probability on this interval equal to the probability that the discrete negative binomial (DNB) random variable X takes value x, $x = 0,1,2,\ldots$, i.e., $f(y;p,r) = nb(x;p,r)$.

Lemma 2: The PCS remains unchanged if each of the k DNB populations is replaced by the k CNB populations, i.e., PCS (CNB) = PCS (DNB).

Proof: Define $Y_{(i)}$ = the CNB random variable associated with the population with parameters $p_{[i]}$ and r, and $X_{(i)}$ = nearest integer to $Y_{(i)}$ for $i = 1,2,\ldots,k$. Thus $X_{(i)}$ is a DNB random variable with the same parameter as $Y_{(i)}$. Then for any configuration with $p_{[k]} > p_{[k-1]}$,

$$\text{PCS (CNB)} = \Pr\,(Y_{(k)} < Y_{(i)},\ i = 1,2,\ldots,k-1)$$

$$= \sum_{x_{(k)}=0}^{\infty} \int_{x_{(k)}-1/2}^{x_{(k)}+1/2} \Pr\,(y_{(k)} < Y_{(i)},\ i = 1,2,\ldots,k-1) f(y_{(k)};p_{[k]},r)\,dy_{(k)}.$$

Within any interval $(x_{(k)} - \frac{1}{2}, x_{(k)} + \frac{1}{2})$ we have

$$\Pr(y_{(k)} < Y_{(1)}, y_{(k)} < Y_{(2)}, \ldots, y_{(k)} < Y_{(k-1)})$$

$$= \Pr(x_{(k)} < X_{(1)}, x_{(k)} < X_{(2)}, \ldots, x_{(k)} < X_{(k-1)})$$

$$+ \frac{1}{2} \sum_{t=1}^{k-1} \Pr(X_{(t)} = x_{(k)}) \Pr(x_{(k)} < X_{(j)}, j = 1, 2, \ldots, k-1, j \neq t)$$

$$+ \ldots + \frac{1}{k} \Pr(X_{(1)} = X_{(2)} = \ldots = X_{(k-1)} = x_{(k)}) \ .$$

Then using this equality and simplifying the expression for PCS the required result is obtained.

Lemma 3: Let $F(y;p,r)$ be the cumulative distribution function of the CNB random variable. For any real $r>0$ and any y, $F(y;p,r)$ is a nondecreasing function of p. In particular, for $1/2<y<\infty$ it is a strictly increasing function of p.

Proof: Define $x = x(y) =$ the integer part of $(y+0.5)$ and $\theta = \theta(y) =$ the fractional part of $(y+0.5)$. Then the inverse function y is a single valued function of the pair (x,θ) such that $y(x,\theta) = x+\theta-0.5$ where $x = 0,1,2,\ldots$ and $0 \leq \theta \leq 1$. Thus $F(y;p,r) \equiv 0$ in p for $y \leq -1/2$. For any $y = y_0 > -1/2$, $F(y_0;p,r) = \int_{-1/2}^{y_0} f(y;p,r)\,dy = \sum_{x=0}^{x(y_0)-1} f(x;p,r) + \theta(y_0) f(x(y_0);p,r) = H(x_0;p,\theta_0)$. Therefore the result follows from Lemma 1.

Theorem: For a fixed $p_{[k]}$, the probability of correct selection is a strictly increasing function of each of the differences $p_{[k]} - p_{[i]}$ $(i = 1,2,\ldots,k-1)$.

Proof: Using Lemma 2, for any configuration with $p_{[k]} > p_{[k-1]}$

$$\text{PCS} = \int_{-1/2}^{\infty} \prod_{i=1}^{k-1} [1 - F(y;p_{[i]},r)] f(y;p_{[k]},r)\,dy \ .$$

Thus the result follows using Lemma 3.

The PCS being a continuous and bounded function over a closed interval, it attains its minimum at some point $p^L_{[k]}(\delta^*,n,k)$ in the interval $(\delta^*,1)$. Thus the configuration Ω_B with $\delta = \delta^*$ and $p_{[k]} = p^L_{[k]}$ is the LFC and it depends on the common sample size n.

4 The Smallest Sample Size Required to Meet Specifications

Due to the dependency of the population variance on the unknown mean, the standard error can not be determined. Therefore the PCS depends on δ^*, n, k and the location of $p_{[k]}$ in the parameter space. Using a conservative approach like Sobel and Huyett (1957) the PCS can be minimised in two stages. In the first stage, for a fixed $p_{[k]}$ and $\delta \geq \delta^*$, each $p_{[i]}$, $(i = 1,2,\ldots,k-1)$ is replaced by $p_{[k]} - \delta^*$. In the second stage, the PCS is minimized with respect to $p_{[k]}$. This procedure is repeated for each n until the smallest sample size for which the in-

Table 1. Number of Units Required per Process to Meet Specifications (P^*, δ^*), for $r = 1$

$\delta^* \backslash P^*$	0.75	0.80	0.90	0.95	0.98	0.99	0.75	0.80	0.90	0.95	0.98	0.99
	2[a]	2	3	4	6	7	3[d]	4	5	6	7	8
0.55	3[b]	3	4	5	6	7	4[e]	4	5	6	8	9
	3[c]	3	4	5	7	8	4[f]	4	5	6	8	9
	2	3	4	5	7	8	4	4	5	7	9	10
0.50	3	3	4	6	8	9	4	5	6	7	9	11
	3	4	5	6	8	9	4	5	6	8	10	11
	3	3	5	6	8	10	4	5	6	8	10	12
0.45	3	4	5	7	9	11	5	5	7	9	11	13
	4	4	6	8	10	12	5	6	8	9	12	14
	3	4	6	8	10	13	5	6	8	10	13	15
0.40	4	4	7	9	12	14	6	6	9	11	14	16
	4	5	7	9	12	15	6	7	9	12	15	17
	4	4	7	10	13	16	6	7	10	13	17	20
0.35	5	6	8	11	15	18	7	8	11	14	18	21
	5	6	9	12	16	19	8	9	12	15	19	22
	5	6	9	13	18	22	8	9	13	18	23	27
0.30	6	7	11	15	20	24	9	11	15	19	25	29
	7	8	12	16	22	26	10	12	16	21	26	30
	6	8	13	18	26	32	11	13	19	25	33	39
0.25	8	10	15	21	29	35	13	15	21	27	35	42
	10	11	17	23	31	37	14	16	23	29	37	44
	9	11	20	28	40	49	16	19	29	38	51	61
0.20	12	15	24	33	45	54	19	22	32	42	55	65
	14	17	27	36	48	58	21	25	35	45	58	68
	15	19	34	49	71	87	28	33	50	68	90	107
0.15	20	25	41	57	79	96	33	39	57	74	97	115
	24	30	46	63	85	102	37	43	61	79	103	120
	32	42	75	110	158	195	60	73	112	150	202	240
0.10	44	55	91	128	177	215	72	85	125	165	218	257
	53	65	103	141	191	229	80	94	136	177	230	269
	124	164	297	437	629	777	235	285	440	596	802	957
0.05	171	216	360	507	705	857	278	331	494	655	866	1024
	207	254	405	557	760	913	312	367	535	699	913	1073

[a] for $k = 3$; [b] for $k = 4$; [c] for $k = 5$; [d] for $k = 6$; [e] for $k = 8$; [f] for $k = 10$.

Table 2. Number of Units Required per Process to Meet Specifications (P^*, δ^*), for $r = 2$

$\delta^* \backslash P^*$	0.75	0.80	0.90	0.95	0.98	0.99	0.75	0.80	0.90	0.95	0.98	0.99
	1[a]	1	2	2	3	4	2[d]	2	3	3	4	4
0.55	2[b]	2	2	3	3	4	2[e]	2	3	3	4	5
	2[c]	2	2	3	4	4	2[f]	2	3	3	4	5
	1	2	2	3	4	4	2	2	3	4	5	5
0.50	2	2	2	3	4	5	2	3	3	4	5	6
	2	2	3	3	4	5	2	3	3	4	5	6
	2	2	3	3	4	5	2	3	3	4	5	6
0.45	2	2	3	4	5	6	3	3	4	5	6	7
	2	2	3	4	5	6	3	3	4	5	6	7
	2	2	3	4	5	7	3	3	4	5	7	8
0.40	2	2	4	5	6	7	3	3	5	6	7	8
	2	3	4	5	6	8	3	4	5	6	8	9
	2	2	4	5	7	8	3	4	5	7	9	10
0.35	3	3	4	6	8	9	4	4	6	7	9	11
	3	3	5	6	8	10	4	5	6	8	10	11
	3	3	5	7	9	11	4	5	7	9	12	14
0.30	3	4	6	8	10	12	5	6	8	10	13	15
	4	4	6	8	11	13	5	6	8	11	13	15
	3	4	7	9	13	16	6	7	10	13	17	20
0.25	4	5	8	11	15	18	7	8	11	14	18	21
	5	6	9	12	16	19	7	8	12	15	19	22
	5	6	10	14	20	25	8	10	15	19	26	31
0.20	6	8	12	17	23	27	10	11	16	21	28	33
	7	9	14	18	24	29	11	13	18	23	29	34
	8	10	17	25	36	44	14	17	25	34	45	54
0.15	10	13	21	29	40	48	17	20	29	37	49	58
	12	15	23	32	43	51	19	22	31	40	52	60
	16	21	38	55	79	98	30	37	56	75	101	120
0.10	22	28	46	64	89	108	36	43	63	83	109	129
	27	33	52	71	96	115	40	47	68	89	115	135
	62	82	149	219	315	389	118	143	220	298	401	479
0.05	86	108	186	254	353	429	139	166	247	328	433	512
	104	127	203	279	380	457	156	184	268	350	457	537

[a] for $k = 3$; [b] for $k = 4$; [c] for $k = 5$; [d] for $k = 6$; [e] for $k = 8$; [f] for $k = 10$.

fimum of PCS exceeds P^* is obtained. The optimal sample sizes required to meet specification (P^*, δ^*) are listed in Tables 1 – 5. Additional tables and plots of sample sizes and the PCS are given by Nagardeolekar (1988). Using these tables, the required number of observations per population for specified (P^*, δ^*) may be determined. For a fixed known n, these tables may be reversed to obtain the PCS that will be achieved by the procedure. The values not listed in tables can be obtained by interpolation, which can be carried out easily using the graphs of the PCS, presented by Nagardeolekar (1988), for given n and δ.

Table 3. Number of Units Required per Process to Meet Specifications (P^*, δ^*), for $r = 3$

$\delta^* \backslash P^*$	0.75	0.80	0.90	0.95	0.98	0.99	0.75	0.80	0.90	0.95	0.98	0.99
	1[a]	1	1	2	2	3	1[d]	2	2	2	3	3
0.55	1[b]	1	2	2	2	3	2[e]	2	2	2	3	3
	1[c]	1	2	2	3	3	2[f]	2	2	2	3	3
	1	1	2	2	3	3	2	2	2	3	3	4
0.50	1	1	2	2	3	3	2	2	2	3	3	4
	1	2	2	2	3	3	2	2	2	3	4	4
	1	1	2	2	3	4	2	2	2	3	4	4
0.45	1	2	2	3	3	4	2	2	3	3	4	5
	2	2	2	3	4	4	2	2	3	3	4	5
	1	2	2	3	4	5	2	2	3	4	5	5
0.40	2	2	3	3	4	5	2	2	3	4	5	6
	2	2	3	3	4	5	2	3	3	4	5	6
	2	2	3	4	5	6	2	3	4	5	6	7
0.35	2	2	3	4	5	6	3	3	4	5	6	7
	2	2	3	4	6	7	3	3	4	5	7	8
	2	2	3	5	6	8	3	3	5	6	8	9
0.30	2	3	4	5	7	8	3	4	5	7	9	10
	3	3	4	6	8	9	4	4	6	7	9	10
	2	3	5	6	9	11	4	5	7	9	11	13
0.25	3	4	5	7	10	12	5	5	7	9	12	14
	4	4	6	8	11	13	5	6	8	10	13	15
	3	4	7	10	14	17	6	7	10	13	17	21
0.20	4	5	8	11	15	18	7	8	11	14	19	22
	5	6	9	12	16	20	7	9	12	15	20	23
	5	7	12	17	24	29	10	11	17	23	30	36
0.15	7	9	14	19	27	32	11	13	19	25	33	39
	8	10	16	21	29	34	13	15	21	27	35	40
	11	14	25	37	53	65	20	25	38	50	68	80
0.10	15	19	31	43	59	72	24	29	42	55	73	86
	18	22	35	47	64	77	28	32	46	59	77	90
	42	55	99	146	210	259	79	95	147	199	268	319
0.05	57	72	120	169	235	286	93	111	165	219	289	342
	69	85	135	186	254	305	104	123	179	233	305	358

[a] for $k = 3$; [b] for $k = 4$; [c] for $k = 5$; [d] for $k = 6$; [e] for $k = 8$; [f] for $k = 10$.

Lemma 4: A sufficient condition for the existence of the required smallest sample size is

$$\lim_{n \to \infty} \frac{[\mathrm{Var}(T_{(1)}) + \mathrm{Var}(T_{(k)})]}{[E(T_{(1)}) - E(T_{(k)})]^2} = 0 \ ,$$

provided the infimum of the PCS exists for some $p_{[k]} = p^0_{[k]}$.

Table 4. Number of Units Required per Process to Meet Specifications (P^*, δ^*), for $r = 4$

$\delta^* \backslash P^*$	0.75	0.80	0.90	0.95	0.98	0.99	0.75	0.80	0.90	0.95	0.98	0.99
	1[a]	1	1	1	2	2	1[d]	1	2	2	2	2
0.55	1[b]	1	1	2	2	2	1[e]	1	2	2	2	3
	1[c]	1	1	2	2	2	1[f]	1	2	2	2	3
	1	1	1	2	2	2	1	1	2	2	3	3
0.50	1	1	1	2	2	3	1	2	2	2	3	3
	1	1	2	2	2	3	1	2	2	2	3	3
	1	1	2	2	2	3	1	2	2	2	3	3
0.45	1	1	2	2	3	3	2	2	2	3	3	4
	1	1	2	2	3	3	2	2	2	3	3	4
	1	1	2	2	3	4	2	2	2	3	4	4
0.40	1	1	2	3	3	4	2	2	3	3	4	4
	1	2	2	3	3	4	2	2	3	3	4	5
	1	1	2	3	4	4	2	2	3	4	5	5
0.35	2	2	2	3	4	5	2	2	3	4	5	6
	2	2	3	3	4	5	2	3	3	4	5	6
	2	2	3	4	5	6	2	3	4	5	6	7
0.30	2	2	3	4	5	6	3	3	4	5	7	8
	2	2	3	4	6	7	3	3	4	6	7	8
	2	2	4	5	7	8	3	4	5	7	9	10
0.25	2	3	4	6	8	9	4	4	6	7	9	11
	3	3	5	6	8	10	4	4	6	8	10	11
	3	3	5	7	10	13	4	5	8	10	13	16
0.20	3	4	6	9	12	14	5	6	8	11	14	17
	4	5	7	9	12	15	6	7	9	12	15	17
	4	5	9	13	18	22	7	9	13	17	23	27
0.15	5	7	11	15	20	24	9	10	15	19	25	29
	6	8	12	16	22	26	10	11	16	20	26	30
	8	11	19	28	40	49	15	19	28	38	51	60
0.10	11	14	23	32	45	54	18	22	32	42	55	65
	14	17	26	36	48	58	20	24	34	45	58	68
	31	41	75	110	158	195	59	72	110	149	201	240
0.05	43	54	90	127	177	215	70	83	124	164	217	256
	52	64	102	140	190	229	78	92	134	175	229	269

[a]for $k = 3$; [b]for $k = 4$; [c]for $k = 5$; [d]for $k = 6$; [e]for $k = 8$; [f]for $k = 10$.

Proof: Under the assumptions, not satisfied by the negative binomial as defined here, Mahamunulu (1967) (Section 7) has presented a similar condition. This result can be derived by following the proof given by the same.

It can be easily checked that, when sampling from negative binomial populations the smallest n exists for which the PCS $\geq P^*$. The effect of r on the number of units required per process is also studied. Sample sizes per process to meet the specification (P^*, δ^*) for $r = 1(1)5$ are listed in Tables 1 – 5. Notice that n

Table 5. Number of Units Required per Process to Meet Specifications (P^*, δ^*), for $r = 5$

$\delta^* \backslash P^*$	0.75	0.80	0.90	0.95	0.98	0.99	0.75	0.80	0.90	0.95	0.98	0.99
	1[a]	1	1	1	2	2	1[d]	1	1	2	2	2
0.55	1[b]	1	1	1	2	2	1[e]	1	1	2	2	2
	1[c]	1	1	1	2	2	1[f]	1	1	2	2	2
	1	1	1	1	2	2	1	1	1	2	2	2
0.50	1	1	1	2	2	2	1	1	2	2	2	3
	1	1	1	2	2	2	1	1	2	2	2	3
	1	1	1	2	2	3	1	1	2	2	2	3
0.45	1	1	1	2	2	3	1	1	2	2	3	3
	1	1	2	2	2	3	1	2	2	2	3	3
	1	1	2	2	2	3	1	2	2	2	3	3
0.40	1	1	2	2	3	3	2	2	2	3	3	4
	1	1	2	2	3	3	2	2	2	3	3	4
	1	1	2	2	3	4	2	2	2	3	4	4
0.35	1	2	2	3	3	4	2	2	3	3	4	5
	1	2	2	3	4	4	2	2	3	3	4	5
	1	2	2	3	4	5	2	2	3	4	5	6
0.30	2	2	3	3	4	5	2	3	3	4	5	6
	2	2	3	4	5	6	2	3	4	5	6	6
	2	2	3	4	6	7	3	3	4	5	7	8
0.25	2	2	3	5	6	7	3	3	5	6	7	9
	2	3	4	5	7	8	3	4	5	6	8	9
	2	3	4	6	8	10	4	4	6	8	11	13
0.20	3	3	5	7	9	11	4	5	7	9	11	13
	3	4	6	8	10	12	5	5	7	9	12	14
	3	4	7	10	15	18	6	7	10	14	18	22
0.15	4	5	9	12	16	20	7	8	12	15	20	23
	5	6	10	13	17	21	8	9	13	16	21	24
	7	9	15	22	32	39	12	15	23	30	41	48
0.10	9	11	19	26	36	43	15	17	25	33	44	52
	11	13	21	29	39	46	16	19	28	36	46	54
	25	33	60	88	126	156	47	57	88	120	161	192
0.05	35	44	72	102	141	172	56	67	99	131	174	205
	42	51	81	112	152	183	63	74	107	140	183	215

[a] for $k = 3$; [b] for $k = 4$; [c] for $k = 5$; [d] for $k = 6$; [e] for $k = 8$; [f] for $k = 10$.

decreases as r is increased keeping k, P^* and δ^* constant. When r is increased from 1 to 2, the sample sizes are reduced by almost fifty percent. The geometric distribution is a particular case of the negative binomial. Therefore this procedure for $r = 1$ can be used for selecting the best geometric population.

Acknowledgements: The authors would like to extend sincere appreciation to J.D. Becker, Applications Analyst, Boeing, from the University of South Alabama, Mobile, for his invaluable assistance in setting up programs on the Alabama Supercomputer to carry out computations of the tables.

References

Bechhofer RE, Kiefer J, Sobel M (1968) Sequential Identification and Ranking Procedures. The Univ of Chicago Press, Chicago

Chambers ML, Jarratt P (1964) Use of Double Sampling for Selecting Best Population. Biometrika 51:49–64

Clark SJ, Perry JN (1989) Estimation of the Negative Binomial Parameter k by Maximum Quasi-Likelihood. Biometrics 45:309–316

Gupta SS, Nagel K (1971) In: Gupta SS, Yackel J (eds) On some Contributions to Multiple Decision Theory. Statistical Decision Theory and Related Topics, pp 79–95

Mahamunulu DM (1967) Some Fixed-Sample Ranking and Selection Problems. Ann Math Statist 38:1079–1091

Nagardeolekar MS (1988) Fixed Sample Selection Procedures and an Approximate Kiefer-Weiss Solution for Negative Binomial Populations. Ph D Dissertation, Okla-State Univ., Stillwater, OK

Shenton LR, Bowman KO (1967) Remarks on Large Sample Estimators for some Discrete Distributions. Technometrics 9:587–598

Sobel M, Huyett MJ (1957) Selecting the Best one of Several Binomial Populations. Bell Syst Tech J 36:537–576

Williamson E, Brentherton MH (1964) Tables of Logarithmic Series Distribution. Ann Math Statist 35:284–297

Willson LJ, Folks JL, Young JH (1984) Multistage Estimation Compared with Fixed-Sample-Size Estimation of the Negative Binomial Parameter k Biometrics 40:109–117

Quenouille MH (1949) A Relation Between the Logarithmic Poisson and Negative Binomial Series. Biometrics 5:162–164

Received 10 January 1992

Book Review

Johnson NL, Kotz S, Wu X: *"Inspection Errors for Attributes in Quality Control"*. Monographs on Statistics and Applied Probability 44. Chapman & Hall, 1991, xi+212 pp., 25.00

In this book the very important problem of the influence of inspection errors on acceptance sampling and on the identification of nonconforming items is investigated. In the Introduction the authors give an illustrative and motivating example: Let $p = 0.99$ be the probability that a nonconforming item is classified as nonconforming and $p' = 0.01$ be the probability that a conforming item is classified as nonconforming. If the proportion of nonconforming items in the population is 0.01 then 50% among all items declared nonconforming are conforming!

The book is divided into three parts. Here, the table of contents:

Although this monograph is based on about twenty papers by the authors they included the work of other authors in this field (see the extensive bibliography). In my opinion the authors found a good compromise between theoretical considerations and applications in the field of quality control. Here the computer programs noted in Appendix A should be mentioned.

As for Part One the influence of inspection errors on acceptance sampling is carefully investigated. I think sampling plans should be constructed in such a way that inspection errors are taken into consideration – and compensated. This might be a field of further research. Part Two of the book is dedicated to the identification of nonconforming items in the presence of inspection errors, where most emphasis is put on group testing with important applications especially in the medical area. In Appendix B the famous 1943 Dorfman paper on group testing is reprinted. In Chapter 11 in Part Three the problem of estimating the inspection error probabilities is investigated and discussed – a very important and challenging field of research.

All in all I recommend this book to everybody (researches, practitioners, students) who is interested in acceptance sampling or group testing.

München *B. Arnold*

Metrika (1993) 40:37–50

Metrika

A New Interpretation of Optimality for *E*-Optimal Designs in Linear Regression Models

HOLGER DETTE

Institut für Mathematische Stochastik, Universität Göttingen, Lotzestr. 13, 3400 Göttingen, Germany

Abstract: The optimal design problem for the estimation of several linear combinations $c_l'\theta$ ($l = 1, \ldots, m$) is considered in the usual linear regression model $y = f'(x)\theta$ ($f(x) \in \mathbb{R}^k$, $\theta \in \mathbb{R}^k$). An optimal design minimizes a (weighted) p-norm of the variances of the least squares estimates for the different linear combinations $c_l'\theta$. A generalized Elfving theorem is used to derive the relation of the new optimality criterion to the *E*-optimal design problem. It is shown that the *E*-optimal design for the parameter θ minimizes such a (weighted) p-norm whenever the vector $c = (c_1', \ldots, c_k')'$ is an inball vector of a symmetric convex and compact "Elfving set" in $\mathbb{R}^{k^2}$.

1 Introduction

Consider m linear regression models of the form

$$g_l(x) = f_l'(x)\theta_l \quad l = 1, \ldots, m$$

where $f_l'(x) = (f_{l1}(x), \ldots, f_{lk_l}(x))$ is the vector of regression functions ($l = 1, \ldots, m$). The control variable x takes values in a compact space $\mathcal{X}$ with sigma field $\mathcal{B}$ including all one point sets. $\theta_l = (\theta_{l1}, \ldots, \theta_{lk_l})'$ is the vector of unknown parameters and the functions $f_{l_1}(x), \ldots, f_{lk_l}(x)$ are assumed to be real valued and continuous functions on the design space $\mathcal{X}$. The different models are collected in the set

$$\mathcal{F}_m = \{g_l \mid g_l(x) = f_l'(x)\theta_l,\ l = 1, \ldots, m\}\ . \tag{1.1}$$

For every $x \in \mathcal{X}$ a random variable $Y(x)$ with (unknown) mean $g(x) \in \mathcal{F}_m$ and variance $\sigma^2 > 0$ can be observed where different observations are assumed to be

0026–1335/93/1/37–50 $2.50

independent. A design is a probability measure on $\mathcal{X}$ and the information matrix of ξ in the model g_l is defined by

$$M_l(\xi) = \int_{\mathcal{X}} f_l(x) f_l'(x) d\xi(x) \qquad l = 1, \ldots, m \ .$$

Throughout this paper we assume that there exists a finite support design ξ on $\mathcal{X}$ such that all information matrices $M_l(\xi)$ are regular. If ξ is supported at s points $x_1, \ldots, x_s$ with masses $\xi(x_i) = \frac{n_i}{n}$ the experimenter takes n_i *uncorrelated observations* at each point x_i $\left(i = 1, \ldots, s, \ \sum_{i=1}^{s} n_i = n \right)$ and the covariance matrix of the least squares estimator for θ_l (in the model $g_l(x)$) is proportional to the inverse of the information matrix $M_l(\xi)$ (see e.g. Krafft (1978)). This paper deals with the E-optimality and the c_l-optimality criterion for the model g_l. An E-optimal design (for the model g_l) maximizes the minimum eigenvalue $\lambda_{\min}(M_l(\xi))$ of the information matrix $M_l(\xi)$ while a c_l-optimal design minimizes $c_l' M_l^-(\xi) c_l$ for a given vector $c_l \in \mathbb{R}^{k_l}$. Here $M_l^-(\xi)$ denotes any generalized inverse of $M_l(\xi)$ and we assume that the vector c_l is estimable by the design ξ, i.e. $c_l \in range(M_l(\xi))$ (see Pukelsheim (1980)). The c_l-optimality is a special case of the following more general definition of optimality. Consider a number $p \in [-\infty, 1]$, m vectors $c_l \in \mathbb{R}^{k_l}$ and numbers $\beta_l > 0$. The vectors c_l are collected in the "grand" vector $(c_1', \ldots, c_m')' \in \mathbb{R}^k$ $\left(k = \sum_{l=1}^{m} k_l \right)$ and the vector $\beta = (\beta_1, \ldots, \beta_m)$ is called a "prior" for the class $\mathcal{F}_m$ in (1.1) where we assume that the components of β sum to one. For $p \in [-\infty, 1] \backslash \{-\infty, 0\}$ define

$$\Phi_{p,\beta}^{c}(\xi) = \left[\sum_{l=1}^{m} \beta_l (c_l' M_l^-(\xi) c_l)^{-p} \right]^{1/p}$$

and let for $p = 0$ and $p = -\infty$

$$\Phi_{0,\beta}^{c}(\xi) = \prod_{l=1}^{m} (c_l' M_l^-(\xi) c_l)^{-\beta_l} \quad \text{and} \quad \Phi_{-\infty,\beta}^{c}(\xi) = \min_{l=1}^{m} (c_l' M_l^-(\xi) c_l)^{-1}$$

respectively. It is easy to see that $\Phi_{p,\beta}^{c}(\xi) = m^{1/p} \Phi_{p,\beta^*}^{\tilde{c}}(\xi)$ where

$$\beta^* = \left(\frac{1}{m}, \ldots, \frac{1}{m} \right) \quad \text{and} \quad \tilde{c} = (\beta_1^{-1/2p} c_1', \ldots, \beta_m^{-1/2p} c_m')' \quad (p \neq 0, -\infty) \ .$$

Definition 1.1: A design ξ for which all linear combinations $c_l'\theta_l$ are estimable is called $\Phi_{p,\beta}^c$-optimal for the class $\mathcal{F}_m$ with respect to the prior β if ξ maximizes the function $\Phi_{p,\beta}^c(\xi)$. If all models in the set $\mathcal{F}_m$ in (1.1) are identical (i.e. $f_1(x) = f_l(x)$, $l = 1, \ldots, m$) the maximizing design is called $\Phi_{p,\beta}^c$-optimal with respect to the prior β.

Note that if $f_l = f_1$ $(l = 1, \ldots, m)$ and $p = -1$ the $\Phi_{p,\beta}^c$-optimality criterion gives the $A'\theta_1$-optimality (defined as in Karlin and Studden (1966)) when putting $A = A_\beta = (\sqrt{\beta_1}c_1, \ldots, \sqrt{\beta_m}c_m) \in \mathbb{R}^{k_1 \times m}$ (see also Studden (1971) for some properties of optimal designs with respect to this criterion).

The Definition 1.1 is motivated by the following point of critism on the "classical" optimal design theory. This theory is based on the fact that the underlying model (namely g_l) is known by the experimenter which is not very realistic from a practical point of view. In scientific experiments it is often the case that a suitable (linear) model for the description of the dependency of Y from x is not known before experiments have been conducted. However, in some cases the experimenter does not know the "true" model exactly but he knows that the model belongs to the given class $\mathcal{F}_m$ defined by (1.1) (e.g. the experimenter is sure that a linear or quadratic polynomial is appropriate). Assume that in every model $g_l \in \mathcal{F}_m$ the linear combinations $c_l'\theta_l$ $(l = 1, \ldots, m)$ (where $c_1, \ldots, c_m$ are given vectors $c_l \in \mathbb{R}^{k_l}$) or different linear combinations in the same model have to be estimated (in this case all vectors f_l would be the same). Because the variances of the least squares estimates for the linear combinations $c_l'\theta_l$ are proportional to $c_l' M_l^- (\xi) c_l$ $(l = 1, \ldots, m)$ an optimal design with respect to one of the criteria of Definition 1.1 allows "precise" estimates of all linear combinations in the given models of $\mathcal{F}_m$. The numbers β_l are used to reflect the experimenters belief about the adequacy of the models g_l or the importance of the linear combinations $c_l'\theta_l$. In this sense the $\Phi_{p,\beta}^c$-optimality criterion can be viewed as a model robust version of the classical c_l-optimality criterion.

Note that the efficiency of a $\Phi_{p,\beta}^c$-optimal design for estimating one linear combination $c_{l_0}'\theta_{l_0}$ in the model g_{l_0} will depend sensitively on the choice of the prior β_{l_0}. Thus a "bad" design for estimating $c_{l_0}'\theta_{l_0}$ in the model g_{l_0} may be judged as a good solution of the $\Phi_{p,\beta}^c$-optimal design problem just by choosing a relative small number β_{l_0}. A statistical justification of a design performing well for the estimation of completely independent parameters in different models can be given under special circumstances. As an example we will consider the following two regression setups. Assume that an estimate of the (unknown) regression function $g(x) \in \mathcal{F}_m$ at a given point $x_0 \notin x$ is desired. In this case we choose $c_l = f_l(x_0)$ $(l = 1, \ldots, m)$ and obtain an extrapolation design robust with respect to the model assumptions. If the experimenter has a very precise knowledge about the underlying model and wants to estimate the regression function at the points $x_1, \ldots, x_m$ we choose $c_l = f_1(x_l)$ $(l = 1, \ldots, m)$ and all models g_l equal g_1. Secondly, consider the case of nested regression models (i.e. $f_l(x) = (f_{11}(x), \ldots, f_{1l}(x))'$, $l = 1, \ldots, m$) and define $c_l = (0, \ldots, 0, 1)' \in \mathbb{R}^l$. In this case an $\Phi_{p,\beta}^c$-optimal design may be useful to decide how many regression functions have to be included into the model.

For the classical c_1-optimality criterion (i.e. $m = 1$ and $\beta_1 = 1$ in Definition 1.1) the theorem of Elfving (1952) is an important tool for the (geometric) characterization of the c_1-optimal design (see also Karlin and Studden (1966) and Pukelsheim (1981) for more details). A c_1-optimal design can be characterized as a design which allows a representation of the intersection of the half line $\{\lambda c_1 \mid \lambda > 0\}$ with the boundary of the Elfving set

$$\mathcal{R}_1 = \text{co}\,(\{f_1(x) \mid x \in \mathcal{X}\} \cup \{-f_1(x) \mid x \in \mathcal{X}\}) \subseteq \mathbb{R}^{k_1} \tag{1.2}$$

where co $(\mathcal{A})$ denotes the convex hull of a set $\mathcal{A} \subseteq \mathbb{R}^{k_1}$. In Section 2 we will present a generalization of Elfving's Theorem for the $\Phi^c_{p,\beta}$-optimal design problem of Definition 1.1. In the same section the relationship between $\Phi^c_{p,\beta}$-optimal designs with respect to different values of $p \in (-\infty, 1]$ is given and a sufficient condition is stated when a design is $\Phi^c_{p,\beta}$-optimal for all $p \in [-\infty, 1]$. Section 3 deals with the E-optimality criterion and its relation to the $\Phi^c_{p,\beta}$-optimal design problem where all functions f_l are identical (i.e. $f_1(x) = f_l(x)$, $l = 1, \ldots, m$). We show that the E-optimal design is $\Phi^c_{p,\beta}$-optimal with respect to a prior β, where $p \in (-\infty, 1]$ is arbitrary, $c = (c'_1, \ldots, c'_{k_1})'$ is any inball vector of the generalized Elfving set and the l-th weight of the prior β is proportional to $(c'_l c_l)^{p+1}$ $(l = 1, \ldots, k_1)$.

2 $\Phi^c_{p,\beta}$-Optimal Designs

In the following we will define a generalized Elfving set $\mathcal{R}^\beta_m$ corresponding to the $\Phi^c_{p,\beta}$-optimal design problem by

$$\mathcal{R}^\beta_m := \text{co}\left(\left\{(\varepsilon_1 f'_1(x), \ldots, \varepsilon_m f'_m(x))' \mid x \in \mathcal{X},\ \varepsilon_l \in \left[-\frac{1}{\sqrt{\beta_l}}, \frac{1}{\sqrt{\beta_l}}\right],\ \sum_{l=1}^m \beta_l \varepsilon_l^2 = 1\right\}\right) \tag{2.1}$$

which is a convex, symmetric and compact subset of $\mathbb{R}^k \left(k = \sum_{l=1}^m k_l\right)$ containing the point 0. Note that in the case $m = 1$ $(\beta_1 = 1)$ (2.1) gives exactly the set defined in (1.2). In general the structure of $\mathcal{R}^\beta_m$ is very complicated and is illustrated by some examples in a paper of Dette (1992). We are now able to state an analogous geometric characterization of the $\Phi^c_{p,\beta}$-optimal design problem for

the class $\mathcal{F}_m$ with respect to a prior β as given by Elfving (1952). The proof of the following theorem is performed by similar arguments as given in Dette (1992) (where the case $p = 0$ is considered) and therefore omitted.

Theorem 2.1: Let $p \in (-\infty, 1]$, *a design* $\xi = \begin{Bmatrix} x_v \\ p_v \end{Bmatrix}_{v=1}^{s}$ *(for which* $c_l'\theta_l$ *is estimable* $l = 1, \ldots, m$*) is* $\Phi_{p,\beta}^c$*-optimal for the class* $\mathcal{F}_m$ *with respect to the prior* β *if and only if there exist positive numbers* $\gamma_1, \ldots, \gamma_m$ *and numbers* $\varepsilon_{11}, \ldots, \varepsilon_{1s}, \varepsilon_{21}, \ldots, \varepsilon_{2s}, \ldots, \varepsilon_{m1}, \ldots, \varepsilon_{ms}$ *such that the following properties* (A), (B), (C) *and* (D) *hold.*

$$\gamma_l c_l = \sum_{v=1}^{s} p_v \varepsilon_{lv} f_l(x_v) \qquad l = 1, \ldots, m \ . \tag{A}$$

The point $(\gamma_1 c_1', \ldots, \gamma_m c_m')'$ *is a boundary point of the set* $\mathcal{R}_m^\beta$ *with a supporting hyperplane* $(\beta_1 d_1', \ldots, \beta_m d_m')'$. (B)

$$\gamma_l^2 \left(\frac{c_l' d_l}{\gamma_l} \right)^{p+1} = \left[\sum_{j=1}^{m} \beta_j \left(\frac{\gamma_j}{c_j' d_j} \right)^p \right]^{-1} \qquad l = 1, \ldots, m \tag{C}$$

$$\sum_{l=1}^{m} \beta_l \varepsilon_{lv}^2 = 1 \qquad v = 1, \ldots, s \ . \tag{D}$$

Remark 2.2: It follows from the proof of Theorem 2.1 that the quantities γ_l, ε_{lv} and d_l are given by

$$\gamma_l^{-2} = (c_l' M_l^- (\xi) c_l)^{p+1} \cdot \sum_{j=1}^{m} \beta_j (c_j' M_j^- (\xi) c_j)^{-p} \qquad (l = 1, \ldots, m) \tag{2.2}$$

$$d_l = \gamma_l G_l c_l \qquad (l = 1, \ldots, m) \tag{2.3}$$

$$\varepsilon_{lv} = f_l'(x_v) d_l \qquad (l = 1, \ldots, m, \ v = 1, \ldots, s)$$

where the matrices G_l are "suitable" generalized inverses of $M_l(\xi)$, $l = 1, \ldots, m$ (see Dette (1992) for more details). Note also that the numbers $\gamma_1, \ldots, \gamma_m$ are uniquely determined which is a consequence of (2.2) and the concavity of the function $\Phi_{p,\beta}^c(\xi)$.

Remark 2.3: The case $p = -1$ plays a particular role in Theorem 2.1 because in this case all quantities γ_l in the general Elfving Theorem coincide, that is

$$\gamma^{-2} = \gamma_l^{-2} = \sum_{j=1}^{m} \beta_j c_j' M_j^- (\xi) c_j \qquad l = 1, \ldots, m \ .$$

Here the $\Phi^c_{-1,\beta}$-optimal design can be determined (as in the classical case $m = 1$) finding a representation (A) of the point where the line $\{\lambda (c_1', \ldots, c_m')' \mid \lambda > 0\}$ intersects the boundary of the Elfving set $\mathscr{R}_m^\beta$. For the choice of identical models we thus get the Elfving theorem for $A'\theta$-optimality given in Studden (1971) where $A = A_\beta = (\sqrt{\beta_1} c_1, \ldots, \sqrt{\beta_m} c_m)$.

Theorem 2.4: Let $p_1 \in (-\infty, 1]$ and ξ denote a $\Phi^c_{p_1,\beta}$-optimal design for the class $\mathscr{F}_m$ with respect to the prior $\beta = (\beta_1, \ldots, \beta_m)$ where $\gamma_1, \ldots, \gamma_m, d_1, \ldots, d_m$ are the quantities of the Elfving Theorem 2.1. For every $p_2 \in (-\infty, 1]$ the design ξ is also $\Phi^c_{p_2,\tilde{\beta}}$-optimal for the class $\mathscr{F}_m$ with respect to the prior $\tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_m)$ where

$$\tilde{\beta}_l = \beta_l \frac{\gamma_l^{p_1 - p_2} (c_l' d_l)^{p_2 - p_1}}{\sum_{l=1}^{m} \beta_l \gamma_l^{p_1 - p_2} (c_l' d_l)^{p_2 - p_1}} \qquad (l = 1, \ldots, m) \ .$$

If $p \in (-\infty, 1]$ and ξ denotes a $\Phi^c_{p,\beta}$-optimal design for the class $\mathscr{F}_m$ with respect to the prior β such that the quantities $c_l' d_l / \gamma_l$ are constant, then the design ξ is $\Phi^c_{q,\beta}$-optimal for the class $\mathscr{F}_m$ with respect to the prior β whenever $q \in (-\infty, 1]$. If, additionally, the matrices $M_1(\xi), \ldots, M_m(\xi)$ are nonsingular the design ξ is also $\Phi^c_{-\infty,\beta}$-optimal for the class $\mathscr{F}_m$.

Proof: Let $k = \sum_{l=1}^{m} k_l$ and

$$\tilde{M}(\xi) = \begin{pmatrix} M_1(\xi) & & \\ & \ddots & \\ & & M_m(\xi) \end{pmatrix} \in \mathbb{R}^{k \times k} \ ,$$

$$\tilde{K} = \begin{pmatrix} \tilde{c}_1 & & \\ & \ddots & \\ & & \tilde{c}_m' \end{pmatrix} \in \mathbb{R}^{m \times k} \ ,$$

(all other entries in the matrices are 0) $\tilde{c}_l = \beta_l^{-1/2p} c_l$ $(l = 1, \ldots, m)$ and $C_{\tilde{K}}(\tilde{M}) = \min\{L\tilde{M}L' \mid L\tilde{K}' = I_m\}$, then minimizing $\Phi^c_{p,\beta}$ over all designs ξ is equivalent to maximizing the (unweighted) p-mean $(\text{trace}\, (C_{\tilde{K}}(\tilde{M})^p)^{1/p}$ in the convex compact set of associated matrices $\tilde{M}(\xi)$. Thus the first part of the theorem follows im-

mediately by computing gradients of $\Phi^c_{p,\beta}$ via standard rules of differentiation (see e.g. Gaffke (1985, 1987)). For the second part we use (2.3) and obtain by an application of the first part ($p_1 = p$, $p_2 = q$) that the design ξ is $\Phi^c_{q,\beta}$-optimal for the class $\mathscr{F}_m$ with respect to the prior β whenever $q \in (-\infty, 1]$. Finally, the case $q = -\infty$ follows considering the limit $q \to -\infty$. ■

We will finish this section showing that *every* design which allows a representation of a boundary point $\gamma c = (\gamma_1 c_1', \ldots, \gamma_m c_m')'$ of the generalized Elfving set $\mathscr{R}^\beta_m$ is already $\Phi^c_{p,\beta}$-optimal for a special prior β.

Theorem 2.5: Let $c = (c_1', \ldots, c_m')'$ *and* $\gamma_c = (\gamma_1 c_1', \ldots, \gamma_m c_m')'$ *denote an arbitrary boundary point of the set* $\mathscr{R}^\beta_m$ *with supporting hyperplane* $(\beta_1 d_1', \ldots \beta_m d_m')'$. *Let*

$$\xi = \begin{Bmatrix} x_\nu \\ p_\nu \end{Bmatrix}_{\nu=1}^{s}$$

denote a design which allows a representation

$$\gamma_l c_l = \sum_{\nu=1}^{s} p_\nu \varepsilon_{l\nu} f_l(x_\nu) \qquad (l = 1, \ldots, m) ,$$

such that for $\nu = 1, \ldots, s$ $\sum_{l=1}^{m} \beta_l \varepsilon_{l\nu}^2 = 1$. *For every* $p \in (-\infty, 1]$ *the design* ξ *is* $\Phi^c_{p,\beta}$*-optimal for the class* $\mathscr{F}_m$ *with respect to the prior* $\tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_m)$ *where*

$$\tilde{\beta}_l = \beta_l \gamma_l^2 \left(\frac{c_l' d_l}{\gamma_l} \right)^{p+1} \left[\sum_{j=1}^{m} \beta_j \gamma_j^2 \left(\frac{c_j' d_j}{\gamma_j} \right)^{p+1} \right]^{-1} \qquad (l = 1, \ldots, m) .$$

Proof: By an application of Theorem 2.1 in the case $p = -1$ we obtain that the design ξ is $\phi^{\tilde{c}}_{-1,\beta}$-optimal for the class $\mathscr{F}_m$ with respect to the prior β where $\tilde{c} = (\gamma_1 c_1', \ldots, \gamma_m c_m')'$ (note that in the case $p = -1$ the condition (C) of Theorem 2.1 is evident by the property (B)). The assertion now follows directly from Theorem 2.4 putting $p_1 = -1$ and $p_2 = p$. ■

3 E-Optimal Designs

In this section we will investigate the relation between the E-optimality criterion and the model robust criterion given in Definition 1.1. In what follows $\lambda_{\min}$ always denote the minimum eigenvalue of the information matrix corresponding

to the E-optimal design and $\lambda_{\min}(A)$ denotes the minimum eigenvalue of a symmetric matrix $A \in \mathbb{R}^{k_1 \times k_1}$. Pukelsheim (1980) showed that ξ is E-optimal iff there exists a matrix

$$E = \sum_{i=1}^{k_0} \alpha_i z_i z_i' \in \mathbb{R}^{k_1 \times k_1} \tag{3.1}$$

such that $f_1'(x) E f_1(x) \leq \lambda_{\min}(M_1(\xi))$ for all $x \in \mathcal{X}$. The vectors z_i in the representation (3.1) are normalized eigenvectors ($||z_i||_2 = 1$) corresponding to the minimum eigenvalue $\lambda_{\min}$ of $M_1(\xi)$, the α_i are positive numbers with sum 1 and $k_0 = \operatorname{rank}(E) \leq k_1$. Throughout this section we will consider the model robust optimality criterion of Definition 1.1 where all models $g_l \in \mathcal{F}_m$ are identical (that is $f_l(x) = f_1(x)$, $l = 1, \ldots, m$). The following result generalizes Theorem 3.2 of Dette and Studden (1992) ($p = -1$).

Theorem 3.1: For every $p \in [-\infty, 1]$ the design ξ is $\Phi^c_{p,\beta}$-optimal with respect to the prior α where $\alpha = (\alpha_1, \ldots, \alpha_{k_0})$ and the vector $c = (z_1', \ldots, z_{k_0}')'$ are defined by (3.1).

For the following investigations of the geometric properties of the E-optimal design we will consider the corresponding Elfving set $\mathcal{R}^{\alpha}_{k_0}$ defined in (2.1) and its "inball radius"

$$r_{k_0} := \min\{||c||_2 \mid c \in \partial \mathcal{R}^{\alpha}_{k_0}\} . \tag{3.2}$$

Because $\mathcal{R}^{\alpha}_{k_0}$ is a convex subset of $\mathbb{R}^{k_0 k_1}$ we obtain an equivalent representation of (3.2) by the distances of covering halfspaces to $\mathcal{R}^{\alpha}_{k_0}$

$$r_{k_0} = \min\left\{ \frac{1}{\sqrt{\sum_{l=1}^{k_0} d_l' d_l}} \,\middle|\, d_l \in \mathbb{R}^{k_1}, \left| \sum_{l=1}^{k_0} \varepsilon_l d_l' f_1(x) \right| \leq 1, \ \forall x \in \mathcal{X}, \ \sum_{l=1}^{k_0} \alpha_l \varepsilon_l^2 = 1 \right\}. \tag{3.3}$$

Every vector $c \in \partial \mathcal{R}^{\alpha}_{k_0}$ with $||c||_2 = r_{k_0}$ is called an "inball vector" of $\mathcal{R}^{\alpha}_{k_0}$.

Theorem 3.2: Let E be defined by (3.1) then the point $\sqrt{\lambda_{\min}}(z_1', \ldots, z_{k_0}')'$ is a boundary point of the set $\mathcal{R}^{\alpha}_{k_0}$ with supporting hyperplane $\frac{1}{\sqrt{\lambda_{\min}}}(\alpha_1 z_1', \ldots, \alpha_{k_0} z_{k_0}')'$. For the inball radius r_{k_0} of $\mathcal{R}^{\alpha}_{k_0}$ we have the estimates

$$\lambda_{\min} \leq r_{k_0}^2 \leq k_0 \lambda_{\min} \ . \tag{3.4}$$

If, additionally, all weights α_l in (3.1) are equal the vector $c = \sqrt{\lambda_{\min}}\,(z_1', \ldots, z_{k_0}')'$ defines an inball vector of $\mathcal{R}_{k_0}^{\alpha^*}\left(\alpha^* = \left(\frac{1}{k_0}, \ldots, \frac{1}{k_0}\right)\right)$ and the inball radius is given by $r_{k_0} = k_0 \lambda_{\min}$.

Proof: Consider the bijective linear mapping

$$\mathrm{vec}_\alpha(A) = \left(\frac{1}{\sqrt{\alpha_1}} a_1', \ldots, \frac{1}{\sqrt{\alpha_{k_0}}} a_{k_0}'\right)', \quad A = (a_1, \ldots, a_{k_0}) \in \mathbb{R}^{k_1 \times k_0}$$

from $\mathbb{R}^{k_1 \times k_0}$ onto $\mathbb{R}^{k_1 k_0}$. It is easy to see that

$$\mathrm{vec}_\alpha(S_{k_0}) = \mathcal{R}_{k_0}^{\alpha}, \quad \mathrm{vec}_\alpha(\partial S_{k_0}) = \partial \mathcal{R}_{k_0}^{\alpha} \tag{3.5}$$

where $S_{k_0} := \mathrm{co}\{f_1(x)\varepsilon' \mid x \in \mathcal{X},\ \varepsilon \in \mathbb{R}^{k_0},\ \|\varepsilon\|_2 = 1\} \subseteq \mathbb{R}^{k_1 \times k_0}$ denotes the Elfving set investigated by Studden (1971). By straight forward calculations it follows that $A = (\sqrt{\alpha_1} a_1, \ldots, \sqrt{\alpha_{k_0}} a_{k_0})$ is a covering halfspace to S_{k_0} if and only if $a = (\alpha_1 a_1', \ldots, \alpha_{k_0} a_{k_0}')'$ is a covering halfspace to $\mathcal{R}_{k_0}^{\alpha}$. Dette and Studden (1992) showed that the squared inball radius of S_{k_0} is given by $\lambda_{\min}$. Thus we obtain from

$$\lambda_{\min} \leq \frac{1}{\mathrm{trace}\,(A'A)} = \frac{1}{\sum_{l=1}^{k_0} \alpha_l a_l' a_l} \leq \frac{1}{a'a}$$

and from (3.3) the inequality $\lambda_{\min} \leq r_{k_0}^2$. It also follows from Theorem 3.3 of Dette and Studden (1992) that the point $Z = \sqrt{\lambda_{\min}}\,(\sqrt{\alpha_1} z_1, \ldots, \sqrt{\alpha_{k_0}} z_{k_0})$ defines an inball vector of the set S_{k_0} (here z_l, α_l and k_0 are the quantities of the representation of the matrix E in (3.1)). By (3.5) $\mathrm{vec}_\alpha(Z)$ is a boundary point of the set $\mathcal{R}_{k_0}^{\alpha}$ with norm $\|\mathrm{vec}_\alpha(Z)\|_2^2 = k_0 \lambda_{\min}$ which proves the right inequality of (3.4) (note that $\mathrm{vec}_\alpha(Z)$ is not necessarily an inball vector of $\mathcal{R}_{k_0}^{\alpha}$ because the supporting hyperplane $d = \frac{1}{\sqrt{\lambda_{\min}}}\,(\alpha_1 z_1', \ldots, \alpha_{k_0} z_{k_0}')'$ to $R_{k_0}^{\alpha}$ at the point $\mathrm{vec}_\alpha(Z)$ has in general a different direction as the vector $\mathrm{vec}_\alpha(Z)$).

In the case of equal $\alpha_l = \frac{1}{k_0}$ $(l = 1, \ldots, k_0)$ we have $||\text{vec}_\alpha(A)||_2^2 = k_0 ||A||_2^2$ and obtain from (3.5) the second assertion of the theorem. ■

The results stated so far are more of theoretical interest for the determination of the E-optimal design because they assume the knowledge of $k_0, \alpha_1, \ldots, \alpha_{k_0}$ in the representation (3.1) which is usually not available without knowing the E-optimal design. Nevertheless they provide a first insight into the geometric structure of the problem and also indicate the usefulness of a representation with equal weights $\alpha_l = \frac{1}{k_0}$ $(l = 1, \ldots, k_0)$. The following Lemma shows that there always exists such a representation.

Lemma 3.3: For every matrix E in (3.1) there always exists a representation with identical weights α_l.

Proof: Considering a spectral decomposition of the matrix E in (3.1) we obtain a representation with orthonormal vectors z_j $(k_0 = \text{rank}(E) \leq k_1)$. Let $k^* = 2^{k_0 - 1}$ and

$$F = \begin{pmatrix} 1 & 1 & \ldots & 1 & 1 \\ 1 & 1 & \ldots & 1 & -1 \\ 1 & 1 & \ldots & -1 & 1 \\ 1 & 1 & \ldots & -1 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & \ldots & 1 & 1 \\ 1 & -1 & \ldots & 1 & -1 \\ 1 & -1 & \ldots & -1 & 1 \\ 1 & -1 & \ldots & -1 & -1 \end{pmatrix} \in \mathbb{R}^{k^* \times k_0}$$

where all $2^{k_0 - 1}$ components of the first column of F are 1, the first $2^{k_0 - 2}$ components of the second column are 1 while the second $2^{k_0 - 2}$ components of this column are -1 and so on.... Defining new vectors $\tilde{z}_j \in \mathbb{R}^{k_1}$ by

$$(\tilde{z}_1, \ldots, \tilde{z}_{k^*}) = (\sqrt{\alpha_1} z_1, \ldots, \sqrt{\alpha_{k_0}} z_{k_0}) F'$$

(e.g. $\tilde{z}_1 = \sum_{i=1}^{k_0} \sqrt{\alpha_i} z_i$, $\tilde{z}_{k^*} = \sqrt{\alpha_1} z_1 - \sum_{i=2}^{k_0} \sqrt{\alpha_i} z_i$) we have

$$\frac{1}{k^*}\sum_{j=1}^{k^*}\tilde{z}_j\tilde{z}_j' = \frac{1}{k^*}(\tilde{z}_1,\ldots,\tilde{z}_{k^*})\begin{pmatrix}\tilde{z}'\\ \vdots\\ \tilde{z}'_{k^*}\end{pmatrix}$$

$$= \frac{1}{k^*}(\sqrt{\alpha_1}z_1,\ldots,\sqrt{\alpha_{k_0}}z_{k_0})F'F\begin{pmatrix}\sqrt{\alpha_1}z_1'\\ \vdots\\ \sqrt{\alpha_{k_0}}z'_{k_0}\end{pmatrix}$$

$$= (\sqrt{\alpha_1}z_1,\ldots,\sqrt{\alpha_{k_0}}z_{k_0})\begin{pmatrix}\sqrt{\alpha_1}z_1'\\ \vdots\\ \sqrt{\alpha_{k_0}}z'_{k_0}\end{pmatrix} = \sum_{i=1}^{k_0}\alpha_i z_i z_i' = E\ .$$

The vectors $\tilde{z}_j$ are eigenvectors of the information matrix of the E-optimal design corresponding to the minimum eigenvalue $\lambda_{\min}$. From the orthonormality of the vectors z_j we obtain $||\tilde{z}_j||_2 = 1$ which proves the assertion of Lemma 3.3. ■

In view of Theorem 3.2 and Lemma 3.3 it now makes sense to investigate the inball radii of the sequence of Elfving sets $\left(\alpha^* = \left(\frac{1}{m},\ldots,\frac{1}{m}\right)\right)$

$$\mathcal{R}_m^{\alpha^*} = \operatorname{co}\left(\left\{(\varepsilon_1 f_1'(x),\ldots,\varepsilon_m f_1'(x))' \mid x\in\mathcal{X},\ \varepsilon_l\in[-\sqrt{m},\sqrt{m}],\ \sum_{l=1}^{m}\varepsilon_l^2 = m\right\}\right)$$

$(m = 1,2,\ldots)$ for the determination of the minimum eigenvalue of the E-optimal information matrix. By the results stated so far we have equality $r_{k^*}^2 = k^*\lambda_{\min}$ for some $k^* \le 2^{k_1-1}$ (note that $k_0 = \operatorname{rank}(E) \le k_1$). The following theorem shows that this identity holds already for the set $\mathcal{R}_{k_1}^{\alpha^*}$.

Theorem 3.4: Let r_m denote the inball radius of $\mathcal{R}_m^{\alpha^*}$, then we have $r_m^2 = m\lambda_{\min}$ for all $m \ge k_0$. Whenever $m \ge k_0$ and $c = (c_1',\ldots,c_m')'$ is an inball vector of the set $\mathcal{R}_m^{\alpha^*}$ the E-optimal design is $\Phi_{p,\beta}^c$-optimal with respect to the prior $\beta = (\beta_1,\ldots,\beta_m)$ where

$$\beta_l = \frac{(c_l'c_l)^{p+1}}{\sum_{l=1}^{m}(c_l'c_l)^{p+1}} \qquad l = 1,\ldots,m\ . \tag{3.6}$$

Proof: Observing the reasoning in the proof of Theorem 3.2 we see that the Elfving set S_m is a scaled version of the Elfving set $\mathcal{R}_m^{\alpha^*}$ and obtain for the corresponding inball radii

$$s_m = \frac{1}{\sqrt{m}} r_m \quad \text{for all} \quad m \in \mathbb{N} .$$

By Theorem 3.3 of Dette and Studden (1992) we have for $m \geq k_0$ $r_m^2 = m\lambda_{\min}$ which proves the first part of the assertion. For the second part let $m \geq k_0$ and $c = (c_1', \ldots, c_m')'$ denote an inball vector of the set $\mathcal{R}_m^{\alpha *}$ (i.e. $c \in \partial \mathcal{R}_m^{\alpha *}$ and $||c||_2^2 = m\lambda_{\min}$). Using (3.5) it follows that $\tilde{C} = \text{vec}_{\alpha *}^{-1}(c) = \frac{1}{\sqrt{m}}(c_1, \ldots, c_m) \in \mathbb{R}^{k_1 \times m}$ is a boundary point of S_m with norm $||\tilde{C}||_2^2 = \text{trace}\,(\tilde{C}'\tilde{C}) = \frac{1}{m}||c||_2^2 = \lambda_{\min}$. Thus the matrix $\tilde{C}$ is also an inball vector of S_m and we obtain from Theorem 3.4 of Dette and Studden (1992) that the E-optimal design ξ is an optimal design for $\tilde{C}'\theta_1$. This implies that the E-optimal design is also $\Phi_{-1,\beta}^c$-optimal with respect to the prior $\beta^* = \left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$ and the quantities of the Elfving Theorem 2.1 are given by $\gamma_l = 1$ and $d_l = \frac{c_l}{\lambda_{\min}}$ $(l = 1, \ldots, m)$. By an application of Theorem 2.4 it now follows that for every $p \in (-\infty, 1]$ the E-optimal design is also $\Phi_{p,\beta}^c$-optimal with respect to the prior $\beta = (\beta_1, \ldots, \beta_m)$ given in (3.6). ■

Using the above results the E-optimal design can now be determined in the following way. In a first step we have to find an inball vector c and its supporting hyperplane of the set $\mathcal{R}_{k_1}^{\alpha *}$ (note that the number $k_0 = \text{rank}\,(E)$ in the Theorem 3.4 is unkown but always $\leq k_1$). This problem is equivalent to the calculation of the inball vector and its supporting hyperplane of the set S_{k_1} (which is a scaled version of $\mathcal{R}_{k_1}^{\alpha *}$). In a second step we have to find the $\Phi_{p,\beta}^c$-optimal designs with respect to the prior β in (3.6) using the results on $\Phi_{p,\beta}^c$-optimality. By Theorem 3.4 the E-optimal design has to be among these designs. In general the computational difficulties in this part will depend sensitively on the choice of the parameter p in the optimality criterion. Usually the cases $p = 0$ (for $p = 0$ the criterion is invariant with respect to different scalings of the vectors c_l) and $p = -1$ (for $p = -1$ all numbers γ_l in Theorem 2.1 are equal, see Remark 2.3) can be treated easier compared to other choices of p. A couple of examples for this procedure in the case $p = -1$ can be found in Dette and Studden (1992) and we will present here a simple example where the choice $p = 0$ might be useful.

Example 3.5: Consider the weighted polynomial regression model $f(x) = \sqrt{b^2 - x^2} \times (1, x)'$ on the interval $[-b, b]$ where $b > \sqrt{2}$. Observing (3.3) it is straight forward to show that an inball vector of $\mathcal{R}_2^{\alpha *}$ is given by $c = (c_1', c_2')' = \sqrt{2}\,(0, 1, \sqrt{b^2 - 2}, 0)'$. From standard arguments of optimal design theory it follows that

there exists a symmetric E-optimal design. Expressing the information matrix of such a symmetric design in terms of canonical moments of ξ (see Studden (1980), p. 1336) we obtain ($q_2 = 1-p_2$, $q_4 = 1-p_4$)

$$M_2(\xi) = \begin{pmatrix} b^2 q_2 & 0 \\ 0 & b^4 q_2 p_2 q_4 \end{pmatrix}.$$

Therefore a $\Phi^c_{0,\beta}$-optimal design with respect to the prior $\beta = (1/(b^2-1), (b^2-2)/(b^2-1))$ maximizes the product $q_2(p_2q_4)^{1/(b^2-1)}$ and we obtain for the canonical moments of the maximizing measure $p_2 = 1/b^2$, $p_4 = 0$. Using the results of Studden (1982) (transformed to the interval $[-b, b]$) we obtain that the unique $\Phi^c_{0,\beta}$-optimal design puts equal masses at the points -1 and 1 (independent of the length of the interval $[-b, b]$). By Theorem 3.4 this design must be E-optimal too. ■

If the E-optimal design and the matrix E in (3.1) are known it follows from Theorem 3.1 that the E-optimal design also maximizes every (weighted) p-mean of the inverse variances of the least squares estimates for the linear combinations $z_l'\theta_1$ ($l = 1, \ldots, k_0$) and we obtain a new interpretation of optimality for the E-optimal design. It is remarkable that by Theorem 3.4 the E-optimal design is not necessarily $\Phi^c_{p,\beta}$-optimal with respect to the uniform prior $\alpha^* = \left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$ (although c is an inball vector of the set $\mathcal{R}^{\alpha^*}_m$ ($m \geq k_0$)). The weights of the prior in the $\Phi^c_{p,\beta}$-optimality criterion will usually depend on the length of the components c_l of the inball vector $c = (c_1', \ldots, c_m')'$ and these are not necessarily all equal (see Example 3.5).

Acknowledgements: This paper is based on a part of the author's Habilitationsschrift written at the Universität Göttingen. The author would like to thank an anonymous referee for his careful reading of the manuscript and his helpful comments.

References

Dette H (1992) Elfving's Theorem for D-Optimality. Ann Statist, to appear (Technical Report # 16-91, Department of Statistics, Purdue University)

Dette H, Studden WJ (1992) Geometry of E-Optimality. Ann Statist, to appear (Technical Report # 31-91, Department of Statistics, Purdue University)

Elfving G (1952) Optimum Allocation in Linear Regression. Ann Math Statist 23:255–262

Gaffke N (1985) Directional Derivatives of Optimality Criteria at Singular Matrices in Convex Design Theory. Statistics 16:373–388

Gaffke N (1987) Further Characterizations of Design Admissibility for Partial Parameter Estimation in Linear Regression Models. Ann Statist 15:942–957

Karlin S, Studden WJ (1966) Optimal Experimental Designs. Ann Math Statist 37:783–815

Krafft O (1978) Lineare statistische Modelle und optimale Versuchspläne. Vandenhoeck & Ruprecht, Göttingen

Pukelsheim F (1980) On Linear Regression Designs which Maximize Information. J Statist Planning and Inference 4:339–364

Pukelsheim F (1981) On c-Optimal Design Measures. Math Operationsforsch. Statist Ser Statist 12:13–20

Studden WJ (1971) Elfving's Theorem and Optimal Designs for Quadratic Loss. Ann Math Statist 42:1613–1621

Studden WJ (1980) D_s-Optimal Designs for Polynomial Regression Using Continued Fractions. Ann Statist 8:1132–1141

Studden WJ (1982) Optimal Desings for Weighted Polynomial Regression Using Canonical Moments. In: Statistical Decision Theory and Related Topics III. Academic Press, New York, pp 335–350

Received 16 December 1991
Revised version 18 May 1992

Metrika (1993) 40:51–61

Metrika

Uniformly Most Accurate Equivariant Prediction Limit

YOSHIKAZU TAKADA

Department of Mathematics, Faculty of Science, Kumamoto University, Kumamoto 860, Japan

Summary: Prediction limits are widely used for reliability problems and other related problems. The determination of prediction limits has been extensively investigated, but few optimal properties of these limits have been explored. This paper introduces a concept of uniform accuracy in order to compare equivariant prediction limits and show that the prediction limits used for the normal distribution and the exponential distribution are uniformly most accurate equivariant.

Key words: Prediction limit, location-scale family, equivariance, uniform accuracy, median unbiasedness, *s*-out-of-*m* system.

1 Introduction

Suppose that the observed variables $X = (X_1, \ldots, X_r)$ and the predicted variable Y have a joint distribution, which depends on an unknown parameter θ. We consider the problem of obtaining a lower or upper prediction limit for Y. The discussion of lower and upper prediction limits is completely parallel, and therefore it is enough to consider the case of a lower prediction limit.

Let $\delta = \delta(X)$ be a statistic such that

$$P_\theta\{Y \geq \delta(X)\} \geq 1 - \alpha \tag{1.1}$$

for all θ, then it is called a lower prediction limit for Y at confidence level $1-\alpha$.

For certain parametric classes of distributions such as the normal distribution and the exponential distribution, several authors have obtained prediction limits. Good reviews are provided by Hahn and Nelson (1973) and Patel (1989). But few

0026–1335/93/1/51–61 $2.50

theoretical justifications for the limits have been considered. See Chapter 5 of Takeuchi (1975) and Takada (1985). In order to compare prediction limits, this paper introduces the concept of a uniformly most accurate (UMA) equivariant prediction limit which is analogous to a UMA confidence limit (see Chapter 3.5 of Lehmann, 1986), and provides the conditions for a prediction limit to be UMA equivariant for the location-scale family.

Suppose that X and Y have a joint density given by

$$\sigma^{-(r+1)} f\{(x_1-\mu)/\sigma, \ldots, (x_r-\mu)/\sigma, (y-\mu)/\sigma\} \tag{1.2}$$

for some known function f, where $\theta=(\mu,\sigma)$, with $\sigma>0$, is unknown. Then it seems reasonable to confine our attention to (location-scale) equivariant prediction limits, that is, δ for which

$$\delta(aX_1+b, \ldots, aX_r+b) = a\delta(X)+b \tag{1.3}$$

for all $a>0$ and all b. The equivariance requires that the prediction limit remains unchanged under the location-scale transformation. A prediction limit $\tilde{\delta}$ is said to be UMA equivariant at confidence level $1-\alpha$ if within the class of all equivariant prediction limits δ subject to (1.1) it minimizes

$$P_\theta\{Y-\gamma \geq \delta(X)\}$$

for all $\gamma>0$ and all θ. That is, for every preassigned margin $\gamma>0$, $\tilde{\delta}(X)$ (uniformly) minimizes the probability that the predicted variable Y exceeds the lower prediction limit $\delta(X)$ by this margin γ.

In Section 2 the condition for a prediction limit to be UMA equivariant is given. In Section 3 it is shown that the prediction limits used for the normal family and the exponential family are UMA equivariant.

2 UMA Equivariant Prediction Limit

Let $\Delta=\gamma/\sigma$. Then it follows from (1.2) and (1.3) that for an equivariant prediction limit δ

$$P_\theta\{Y-\gamma \geq \delta(X)\} = \int \ldots \int \left\{ \int_{\delta(x)+\Delta}^{\infty} f(x_1, \ldots, x_r, y)\, dy \right\} dx_1 \ldots dx_r \,. \tag{2.1}$$

In order to obtain the condition for a prediction limit to be UMA equivariant, we need the following lemma.

Lemma: For an equivariant prediction limit δ,

$$P_\theta\{Y-\gamma \geq \delta(X)\} = \int \cdots \int \left\{ \int_{\delta(z_1,\ldots,z_{r-2},1,0)}^{\infty} F_\Delta(z_1,z_2,\ldots,z_{r-2},y)\,dy + \int_{\delta(z_1,\ldots,z_{r-2},-1,0)}^{\infty} G_\Delta(z_1,z_2,\ldots,z_{r-2},y)\,dy \right\} dz_1 \ldots dz_{r-2} ,$$

where

$$F_\Delta(z_1,\ldots,z_{r-2},y) = \int_{-\infty}^{\infty} \left\{ \int_0^{\infty} t^{r-1} f(tz_1+u,\ldots,tz_{r-2}+u,t+u,u,ty+u+\Delta)\,dt \right\} du$$

and

$$G_\Delta(z_1,\ldots,z_{r-2},y) = \int_{-\infty}^{\infty} \left\{ \int_0^{\infty} t^{r-1} f(tz_1+u,\ldots,tz_{r-2}+u,-t+u,u,ty+u+\Delta)\,dt \right\} du .$$

Proof: Changing variables from $(x_1,\ldots,x_r)$ to $(u_1,\ldots,u_{r-1},u)$ with $u_i = x_i - x_r$ $(i = 1,\ldots,r-1)$ and $u = x_r$, it follows from (1.3) and (2.1) that

$$P_\theta\{Y-\gamma \geq \delta(X)\} = \int \cdots \int \left\{ \int_{\delta(u_1,\ldots,u_{r-1},0)}^{\infty} f(u_1+u,\ldots,u_{r-1}+u,u,y+u+\Delta)\,dy \right\} \times du_1 \ldots du_{r-1}\,du$$

$$= \int \cdots \int \left\{ \int_0^{\infty} \left\{ \int_{\delta(u_1,\ldots,u_{r-1},0)}^{\infty} f(u_1+u,\ldots,u_{r-1}+u,u,y+u+\Delta)\,dy \right\} du_{r-1} \right\}$$

$$\times du_1 \ldots du_{r-2} du$$

$$+\int \ldots \int \left\{ \int_{-\infty}^{0} \left\{ \int_{\delta(u_1,\ldots,u_{r-1},0)}^{\infty} f(u_1+u,\ldots,u_{r-1}+u,u,y+u+\Delta)\,dy \right\} du_{r-1} \right\}$$

$$\times du_1 \ldots du_{r-2} du = \mathrm{I}+\mathrm{II} \quad \text{(say)} \ . \tag{2.2}$$

Make the change of variables

$$z_i = u_i/u_{r-1} \quad (i = 1,\ldots,r-2) \ , \quad t = u_{r-1} \ .$$

Then

$$\mathrm{I} = \int \ldots \int$$

$$\times \left\{ \int_{0}^{\infty} t^{r-2} \left\{ \int_{t\delta(z_1,\ldots,z_{r-2},1,0)}^{\infty} f(tz_1+u,\ldots,tz_{r-2}+u,t+u,u,y+u+\Delta)\,dy \right\} dt \right\}$$

$$\times dz_1 \ldots dz_{r-2} du$$

$$= \int \ldots \int \left\{ \int_{\delta(z_1,\ldots,z_{r-2},1,0)}^{\infty} F_\Delta(z_1,\ldots,z_{r-2},y)\,dy \right\} dz_1 \ldots dz_{r-2} \ . \tag{2.3}$$

Similarly,

$$\mathrm{II} = \int \ldots \int \left\{ \int_{\delta(z_1,\ldots,z_{r-2},-1,0)}^{\infty} G_\Delta(z_1,\ldots,z_{r-2},y)\,dy \right\} dz_1 \ldots dz_{r-2} \ . \tag{2.4}$$

Substituting (2.3) and (2.4) into (2.2), the proof is completed.

Theorem 1: Suppose that for any $\Delta > 0$

$$H_\Delta(x_1,\ldots,x_r,y)$$

$$= \frac{\int_{-\infty}^{\infty} \int_{0}^{\infty} \sigma^{-(r+2)} f\{(x_1-\mu)/\sigma,\ldots,(x_r-\mu)/\sigma,(y-\mu)/\sigma+\Delta\}\,d\mu\,d\sigma}{\int_{-\infty}^{\infty} \int_{0}^{\infty} \sigma^{-(r+2)} f\{(x_1-\mu)/\sigma,\ldots,(x_r-\mu)/\sigma,(y-\mu)/\sigma\}\,d\mu\,d\sigma}$$

is non-increasing in y for any fixed $x_1, \ldots, x_r$, and that an equivariant prediction limit δ satisfies

$$H_\Delta(x_1, \ldots, x_r, \delta(x_1, \ldots, x_r)) = \lambda$$

for all $x_1, \ldots, x_r$, where λ is a positive constant which does not depend on $x_1, \ldots, x_r$ and $P_\theta\{Y \geq \delta(X)\} = 1-\alpha$. Then δ is a UMA equivariant lower prediction limit at confidence level $1-\alpha$.

Proof: Let δ' be any other equivariant prediction limit at confidence level $1-\alpha$. Then it follows from Lemma that

$$\begin{aligned} &P_\theta\{Y-\gamma \geq \delta'(X)\} - P_\theta\{Y-\gamma \geq \delta(X)\} \\ &= \int \ldots \int \left\{ \int_{\delta'(z_1, \ldots, z_{r-2}, 1, 0)}^{\delta(z_1, \ldots, z_{r-2}, 1, 0)} F_\Delta(z_1, \ldots, z_{r-2}, y)\, dy \right. \\ &\qquad \left. + \int_{\delta'(z_1, \ldots, z_{r-2}, -1, 0)}^{\delta(z_1, \ldots, z_{r-2}, -1, 0)} G_\Delta(z_1, \ldots, z_{r-2}, y)\, dy \right\} dz_1 \ldots dz_{r-2} \,. \end{aligned} \tag{2.5}$$

It is easy to see that

$$F_\Delta(z_1, \ldots, z_{r-2}, y)/F_0(z_1, \ldots, z_{r-2}, y) = H_\Delta(z_1, \ldots, z_{r-2}, 1, 0, y)$$

and

$$G_\Delta(z_1, \ldots, z_{r-2}, y)/G_0(z_1, \ldots, z_{r-2}, y) = H_\Delta(z_1, \ldots, r_{r-2}, -1, 0, y) \,,$$

where F_0 and G_0 are F_Δ and G_Δ with $\Delta = 0$. Hence it follows from the conditions of the theorem that the right hand side of (2.5) is bounded below by

$$\begin{aligned} &\lambda \int \ldots \int \left\{ \int_{\delta'(z_1, \ldots, z_{r-2}, 1, 0)}^{\delta(z_1, \ldots, z_{r-2}, 1, 0)} F_0(z_1, \ldots, z_{r-2}, y)\, dy \right. \\ &\qquad \left. + \int_{\delta'(z_1, \ldots, z_{r-2}, -1, 0)}^{\delta(z_1, \ldots, z_{r-2}, -1, 0)} G_0(z_1, \ldots, z_{r-2}, y)\, dy \right\} dz_1 \ldots dz_{r-2} \\ &= \lambda\{P_\theta(Y \geq \delta'(X)) - P_\theta(Y \geq \delta(X))\} \geq 0 \,, \end{aligned}$$

which completes the proof.

Similarly for the UMA equivariant upper prediction limit, the following result holds.

Theorem 2: Suppose that for any $\Delta<0$, $H_\Delta(x_1,\ldots,x_r,y)$ is non-decreasing in y for any fixed $x_1,\ldots,x_r$, and that an equivariant prediction limit δ satisfies

$$H_\Delta(x_1,\ldots,x_r,\delta(x_1,\ldots,x_r))=\lambda$$

for all $x_1,\ldots,x_r$, where λ is a positive constant which does not depend on $x_1,\ldots,x_r$ and $P_\theta\{Y\le\delta(X)\}=1-\alpha$. Then δ is a UMA equivariant upper prediction limit at confidence level $1-\alpha$.

The case $\alpha=1/2$ is of particular interest. If δ satisfies

$$P_\theta\{Y\le\delta(X)\}=P_\theta\{Y\ge\delta(X)\}=1/2$$

for all θ, then it is said to be a median unbiased predictor for Y. See Takada (1991). Since for $\gamma_1>0$ and $\gamma_2>0$

$$P_\theta\{-\gamma_1<Y-\delta(X)<\gamma_2\}=1-P_\theta\{Y+\gamma_1\le\delta(X)\}-P_\theta\{Y-\gamma_2\ge\delta(X)\}\ ,$$

the following result (stated in Lehmann 1986, page 95) is an immediate consequence of Theorems 1 and 2 for the particular case $\alpha=1/2$.

Theorem 3: Suppose that for any fixed $x_1,\ldots,x_r$, $H_\Delta(x_1,\ldots,x_r,y)$ is non-increasing in y for $\Delta>0$ and non-decreasing in y for $\Delta<0$, and that a median unbiased equivariant predictor $\tilde{\delta}$ satisfies

$$H_\Delta(x_1,\ldots,x_r,\tilde{\delta}(x_1,\ldots,x_r))=\lambda$$

for all $x_1,\ldots,x_r$, where λ is a positive constant which does not depend on $x_1,\ldots,x_r$. Then $\tilde{\delta}$ maximizes the probability

$$P_\theta\{-\gamma_1<Y-\delta(X)<\gamma_2\}$$

for all $\gamma_1>0$, $\gamma_2>0$ and all θ within the class of all median unbiased equivariant predictors $\delta(X)$.

In other words, for any preassigned margins $\gamma_1 > 0$, $\gamma_2 > 0$, the predictor $\tilde{\delta}(X)$ (uniformly) maximizes the probability that the predicted variable Y lies within the bounds $\delta(X) - \gamma_1$ and $\delta(X) + \gamma_2$.

Remark 1: Though the results here were derived for the location-scale family, the proofs given in this paper hold for a family of distributions which is invariant under a certain group of transformations, e.g. the location family and the scale family.

3 Examples

This section applies the results obtained to the normal distribution and the exponential distribution. In order to compute $H_\Delta(x_1, \ldots, x_r, y)$, make the change of variables from (μ, σ) to (z, σ) with $\mu = y - \sigma z$. Then

$$H_\Delta(x_1, \ldots, x_r, y) = \frac{\int_{-\infty}^{\infty} \int_0^{\infty} \sigma^{-(r+1)} f\{(x_1 - y)/\sigma + z, \ldots, (x_r - y)/\sigma + z, z + \Delta\} dz\, d\sigma}{\int_{-\infty}^{\infty} \int_0^{\infty} \sigma^{-(r+1)} f\{(x_1 - y)/\sigma + z, \ldots, (x_r - y)/\sigma + z, z\} dz\, d\sigma}. \tag{3.1}$$

3.1 Normal Distribution

Suppose that $X_1, \ldots, X_r$ and Y are iid according to the normal distribution with mean μ and variance σ^2, where $\theta = (\mu, \sigma)$, with $\sigma > 0$, is unknown. From (3.1) it is easy to see that

$$H_\Delta(x_1, \ldots, x_r, y) = \frac{\int_0^{\infty} \sigma^{-(r+1)} \exp\{-[r(r+1)^{-1}((\bar{x} - y)/\sigma - \Delta)^2 + S^2/\sigma^2]/2\} d\sigma}{\int_0^{\infty} \sigma^{-(r+1)} \exp\{-[r(r+1)^{-1}(\bar{x} - y)^2/\sigma^2 + S^2/\sigma^2]/2\} d\sigma},$$

with $\bar{x} = \sum_{i=1}^{r} x_i/r$ and $S = \left(\sum_{i=1}^{r} (x_i - \bar{x})^2\right)^{1/2}$, and hence

$$H_{\Delta}(x_1, \ldots, x_r, y) = \frac{\int_0^{\infty} t^{r-1} \exp\{-[r(r+1)^{-1}(ut+\Delta)^2 + t^2]/2\} dt}{\int_0^{\infty} t^{r-1} \exp\{-[r(r+1)^{-1}(ut)^2 + t^2]/2\} dt},$$

with $u = (y - \bar{x})/S$. Then H_{Δ} can be written as

$$H_{\Delta}(x_1, \ldots, x_r, y)$$

$$= \exp\{-r(r+1)^{-1}\Delta^2/2\} \int_0^{\infty} \exp\{-\text{sign}(u) r(r+1)^{-1}\Delta t\} g_u(t) dt , \qquad (3.2)$$

where sign (u) denotes the sign function of u and

$$g_u(t) = \frac{t^{r-1} \exp\{-[r(r+1)^{-1}t^2 + t^2/u^2]/2\}}{\int_0^{\infty} t^{r-1} \exp\{-[r(r+1)^{-1}t^2 + t^2/u^2]/2\} dt} .$$

Note that for $0 < u < u'$ $g_{u'}(t)/g_u(t)$ is increasing in $t > 0$, and for $u < u' < 0$ $g_{u'}(t)/g_u(t)$ is decreasing in $t > 0$. Hence it follows from (3.2) that $H_{\Delta}(x_1, \ldots, x_r, y)$ is non-increasing in y for fixed $x_1, \ldots, x_r$ and $\Delta > 0$ (e.g. Lemma 2 of Lehmann 1986, p. 85) and that $H_{\Delta}(x_1, \ldots, x_r, \bar{x} + cS)$ does not depend on $x_1, \ldots, x_r$, where c is some constant. Hence from Theorem 1

$$\delta(X) = \bar{X} + cS$$

is a UMA equivariant lower prediction limit at confidence level $1 - \alpha$ if the constant c is chosen such that $P_{\theta}\{Y \geq \bar{X} + cS\} = 1 - \alpha$.

In particular for $\alpha = 1/2$, $c = 0$, so that $\bar{X}$ is median unbiased. It is easy to see that the conditions of Theorem 3 are satisfied for $\bar{X}$. Hence $\bar{X}$ maximizes the probability $P_{\theta}\{-\gamma_1 < Y - \delta(X) < \gamma_2\}$ for all $\gamma_1 > 0$, $\gamma_2 > 0$ and all θ within the class of all equivariant median unbiased predictors $\delta(X)$.

Remark 2: It can be shown that $\bar{X}$ is also optimal in the above sense within the class of all location equivariant median unbiased predictors.

3.2 *Exponential Distribution*

Let $X_1<X_2<\ldots<X_n$ be the order statistics in a sample of size n from the exponential distribution with density function $\sigma^{-1}\exp\{-(x-\mu)/\sigma\}$, $x>\mu$ $(E(\mu,\sigma))$, where $\theta=(\mu,\sigma)$, with $\sigma>0$, is unknown. We consider two prediction problems based on $X=(X_1,\ldots,X_r)$ $(1<r\leq n)$.

Consider, first, the problem of predicting the s-th observation in a future independent sample of size m from the same distribution (s-out-of-m system). It follows from (3.1) that

$$H_\Delta(x_1,\ldots,x_r,y)=\frac{\int_0^\infty \sigma^{-(r+1)}\exp\{-(\tilde{S}-n(y-x_1))/\sigma\}\left\{\int_{(y-x_1)/\sigma}^\infty e^{-nz}g(z+\Delta)\,dz\right\}d\sigma}{\int_0^\infty \sigma^{-(r+1)}\exp\{-(\tilde{S}-n(y-x_1))/\sigma\}\left\{\int_{(y-x_1)/\sigma}^\infty e^{-nz}g(z)\,dz\right\}d\sigma}$$

for $x_1<\ldots<x_r$, where $\tilde{S}=\sum_{i=2}^{r}(x_i-x_1)+(n-r)(x_r-x_1)$ and $g(z)$ is the density of Y under $E(0,1)$. Then

$$H_\Delta(x_1,\ldots,x_r,y)=\frac{e^{n\Delta}\int_0^\infty t^{r-1}\exp\{-t(1-n\tilde{u})\}\left\{\int_{t\tilde{u}+\Delta}^\infty e^{-nz}g(z)\,dz\right\}dt}{\int_0^\infty t^{r-1}\exp\{-t(1-n\tilde{u})\}\left\{\int_{t\tilde{u}}^\infty e^{-nz}g(z)\,dz\right\}dt}$$

with $\tilde{u}=(y-x_1)/\tilde{S}$, so that H_Δ can be written as

$$\begin{aligned}H_\Delta(x_1,\ldots,x_r,y)&=e^{n\Delta}\int_0^\infty\{(1-\Psi(t+\Delta))/(1-\Psi(t))\}p_{\tilde{u}}(t)\,dt\ ,\quad \text{if}\quad \tilde{u}>0\ ,\\&=e^{n\Delta}\int_0^\infty\{1-\Psi(-t+\Delta)\}q_{\tilde{u}}(t)\,dt\ ,\qquad\qquad \text{if}\quad \tilde{u}<0\ ,\end{aligned}\tag{3.3}$$

where Ψ denotes the distribution function of the s-th order statistic in a sample of size $n+m$ from $E(0,1)$,

$$p_{\tilde{u}}(t)=t^{r-1}\exp\{-t(\tilde{u}^{-1}-n)\}(1-\Psi(t))/\int_0^\infty t^{r-1}\exp\{-t(\tilde{u}^{-1}-n)\}(1-\Psi(t))\,dt$$

and

$$q_{\tilde{u}}(t) = t^{r-1}\exp\{t(\tilde{u}^{-1}-n)\}/\int_0^\infty t^{r-1}\exp\{t(\tilde{u}^{-1}-n)\}dt\ .$$

Note that for $0<\tilde{u}<\tilde{u}'$ $p_{\tilde{u}'}(t)/p_{\tilde{u}}(t)$ is increasing in $t>0$ and for $\tilde{u}<\tilde{u}'<0$ $q_{\tilde{u}'}(t)/q_{\tilde{u}}(t)$ is decreasing in $t>0$. Since Ψ is a distribution with increasing failure rate (see Theorem 5.8 of Barlow and Proschan, 1975, p. 108), it is easy to see that $(1-\Psi(t+\Delta))/(1-\Psi(t))$ is non-increasing in $t>0$ for $\Delta>0$. Hence it follows from (3.3) that $H_\Delta(x_1,\ldots,x_r,y)$ is non-increasing in y for fixed $x_1<\ldots<x_r$ and $\Delta>0$ and that $H_\Delta(x_1,\ldots,x_r,x_1+c\tilde{S})$ does not depend on $x_1,\ldots,x_r$ if c is some constant. Then Theorem 1 implies that

$$\tilde{\delta}(X) = X_1 + c\tilde{S}$$

is a UMA equivariant lower prediction limit at confidence level $1-\alpha$ if the constant c is chosen such that $P_\theta\{Y\geq X_1+c\tilde{S}\} = 1-\alpha$.

In particular for $\alpha = 1/2$ $\tilde{\delta}$ becomes median unbiased. It is easy to see that the conditions of Theorem 3 are satisfied with $\tilde{\delta}$. Hence $\tilde{\delta}$ maximizes the probability $P_\theta\{-\gamma_1<Y-\delta(X)<\gamma_2\}$ for all $\gamma_1>0$, $\gamma_2>0$ and all θ within the class of all equivariant median unbiased predictors $\delta(X)$.

Next consider the problem of predicting the unobserved value X_s $(r<s\leq n)$ in the sample. It is well known that the conditional distribution of X_s given X is the same as that of the $(s-r)$-th order statistic in a sample of size $n-r$ from $E(x_r,\sigma)$. Hence from (3.1) it is easy to see that

$$H_\Delta(x_1,\ldots,x_r,y) = \frac{\int_0^\infty \sigma^{-(r+1)}e^{-\tilde{S}/\sigma}h\{(y-x_r)/\sigma+\Delta\}d\sigma}{\int_0^\infty \sigma^{-(r+1)}e^{-\tilde{S}/\sigma}h\{(y-x_r)/\sigma\}d\sigma}$$

for $x_1<\ldots<x_r<y$, where h is the density function of the $(s-r)$-th order statistic in a sample of size $n-r$ from $E(0,1)$. Then

$$H_\Delta(x_1,\ldots,x_r,y) = \int_0^\infty t^{r-1}e^{-t}h(t\hat{u}+\Delta)dt/\int_0^\infty t^{r-1}e^{-t}h(t\hat{u})dt$$

with $\hat{u} = (y-x_r)/\tilde{S}$, and hence

$$H_\Delta(x_1,\ldots,x_r,y) = \int_0^\infty \{h(t+\Delta)/h(t)\}f_{\hat{u}}(t)dt\ , \tag{3.4}$$

where

$$f_{\hat{u}}(t) = t^{r-1} e^{-t/\hat{u}} h(t) / \int_0^\infty t^{r-1} e^{-t/\hat{u}} h(t)\, dt \ .$$

Note that for $0<\hat{u}<\hat{u}'$ $f_{\hat{u}'}(t)/f_{\hat{u}}(t)$ is increasing in $t>0$. Since $h(t+\Delta)/h(t)$ is non-increasing in $t>0$ for $\Delta>0$, it follows from (3.4) that $H_\Delta(x_1, \ldots, x_r, y)$ is non-increasing in y for fixed $x_1<\ldots<x_r$ and $\Delta>0$ and that $H_\Delta(x_1, \ldots, x_r, x_r+c\tilde{S})$ does not depend on $x_1, \ldots, x_r$ if c is some constant. Hence from Theorem 1

$$\delta(X) = X_r + c\tilde{S}$$

is a UMA equivariant lower prediction limit at confidence level $1-\alpha$ if the constant c is chosen such that $P_\theta\{X_s > X_r + c\tilde{S}\} = 1-\alpha$.

In particular for $\alpha = 1/2$, $\hat{\delta}$ becomes median unbiased. From (3.4) it is easy to see that the conditions of Theorem 3 are satisfied with $\hat{\delta}$. Hence $\hat{\delta}$ maximizes the probability $P_\theta\{-\gamma_1 < X_s - \delta(X) < \gamma_2\}$ for all $\gamma_1>0$, $\gamma_2>0$ and all θ within the class of all equivariant median unbiased predictors $\delta(X)$.

Acknowledgement: The author wishes to thank the referee for his helpful suggestions and comments.

References

Barlow RE, Proschan F (1975) Statistical Theory of Reliability and Life Testing. Holt, Rinehart & Winston, New York

Hahn GJ, Nelson W (1973) A Survey of Prediction Intervals and Their Applications. J Qual Tech 5:178–188

Lehmann EL (1986) Testing Statistical Hypotheses, 2nd ed. Wiley, New York

Patel JK (1989) Prediction Intervals – A Review. Comm in Statist–Theor Meth 18:2393–2465

Takada Y (1985) Prediction Limit for Observation from the Exponential Distribution. Canadian J Statist 13:325–330

Takada Y (1991) Median Unbiasedness in an Invariant Prediction Problem. Statistics and Probability Letters 12:281–283

Takeuchi K (1975) Statistical Prediction Theory (in Japanese). Baihūkan, Tokyō

Received 26 August 1991
Revised version 26 May 1992

Book Reviews

Le Cam Lucien, Lo Yang Grace: *Asymptotics in Statistics – Some Basic Concepts.* Springer, Berlin Heidelberg New York London Paris Tokyo Hong Kong 1990, p. viii+180, DM 58,–

This book is a new and shorter (compared to [2]) account of Le Cam's asymptotic theory of experiments and related topics. Le Cam's work, which was developed during many years, is centered on a theory which tries to replace complicated statistical models by simpler ones, normal ones for instance, by approximating the given model (experiment) by the simpler one, mainly when the number of observations becomes large. Therefore this is essentially an asymptotic theory which heavily relies on local properties of the likelihood function. The authors are obviously perfectly aware of the problems which are related to asymptotics and the practical relevance of such results. They, indeed, warn the reader against the faithful use of these properties at the beginning of Section 6. These warnings are clearly sensible but they sound rather strange in the middle of a book which is almost entirely devoted to asymptotics. I believe that the natural location of such a warning about the limitation of the asymptotic theory would have been the introduction. Anyway, what would be left of statistics if we would give up asymptotic theory, including all the classical approximations and the applications of the central limit theorem.

Le Cam's work is an essential contribution to the asymptotic theory of statistics. He is well-known for his introduction of contiguity and many developments connected with local asymptotic normality, differentiability in quadratic mean and local asymptotic minimax theorems, but this is only the tip of the iceberg. Many other topics are developed at length in [2] but in such a framework (use of L- and M-spaces instead of the usual probabilistic framework, massive introduction of delicate topological tools) that most of it is certainly not accessible to the average statistician, even the theoretical one.

This book is clearly an attempt to make the essential parts of this theory accessible to anyone who wants to deal with asymptotics. Of course, it is a short book and cannot be compared with [2], but it gives a very nice account of various (and probably the most important) interesting aspects of Le Cam's theory. The most noticeable difference with [2] and a crucial one is the fact that this version is radically more accessible. It avoids the heavy topological machinery and deals with ordinary probabilities. Although it is supposed to be a new version of [1], I found it quite different, much more attractive, and most of the time, easier to read. One can easily discover the main ideas and the principal topics are clearly treated. I suppose that the rewriting benefited from the presence of a co-author and a visible effort to make the book more attractive. I must say that I found it a success and this version is much closer to a textbook than [2]. Nevertheless, even if the authors, in their introduction, mention that they "have attempted to present a few concepts and tools in an elementary manner", I am convinced that most readers will not find the book so "elementary". Indeed if a few parts are, some

are certainly not and the following remark on page 161 gives a better idea of the general level of the book: "except for various details that should be filled in, we consider that Theorem 1 has been proved". I hope that it will also be an incentive for some people to make the additional effort of reading [2] or at least some parts of it.

Of course there is some unavoidable drawback for this reduction. One cannot divide the number of pages by four, use a larger printing size without some loss. The content of the book has been substantially reduced. Also, some delicate points which would need rather long developments, are only sketched or mentioned without proofs with reference to [2]. This phenomenon is more or less important, depending of the sections. Most of the essential points are completely (or almost completely) treated but I suspect that when one comes to some extensions, remarks and various digressions, then only the reader who is already familiar with Le Cam's work will fully benefit from those pages.

After a short introduction which briefly describes the content of the book, the authors come to some important definitions of Wald–Le Cam's theory of experiments. They hopefully avoid the use of L- and M-spaces, nevertheless the concepts of deficiency and related results concerning transitions are not elementary and most results refer to [2]. Proofs are only sketched and I hope that this rather abrupt beginning will not discourage the readers. Actually, the remainder of the book will make a moderate use of these results. One could just have a quick look at this chapter and come back to it later. A complete understanding of what is going on really requires a solid knowledge in topology and some familiarity with Le Cam's work.

The reader should certainly feel happier with Section 3 where the important concept of contiguity is developed in great detail together with the proofs. Classical and very useful properties of Hellinger distances are also derived. I only found a few misprints in the book but two on page 21 where $f \Leftrightarrow a$ should be $f \Rightarrow a$ and similarily for $a \Leftrightarrow e$. I would have preferred if convergence of distributions were defined more precisely (not only compactness) on page 21. These are a few examples of some of the difficulties an unprepared reader could find on his way but he should certainly not bother too much and persevere.

Section 4 is a milestone and will be used frequently in the sequel. It derives the main properties of the asymptotic behaviour of likelihood ratios when the corresponding product probability measures are reasonably close. This section is a new version of Chapter 2 of the original Montreal book and proofs are given in detail although they are rather technical.

Section 5 is devoted to the description of locally asymptotically quadratic families and the construction of relevant estimates. Everything is nicely explained with clear notations. Once again proofs are long and rather technical but given in great detail and the results are by far more general than the (more or less) classical asymptotic treatment of MLE for LAN families. The end of this section introduces more sophisticated results but references to [2] are limited to extensions and do not concern the essential results. There is also an excellent list of references and historical remarks.

Section 6 begins with a warning against illegitimate use of asymptotic theory. As I already mentioned, I found it strange to read it here where it only deals with sequences of i.i.d. variables although those criticisms are certainly valid for the largest part of the book. Properties related to differentiability in quadratic mean and local asymptotic normality are studied in detail. They offer an up-to-date alternative to classical Cramér conditions and the Cramér–Rao inequality as well. Unfortunately, several proofs refer to [2]. I regretted that no proof is given of Lemma 1 in 6.3. Examples are very interesting but sometimes their treatment is rather allusive. 6.4, which gives nonparametric extensions, will probably look harder as well as 6.5 and 6.6 unless the reader is already familiar with this type of arguments. It would also be useful to compare this section with the treatment of DQM and LAN families given by Ibragimov and Has'minskii [3].

Section 7 deals with Bayes estimates, developing both positive and negative results. The authors prove asymptotic normality of posterior distributions and a Bernstein–von Mises theorem in detail. Some counter-examples are also given at the end.

To conclude, I found this short book extremely attractive and interesting although not really elementary. Actually its level is rather unequal but some good knowledge of topology and measure theory is clearly required. The concepts are generally easy to understand but not all proofs are easy to follow, especially when they refer to [2]. Many results are not proved and some are only sketched but if the extensions often suffer from these restrictions, the main points are treated in reasonable detail and the important ideas are very clearly emphathized.

I believe that this book will contribute to make Le Cam's theory popular among much wider circles since it is much more like a textbook than [2]. It develops the basic principles of asymptotics for parametric models (although it is not at all restricted to parametric theory) and I am convinced that its content should now be part of the cultural background of every mathematical statistician. I especially mention Sections 3 and 4 and the first parts of Sections 5, 6 and 7. Some other more delicate parts like Section 2 or 6.5 will, perhaps, be an incentive to read [2] and other related references.

References

[1] Le Cam LM (1969) Théorie Asymptotique de la Décision Statistique. Les Presses de l'Université de Montréal, Montréal

[2] Le Cam LM (1986) Asymptotic Methods in Statistical Decision Theory. Springer, New York

[3] Ibragimov IA, Has'minskii RZ (1981) Statistical Estimation-Asymptotic Theory. Springer, New York

Paris *L. Birgé*

Godehardt E: *Graphs as Structural Models. Advances in Systems Analysis*, Vol. 4. Vieweg, Braunschweig, 1988[1], 1990[2], p. X+214, DM 48,–

This is a book on classification and clustering methods and at the same time a detailed account on classical and recent results in the theory of random graphs and multigraphs. The intimate combination of both is a most interesting feature of this monograph.

Cluster analysis provides methods for subdividing a set of objects into a small number of classes which should be homogeneous and well separated with respect to a given set of data, thereby creating a typology of objects or a structured classification like a hierarchy or a dendrogram. It is well known that this domain has undergone, during the last three decades, a rapid development, and many clustering algorithms have been proposed. The first two chapters of Godehardt's book review these more or less classical methods and discuss their role in the general framework of multivariate statistics and exploratory data analysis.

The main part of the monograph is devoted to the presentation of graph-theoretical classification methods and to the investigation of related probability models with the idea to use them for testing the randomness of a clustering. The most basic model starts with an $n \times n$ matrix of dissimilarities d_{ij} between pairs of objects i and j and a dissimilarity threshold $d > 0$ which distinguishes between similar and dissimilar pairs. Thereby we obtain a similarity graph $G = G(d)$ with a link joining i and j $(i \neq j)$ iff $d_{ij} \leq d$. The connected components, cliques or other subsets of G may be considered as classes or clusters of objects where the level d characterizes the strength of similarity in the classes and can be modified appropriately. The reconstruction of these classes is the aim of many combinatorial, recursive or hierarchical clustering methods like single and complete linkage, (weak or strong) k-linkage, k-overlap, B_k-clustering etc.

The new and major issue in Godehardt's monograph consists in using multigraphs: He assumes that for the same n objects t different similarity graphs $G_1 \ldots, G_t$ are given describing t different features, variables or aspects of the objects (possibly all graphs of the previous type $G(d)$, or defined separately by using metric and nominal variables) and can be considered to be stacked one over the other. These graphs are combined into a single graph G where i and j are 's-linked' iff they are linked in at least s of the t given graphs ($1 \leq s \leq t$ a given integer). Thus we are back to the former model, but with an additional similarity or strength parameter s, and the previous cluster definitions apply with a modified notation (s-components, (k, d, s)-clusters, (k, s)-clusters etc.). Chapter 3 provides special algorithms for the construction of these clusters and of the related dendrograms.

The chapters 4 to 6 are devoted to probability models in this graph-theoretical framework and to the problem of assessing the validity of classifications or clusters obtained by some graph-theoretical algorithm. After a brief review of other probabilistic models, the random graph and random multigraph models are introduced (e.g. the binomial model, the hypergeometric model, the random permutation model). Their mutual relationship and their practical relevance for clustering purposes are discussed (e.g. the equiprobability assumption and the

triangle inequality). The classical random graph theory of Erdös, Rényi, Ling and others for $G = \Gamma_{nN}$ is reviewed and, as an original contribution, generalized to the case of random multigraphs $G = \Gamma_{tnN}$ (where usually the hypergeometric model is adopted with n objects and N randomly distributed edges). This investigation results in various exact or asymptotic formulae for combinatorial numbers, probabilities, expectations or variances for typical clustering-related graph characteristics for Γ_{tnN}, e.g.: the number of trees or s-connected graphs, the s-degree of an arbitrary vertex, the number and size of s-connected components, the distribution of s-isolated vertices, the number of s-isolated trees of size m. The asymptotic results suppose that $n \to \infty$ with $N = N(n) \to \infty$ at a specified rate. We learn, e.g., that for large n, Γ_{tnN} consists almost surely of only one giant s-component, that the number of s-components (minus 1) has an asymptotic Poisson distribution with parameter $\lambda = e^{-c}$ provided that $N(n)$ behaves as $n^{2-1/s} \cdot (c + \log n + o(1))^{1/s}$ for $n \to \infty$; on the other hand, $N(n)$ must grow as fast as $n^{2-2/s} \cdot (c + o(1))^{1/s}$ in order to guarantee asymptotically at least a finite number of s-fold connections in Γ_{tnN}. Other limiting cases for $N(n)$ are considered as well and related to the existence of bounded, but positive limits for the expectation of s-degrees, s-trees, and s-saturated connections. – These Chapters provide an alternative to other probability-based tests for random clustering, e.g. mixture models or point processes. However, whilst these latter methods are primarily based on a clustering situation, the random (multi-) graph model starts from an equidistribution hypothesis without formulating a clustering alternative; therefore power properties cannot be investigated in this model.

Many proofs and formulae have been collected in Chapter 6 which has been added for the second edition of the monograph. This chapter is particularly useful for practical computation. Section 6.7 provides some guidelines and tables which demonstrate the precision of the asymptotic approximation in the finite sample case.

Whilst the emphasis of several chapters lies more on the theoretical side (many results are taken from the author's dissertation and habilitation), it discusses many practical aspects and applied problems, especially in the two introductory chapters (about 70 pages) and in the concluding Chapter 7 where the previous results are applied to the analysis of three data sets from a medical investigation and where the typical exploratory and inferential steps are exemplified in full detail.

The book is a valuable source for classical and recent results in random graph theory and clustering. It contains an up-to-date bibliography with 451 entries (but unfortunately not a subject index) and is written in a very readable and illustrative way. Therefore it can be highly recommended for researchers and practitioners in mathematics, probability and statistics.

Aachen *H. H. Bock*

Metrika (1993) 40:67–94

Metrika

Asymptotic Properties of Minimum Contrast Estimators for Parameters of Boolean Models

LOTHAR HEINRICH

Fachbereich Mathematik, Bergakademie Freiberg, Bernhard-von-Cotta-Str.2, 9200 Freiberg, Germany

Summary: This paper presents a method for the estimation of parameters of random closed sets (racs's) in $\mathbb{R}^d$ based on a single realization within a (large) convex sampling window. The essential idea first applied by Diggle (1981) in a special case consists in defining the estimation by minimizing a suitably defined distance (called contrast function) between the true and the empirical contact distribution function of the racs under consideration, where the most relevant case of Boolean models is discussed in details. The resulting estimates are shown to be strongly consistent (if the racs is ergodic) and asymptotically normal (if the racs is Boolean) when the sampling window expands unboundedly.

Keywords and Phrases: Random closed set, empirical contact distribution, point estimation, Boolean model, Minkowski functionals, asymptotic normality.

AMS 1980 subject classifications: Primary 62 F 10, 62 F 12 Secondary 62 G 30, 62 A 99

1 Introduction

The theory of random closed sets (*racs's*) provides an appropriate tool for describing randomly distributed and irregularly shaped geometrical objects in plane or space. Special classes of *racs's* like germ-grain models, fibre and line processes and random mosaics represent suitable models and allow a reasonable interpretation of such phenomena in several fields of application (e.g. material sciences, texture analysis, biology), see e.g. Stoyan et al. (1987), Hall (1988). Hitherto statistical analysis of such models mainly rests on "the method of moments" and consists essentially in determining statistical quantities of first and second order (for example, of the specific area, boundary length and connectivity number) or of some empirical distance distribution functions providing stationarity and isotropy of the considered model, see Stoyan et al. (1987) [Sect. 6.3]. It is not surprising that the Boolean model (Poisson grain model) in particular that with circular grains is one of the most treated models, see Dupač (1980),

0026–1335/93/2/67–94 $2.50

Diggle (1981). Till now only a few work has been done in the field of asymptotic properties of the estimators, see Baddeley (1980) and Mase (1982).

The main purpose of this paper is to establish a rather general method for obtaining strong consistent estimators of parameters of stationary ergodic *racs* which turn out to be even asymptotically normally distributed in the most relevant and comparatively simple case of Boolean models. Our approach generalizes an earlier quite natural idea of Diggle (1981) who determined in a particular Boolean model point estimations of the unknown parameters by minimizing the integrated squared distance between the true and estimated contact distribution function (*cdf*). For doing this we study in Sect. 3 some properties of the *cdf* of a stationary ergodic *racs* and the asymptotic behaviour of its empirical counterpart. Sect. 4 contains the proof of the strong consistency of minimum contrast (*mc*-) estimates in the general case and the consequences arising in the special situation of Boolean models. The proof of asymptotic normality for *mc*-estimates in Boolean models is carried out in Sect. 5 based on an approximation of this model by *m*-dependent fields. The final Sect. 6 discusses two examples in some detail demonstrating the feasibility of the suggested estimation procedure. Some concluding remarks concern possible improvements and proposals to extend the applicability of our *mc*-estimation procedures.

Next in this section we introduce some basic facts from theory and statistics of *racs* and define *mc*-estimates via a general family of contrast functions and a specified one expressing the "contrast" between the true and the empirical *cdf*.

A *racs* Z in $\mathbb{R}^d$, $d \geq 1$, is defined to be an $(\mathfrak{A}, \sigma_f)$-measurable mapping from a hypothetical probability space $[\Omega, \mathfrak{A}, \mathbb{P}]$ into the measurable space $[\mathscr{F}, \sigma_f]$ of all closed subsets of $\mathbb{R}^d$ equipped with the σ-field generated by the families $\mathscr{F}_K = \{F \in \mathscr{F}: F \cap K \neq \emptyset\}$, $K \in \mathscr{K}$, where $\mathscr{K}$ denotes the class of compact subsets of $\mathbb{R}^d$. We recall the well-known fact that the distribution P_Z induced on $[\mathscr{F}, \sigma_f]$ by Z is entirely determined by the hitting function T_Z,

$$T_Z(K) = P_Z(\mathscr{F}_K) = \mathbb{P}(Z \cap K \neq \emptyset) \quad \text{for} \quad K \in \mathscr{K} \ ,$$

see Matheron (1975). Throughout we only consider stationary *racs's* Z, i.e. $Z + x \overset{d}{=} Z$ for all $x \in \mathbb{R}^d$, where here and henceforth $\overset{d}{=}$ ($\overset{d}{\to}$) means equality (convergence) in distribution.

The *cdf* $H_B(r)$, $r \geq 0$, of a stationary *racs* Z is defined by using a so-called structuring element $B \in \mathscr{K}$ containing the origin 0:

$$H_B(r) := \mathbb{P}(Z \cap rB \neq \emptyset | 0 \notin Z) = 1 - \frac{\mathbb{P}(Z \cap rB = \emptyset)}{\mathbb{P}(0 \notin Z)} = 1 - \frac{1 - T_Z(rB)}{1 - p} \ , \qquad (1.1)$$

where $p := \mathbb{E}|Z \cap [0, 1]^d| = \mathbb{P}(0 \in Z) = T_Z(\{0\})$ is the volume fraction of Z ($|\cdot|$ denotes the d-dimensional Lebesgue measure). We shall see in Sect. 3 that monotonicity of H_B and $H_B(\infty) = 1$ hold in general only under some addi-

tional conditions. Another important quantity for describing the second-order structure of an isotropic *racs* Z is the *covariance* $C(r)$ defined by

$$C(r) = \mathbb{P}(0 \in Z, x \in Z) = 2p - T_Z(\{0, x\}) \quad \text{with} \quad r = \|x\| \ .$$

From now on we suppose that our stationary *racs* model Z incorporates a vector $\vartheta = (\vartheta_0, \vartheta_1, \ldots, \vartheta_s)$ of parameters which lies in a Borel set $\Theta \subset \mathbb{R}^{s+1}$, $s \geq 0$. In other words, the distribution P_Z belongs to a parametrized family $\mathfrak{P} := \{P_\tau, \tau \in \Theta\}$ of probability measures on $[\mathscr{F}, \sigma_f]$. We put $P_\vartheta = P_Z$, $T_\vartheta = T_Z$ and write $H_B(r, \vartheta)$, $p(\vartheta)$, $C(r, \vartheta)$ instead of $H_B(r)$, p, $C(r)$. Our method for estimating the unknown $\vartheta \in \Theta$ is based on a single observation of Z within a sampling region $A_n \subset \mathbb{R}^d$, where $\{A_n, n \in \mathbb{N}\}$ is a convex averaging sequence of subsets in $\mathbb{R}^d$, i.e. A_n is compact and convex, $A_n \subset A_{n+1}$ for $n \in \mathbb{N}$ and $\sup\{r \geq 0: b(x, r) \subset A_n$ for some $x \in A_n\} \uparrow \infty$ as $n \to \infty$, where $b(x, r) = \{y \in \mathbb{R}^d: \|x - y\| < r\}$, see Daley & Vere-Jones (1988).

The definition presented below modifies the classical concept of *mc*- estimation in Pfanzagl (1969) and its extension to stochastic processes in Mühleis & Wittwer (1980) generalizing earlier approaches of Walker (1963) and Ibragimov (1967).

Definition: A family of σ_f*-measurable functions* $U_n(\tau, (\cdot) \cap A_n)|\mathscr{F} \mapsto [-\infty, +\infty]$, $\tau \in \Theta^c$ (Θ^c = *closure of* Θ), $n \in \mathbb{N}$, *is called a family of contrast functions for* $\mathfrak{P} = \{P_\tau, \tau \in \Theta\}$, *if there exists a function* $V|\Theta \times \Theta^c \mapsto \mathbb{R}^1$ *such that*

$$P_\vartheta\left(\left\{F \in \mathscr{F}: \lim_{n \to \infty} U_n(\tau, F \cap A_n) = V(\vartheta, \tau)\right\}\right) = 1 \quad \textit{for all} \quad \vartheta \in \Theta, \tau \in \Theta^c \ , \tag{1.2}$$

and

$$V(\vartheta, \vartheta) < V(\vartheta, \tau) \quad \textit{for all} \quad \vartheta \in \Theta, \tau \in \Theta^c, \vartheta \neq \tau \ . \tag{1.3}$$

A σ_f*-measurable mapping* $\hat{\vartheta}_n((\cdot) \cap A_n)|\mathscr{F} \mapsto \Theta^c$ *is called mc-estimate w.r.t. the family of contrast functions* $\{U_n(\tau, (\cdot) \cap A_n), \tau \in \Theta^c, n \in \mathbb{N}\}$, *if*

$$U_n(\hat{\vartheta}_n(F \cap A_n), F \cap A_n) = \inf_{\tau \in \Theta^c} U_n(\tau, F \cap A_n) \quad \textit{for all} \quad F \in \mathscr{F} \ . \tag{1.4}$$

For brevity put $U_n(\tau) = U_n(\tau, Z \cap A_n)$ and $\hat{\vartheta}_n = \hat{\vartheta}_n(Z \cap A_n)$ considered as random variables on $[\Omega, \mathfrak{A}, \mathbb{P}]$.

There are various possibilities of constructing families of contrast functions for stationary ergodic *racs*. Before discussing some special cases of contrast

functions we formulate a set of conditions which ensure strong consistency of *mc*-estimates.

Lemma 1: Let $\{U_n(\tau, (\cdot) \cap A_n), \tau \in \Theta^c, n \in \mathbb{N}\}$ *be a family of contrast functions as defined in Sect.* 1 *and assume the following conditions are satisfied:*

$$\Theta \text{ is compact} , \tag{1.5}$$

$$V(\vartheta, \cdot) \text{ is continuous on } \Theta \text{ for all } \vartheta \in \Theta , \tag{1.6}$$

$$\text{For any } \varepsilon > 0 \text{ there exists a } \delta = \delta(\varepsilon) > 0 \text{ such that, for all } \vartheta \in \Theta , \tag{1.7}$$

$$\lim_{m \to \infty} P_\vartheta \left(\bigcup_{n \geq m} \left\{ \sup_{\tau_1, \tau_2 \in \Theta : \|\tau_1 - \tau_2\| \leq \delta} |U_n(\tau_1, F \cap A_n) - U_n(\tau_2, F \cap A_n)| \geq \varepsilon \right\} \right) = 0 .$$

Then the sequence $\{\hat{\vartheta}_n(F \cap A_n), n \in \mathbb{N}\}$ *of mc-estimates defined by* (1.4) *is strongly consistent, i.e.*

$$P_\vartheta \left(\left\{ \lim_{n \to \infty} \hat{\vartheta}_n(F \cap A_n) = \vartheta \right\} \right) = 1 \quad \text{for all} \quad \vartheta \in \Theta .$$

An appropriate way for obtaining a contrast function is the use of certain basic functions in statistical analysis of *racs* (e.g. $H_B(r)\, C(r)$) in order to define a distance measuring the discrepancy between such function and its empirical counterpart obtained from a single observation of Z in A_n. To execute the estimation procedure in an effective way it is essential that at least for some classes of *racs*'s an analytical expression of these functions is known and their empirical versions can be computed from the observed set by using an automatic image analyzer.

Examples:

$$U_n^{(1)}(\tau) = \int_0^{r_0} (H_B(r, \tau) - \hat{H}_{B,n}(r))^2 G(dr)$$

and

$$U_n^{(2)}(\tau) = \int_0^{r_0} (C(r, \tau) - \hat{C}_n(r))^2 G(dr) ,$$

where $r_0 > 0$ is a suitable constant, $G(\cdot)$ is a finite weight measure on $[0, r_0]$ with $G(\{0\}) = G(\{r_0\}) = 0$,

$$1 - \hat{H}_{B,n}(r) = \frac{1 - \hat{p}_{B,n}(r)}{1 - \hat{p}_n} \quad \text{for} \quad r \geq 0 \quad \text{with} \quad \hat{p}_n = \frac{|Z \cap A_n|}{|A_n|} , \tag{1.8}$$

$$\hat{p}_{B,n}(r) = \frac{|(Z \oplus (-rB)) \cap A_n^{r_0 B}|}{|A_n^{r_0 B}|} , \qquad A_n^{r_0 B} = \{x \in A_n : x + r_0 B \subseteq A_n\}$$

and

$$\hat{C}_n(r) = \frac{|(Z \cap A_n^{b(0,r_0)}) \cap (Z - x)|}{|A_n^{b(0,r_0)}|} \quad \text{for} \quad r = \|x\| \geq 0 . \tag{1.9}$$

Remark 1: The estimators (1.8) *and* (1.9) *are completely determined by information gained from Z inside of* A_n *whenever* $0 \leq r \leq r_0$ *(minus-sampling). In case one replaces in* (2.4) *resp.* (2.5) *the eroded window* $A_n^{r_0 B}$ *resp.* $A_n^{b(0,r_0)}$ *by* A_n *the corresponding estimators of* $H_B(r)$ *resp.* $C(r)$ *are computable only if information from outside of* A_n *is available (plus-sampling). We mention the fact that all asymptotic properties established in Sections* 2 *and* 3 *remain in force under these changes.*

The empirical functions $\hat{H}_{B,n}(r)$ and $\hat{C}_n(r)$ can indeed be computed in a relatively simple way and they prove strongly consistent and (asymptotically) unbiased provided the observed *racs* is ergodic. The estimates $\hat{\vartheta}_n$ are obtained in this case by a "least squares" method to minimize $U_n^{(1)}$ and $U_n^{(2)}$, respectively.

Let $Z = \bigcup_{i \in \mathbb{N}} (Z_i + X_i)$ be a stationary *Boolean model* (also known as *Poisson grain model*, see Stoyan et al. (1987), Hall (1988) and Mecke et al. (1990) for details) – one of the best studied and most used *racs* models. Here $\{X_i, i \in \mathbb{N}\}$ forms a stationary Poisson point field in $\mathbb{R}^d$ with intensity $0 < \lambda < \infty$ and $\{Z_i, i \in \mathbb{N}\}$ is a sequence of *i. i. d.* copies of a random compact set Z_0 (called typical grain) in $\mathbb{R}^d$ being independent of the Poisson field. If, in addition, Z_0 is convex and its distribution is invariant under rotation about the origin (entailing the isotropy of Z), then $H_B(r)$ is analytically expressible as follows:

$$H_B(r) = 1 - \exp\left\{-\frac{\lambda}{\omega_d} \sum_{k=1}^{d} \binom{d}{k} r^k \mathbb{E} W_k(Z_0) W_{d-k}(B)\right\} \quad \text{for} \quad r \geq 0 , \tag{1.10}$$

for an arbitrary convex structuring element $B \in \mathcal{K}$, where $\omega_d = |b(0, 1)|$ and $W_k(K)$ for $k = 0, 1, \ldots, d$ denote the Minkowski functionals of a convex body K,

for example: $W_0(K) = |K|$, $dW_1(K) = (d-1)$-dimensional Lebesgue measure of ∂K and $W_d(K) = \omega_d$. In the particular case of $d = 2$ this gives: $W_0(K) =$ *area* of K, $W_1(K) = \frac{1}{2}$ *boundary length* of K and $W_2(K) = \pi$.

For the volume fraction p and the covariance $C(r)$ the following expressions can be obtained:

$$p = 1 - \exp\{-\lambda \mathbb{E}|Z_0|\} \quad \text{and}$$

$$C(r) = p^2 + (1-p)^2[\exp\{\lambda \mathbb{E}|Z_0 \cap (Z_0 - x)|\} - 1] \ . \tag{1.11}$$

The shape of $H_B(r)$ and p and the inequality $|\ln x - \ln y| \geq |x - y|$ for $0 < x, y \leq 1$ (which sharpens the contrast) motivates to introduce the contrast function

$$U_n(\tau) = \int_0^{r_0} \left(\ln \frac{1 - H_B(r, \tau)}{1 - \hat{H}_{B,n}(r)} \right)^2 G(dr) + g_0 \left(\ln \frac{1 - p(\tau)}{1 - \hat{p}_n} \right)^2 \tag{1.12}$$

for some $g_0 \geq 0$.

In the remaining part of this paper we shall study this distance function and investigate the asymptotic properties of the corresponding *mc*-estimates.

2 Results

Theorem 1: Let Z be a stationary racs with distribution P_ϑ, $\vartheta \in \Theta$. Assume that the following conditions hold:

$$\Theta \subset \mathbb{R}^{s+1},\ s \geq 0, \text{ is compact.} \tag{2.1}$$

For each $\vartheta \in \Theta$, P_ϑ is ergodic, that is (see Heinrich (1992b)) (2.2)

$$\lim_{n \to \infty} \frac{1}{|A_n|} \int_{A_n} (1 - T_\vartheta((K_1 + x) \cup K_2))\, dx$$

$$= (1 - T_\vartheta(K_1))(1 - T_\vartheta(K_2)) \quad \text{for} \quad K_1, K_2 \in \mathcal{K} \ .$$

$$B \in \mathcal{K} \text{ is star-shaped, i.e. } rB \subseteq B \text{ for every } r \in [0, 1]. \tag{2.3}$$

For all $\vartheta, \tau \in \Theta$, $\vartheta \neq \tau$, $p(\vartheta) \neq p(\tau)$ (*if* $g_0 > 0$) *or there exists some Borel set* $I = I(\vartheta, \tau) \subset [0, r_0]$ *with* $G(I) > 0$ *such that* $H_B(r, \vartheta) \neq H_B(r, \tau)$ *for all* $r \in I$. (2.4)

For each $r \in [0, r_0]$, $\tau \mapsto H_B(r, \tau)$ *and* $\tau \mapsto p(\tau)$ *are continuous with* (2.5)

$$\sup_{\tau \in \Theta} H_B(r_0, \tau) < 1 \quad \text{and} \quad \sup_{\tau \in \Theta} p(\tau) < 1 \ .$$

Then the family of functions defined by (1.12) *is a family of contrast functions satisfying the conditions of Lemma* 1 *with*

$$V(\vartheta, \tau) = \int_0^{r_0} \left(\ln \frac{1 - H_B(r, \vartheta)}{1 - H_B(r, \tau)} \right)^2 G(dr) + g_0 \left(\ln \frac{1 - p(\vartheta)}{1 - p(\tau)} \right)^2 , \tag{2.6}$$

which in turn implies the strong consistency of the mc-estimates $\{\hat{\vartheta}_n, n \in \mathbb{N}\}$ *w.r.t. the contrast functions* (1.12).

In case of stationary Boolean models the conditions of Theorem 1 take on a relatively simple form. To this end write $\vartheta = (\vartheta_0, \theta) \in (0, \infty) \times \mathbb{R}^s$ for the parameter vector, where

P_θ^0 = distribution of the typical grain Z_0

ϑ_0 = intensity of the underlying Poisson point field

$$w_k(\theta) = \int_{\mathcal{K}} W_k(K) P_\theta^0(dK) < \infty \quad \text{for} \quad k = 0, 1, \ldots, d-1 \ .$$

Corollary 1: Let Z *be a stationary Boolean model in* $\mathbb{R}^d$ *with intensity parameter* $\vartheta_0 \in \Theta_0 \subset (0, \infty)$ *and convex compact rotation-invariant (about the origin) typical grain* Z_0 *having distribution* P_θ^0, $\theta \in \Theta^* \subset \mathbb{R}^s$.
Further, assume that

$\Theta = \Theta_0 \times \Theta^*$ *is compact,* (2.7)

B *is a convex compact subset of* $\mathbb{R}^d$ *containing* 0 *with* (2.8)

$$l := \min\{k \geq 0 \colon W_k(B) > 0\} \leq d - 1 \ ,$$

For all $(\vartheta_0, \theta), (\vartheta_0', \theta') \in \Theta_0 \times \Theta^*$: $(\vartheta_0, \theta) = (\vartheta_0', \theta')$ *iff* $\vartheta_0 w_k(\theta) = \vartheta_0' w_k(\theta')$ (2.9)
for $k = 1, \ldots, d-1$ *and (if* $g_0 > 0$*)* $\vartheta_0 w_0(\theta) = \vartheta_0' w_0(\theta')$,
(If $W_0(B) > 0$, $g_0 > 0$ *it suffices to require that, for all* $\theta, \theta' \in \Theta^*$: $\theta = \theta'$
iff $w_k(\theta) = w_k(\theta')$ *for* $k = 0, 1, \ldots, d-1$.)

The mappings $\theta' \mapsto w_k(\theta')$, $k = 0, 1, \ldots, d-1$, *are continuous.* (2.10)

There exist $d - l$ *distinct points* $r_1, \ldots, r_{d-l} \in (0, r_0)$ *such that* (2.11)

$$\min_{1 \le j \le d-l} G((r_j - \varepsilon, r_j + \varepsilon)) > 0 \quad \text{for any} \quad \varepsilon > 0 \ .$$

Then the sequence of mc-estimates $\hat{\vartheta}_n = (\hat{\vartheta}_{0,n}, \hat{\theta}_n)$ *w.r.t. the contrast functions* (1.12) *is strongly consistent as* $n \to \infty$.

The dependence structure of a Boolean model which is determined by the independence properties of the underlying Poisson point $\{X_i, i \in \mathbb{N}\}$ field and the extent of the random compact set Z_0 can approximately be described by m-dependent random fields with large enough m, see Baddeley (1980), Mase (1982), Hall (1988), Heinrich (1988a, b). Based on the central limit theorem for such type of random fields and the mentioned approximation technique we are able to establish asymptotic normality of the *mc*-estimates $\hat{\vartheta}_n = (\hat{\vartheta}_{0,n}, \hat{\theta}_n)$.

Theorem 2: Assume that the conditions of Corollary 1 *are fulfilled, where* (2.10) *is strengthened by the following condition*

The mappings $\gamma \mapsto w_k(\gamma)$, $k = 0, 1, \ldots, d-1$, *are twice continuously differentiable and* $\int_{\mathcal{K}} (W_k(K))^2 P_\theta^0(dK) < \infty$ *for* $k = 0, 1, \ldots, d-1$. (2.12)

Further, let the matrix $S(\vartheta) = (s_{ij}(\vartheta))_{i,j=0}^s$ *with*

$$s_{ij}(\vartheta) = \mathbb{P} - \lim_{n \to \infty} \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \frac{U_n(\vartheta)}{2}$$

$$= \int_0^{r_0} A_i(r, \vartheta) A_j(r, \vartheta) G(dr) + g_0 \frac{\partial}{\partial \vartheta_i} \ln(1 - p(\vartheta)) \frac{\partial}{\partial \vartheta_j} \ln(1 - p(\vartheta))$$

be non-singular for all $\vartheta = (\vartheta_0, \theta) \in \Theta_0 \times \Theta^*$ *(the limit exists* $\mathbb{P}$ *– a.s. for* $U_n(\cdot)$ *from* (1.12)).

Then the sequence of mc-estimates $\hat{\vartheta}_n = (\hat{\vartheta}_{0,n}, \hat{\theta}_n)$ from Corollary 1 is asymptotically normally distributed, more precisely (with obvious notation) as $n \to \infty$,

$$|A_n|^{1/2}(\hat{\vartheta}_n - \vartheta) \xrightarrow{d} \mathcal{N}(0, S^{-1}(\vartheta)\Sigma(\vartheta)S^{-1}(\vartheta)) \ , \tag{2.13}$$

for all inner points $\vartheta = (\vartheta_0, \theta)$ of $\Theta_0 \times \Theta^*$, where the matrix $\Sigma(\vartheta) = (\sigma_{ij}(\vartheta))_{i,j=0}^{s}$ is defined by

$$\begin{aligned}\sigma_{ij}(\vartheta) &= \lim_{n\to\infty} |A_n| \mathbb{E} \frac{\partial}{\partial \vartheta_i} \frac{U_n(\vartheta)}{2} \frac{\partial}{\partial \vartheta_j} \frac{U_n(\vartheta)}{2} \\ &= \int_0^{r_0}\int_0^{r_0} C_B(s, t, \vartheta) A_i(s, \vartheta) A_j(t, \vartheta) G(ds) G(dt) + C_B(0, 0, \vartheta) B_i(\vartheta) B_j(\vartheta) \\ &\quad + \int_0^{r_0} C_B(0, r, \vartheta) \{A_i(r, \vartheta) B_j(\vartheta) + A_j(r, \vartheta) B_i(\vartheta)\} G(dr)\end{aligned}$$

with

$$A_i(r, \vartheta) = \frac{\partial}{\partial \vartheta_i} \ln(1 - H_B(r, \vartheta)) \ ,$$

$$B_i(\vartheta) = g_0 \frac{\partial}{\partial \vartheta_i} \ln(1 - p(\vartheta)) - \int_0^{r_0} \frac{\partial}{\partial \vartheta_i} \ln(1 - H_B(r, \vartheta)) G(dr) \ ,$$

$$C_B(s, t, \vartheta) = \int_{\mathbb{R}^d} [\exp\{\vartheta_0 \mathbb{E}|(Z_0 \oplus (-sB)) \cap (Z_0 \oplus (-tB) + x)|\} - 1] \, dx$$

and (resulting from (1.10) and (1.11))

$$\ln(1 - H_B(r, \vartheta)) = -\vartheta_0 \left\{ \sum_{k=1}^{d-1} r^k a_k w_k(\theta) + r^d W_0(B) \right\}$$

$$\ln(1 - p(\vartheta)) = -\vartheta_0 w_0(\theta) \quad \text{and} \quad a_k = \binom{d}{k} \frac{W_{d-k}(B)}{\omega_d} \ , \quad k = 1, \ldots, d-1 \ .$$

3 Properties of the Empirical Contact Distribution Function

To begin with we state some properties of the *cdf* of a general stationary *racs*. In the most relevant case $d = 2$ the commonly used structuring elements are $B = b^c(0, 1)$, $B = [0, 1]^2$ resp. $B = e_\varphi = \{x = (x_1, x_2)\colon \|x\| \leq 1, \tan \varphi = x_2/x_1\}$, $0 \leq \varphi \leq \pi$, leading to the spherical, quadratic resp. linear *cdf*.

Lemma 2: Let Z be a stationary racs in $\mathbb{R}^d$ and let B be a compact star-shaped structuring element. Then $r \mapsto H_B(r)$ is right-continuous and non-decreasing. Furthermore we have

$$\lim_{r \to \infty} H_B(r) = 1 \text{ if } Z \text{ is mixing with } 0 < p < 1 \text{ or } b(0, \varepsilon) \subset B \text{ for some } \varepsilon > 0 \ .$$

Remark 2: (i) *If B is open then $r \mapsto H_B(r)$ is left-continuous.* (ii) *The mixing condition imposed on Z to hold $H_B(\infty) = 1$ cannot be replaced by the weaker assumption of ergodicity in general as the following example demonstrates: Consider the stationary ergodic racs*

$$Z = \bigcup_{z \in \mathbb{Z}^2} b^c(z, \delta) + X \quad \text{with} \quad 0 < \delta < 1/2$$

in $\mathbb{R}^2$, where X is uniformly distributed on the unit square $[0, 1]^2$. But for the unit line segment e_0 we get

$$\mathbb{P}(Z \cap re_0 = \emptyset) = 1 - 2\delta \text{ for all } r \geq 1 \text{ implying } H_B(\infty) = \frac{2\delta - \pi\delta^2}{1 - \pi\delta^2} < 1 \ .$$

Proof of Lemma 2: The star-shapedness of $B \subset \mathbb{R}^d$ is equivalent to $r_1 B \subseteq r_2 B$ for $r_1 \leq r_2$ which implies $\{Z \cap r_2 B\} \subseteq \{Z \cap r_1 B\}$ and this in turn gives $H_B(r_1) \leq H_B(r_2)$ for $r_1 \leq r_2$.

Now, let r_n, $n \in \mathbb{N}$, be a non-increasing sequence with $\lim_{n \to \infty} r_n = r_0 \geq 0$. Therefore $r_0 B \subseteq \bigcap_{n \in \mathbb{N}} r_n B =: B^*$, where B^* is compact and non-empty. For $x \in B^*$, there exists a sequence b_n, $n \in \mathbb{N}$, so that $x = r_n b_n$ for all $n \in \mathbb{N}$. By the compactness of B we find a subsequence $b_{n'}$, $n' \in \mathbb{N}$, such that $\lim_{n' \to \infty} b_{n'} = b^* \in B$ and finally $x = \lim_{n' \to \infty} r_{n'} b_{n'} = r_0 b^* \in r_0 B$, that is $r_0 B = \bigcap_{n \in \mathbb{N}} r_n B$ whence it follows that

$$\{Z \cap r_0 B = \emptyset\} = \bigcap_{n \in \mathbb{N}} \{Z \cap r_n B = \emptyset\} \quad \text{and so}$$

$$H_B(r_0) = \lim_{n \to \infty} H_B(r_n) = H_B(r_0 + 0) \ .$$

Recall a stationary *racs* Z is mixing iff, for $K_1, K_2 \in \mathcal{K}$,

$$\lim_{\|x\| \to \infty} \mathbb{P}(Z \cap (K_1 \cup (K_2 + x)) = \emptyset) = \mathbb{P}(Z \cap K_1 = \emptyset)\mathbb{P}(Z \cap K_2 = \emptyset) \ ,$$

see Heinrich (1992b).

Let l_b denote the line segment with end-points 0 and $b \in B \backslash \{0\}$ and fix a number r' with $r' < r$. Clearly,

$$\mathbb{P}(Z \cap rB = \emptyset) \leq \mathbb{P}(Z \cap rl_b = \emptyset) \leq \mathbb{P}(Z \cap (r' l_b \cup rb) = \emptyset) \ .$$

Using the mixing property of Z we get $L := \lim_{r \to \infty} \mathbb{P}(Z \cap rl_b = \emptyset) \leq \mathbb{P}(Z \cap r' l_b = \emptyset)(1 - p)$ and, by letting $r' \to \infty$, $L \leq L(1 - p)$ yielding $L = 0$. Thus $H_B(\infty) = 1$. □

The following lemma can be regarded as a counterpart to the Glivenko-Cantelli theorem in classical statistics establishing the a.s. uniform convergence of the empirical distribution function to the sample distribution function when the sample size tends to infinity.

Lemma 3: Let Z be a stationary ergodic racs in $\mathbb{R}^d$ and assume that B is compact and star-shaped. Then the sequence of empirical cdf's $\{\hat{H}_{B,n}(r), n \in \mathbb{N}\}$ (defined by (1.8)) *is an asymptotically unbiased and uniformly strongly consistent estimate for $H_B(r)$, i.e.*

$$\mathbb{P}\left(\lim_{n \to \infty} \sup_{r \geq 0} |\hat{H}_{B,n}(r) - H_B(r)| = 0\right) = 1 \ . \tag{3.1}$$

Proof of Lemma 3: We begin by showing $\mathbb{P}(\lim_{n \to \infty} \hat{H}_{B,n}(r) = H_B(r)) = 1$ for any fixed $r \geq 0$. For this purpose we employ a spatial ergodic theorem formulated and proved in Daley & Vere-Jones (1988), p. 333, which says that, for any σ_f-measurable $f | \mathcal{F} \mapsto \mathbb{R}^1$ satisfying $\mathbb{E}|f(Z)| < \infty$,

$$\lim_{n \to \infty} \frac{1}{|A_n|} \int_{A_n} f(Z - x)\, dx = \mathbb{E}f(Z) \quad \mathbb{P}\text{-a.s.} \ ,$$

provided Z is a stationary ergodic *racs*.

In the special case

$$f(F) = 1_{F\oplus(-rB)}(0) = \begin{cases} 1 & \text{if } F \cap rB \neq \emptyset \\ 0 & \text{otherwise} \end{cases},$$

where $1_A(\cdot)$ denotes the indicator function of the set A, we obtain

$$\frac{|Z \oplus (-rB) \cap A_n|}{|A_n|} = \frac{1}{A_n} \int_{A_n} 1_{Z\oplus(-rB)}(x)\, dx \to \mathbb{P}(Z \cap rB \neq \emptyset) \quad \mathbb{P}\text{-a.s.}$$

for every $r \geq 0$.

Since $\lim_{n\to\infty} |A_n^{r\circ B}|/|A_n| = 1$ which is valid for every convex averaging sequence $\{A_n, n \in \mathbb{N}\}$, the ergodic theorem yields (with the notation from (1.8))

$$\lim_{n\to\infty} \hat{p}_{B,n}(r) = \mathbb{P}(Z \cap rB \neq \emptyset) \quad \mathbb{P}\text{-a.s. for every} \quad r \geq 0$$

so that

$$\lim_{n\to\infty} \hat{H}_{B,n}(r) = H_B(r) \quad \mathbb{P}\text{-a.s. for every} \quad r \geq 0\ . \tag{3.2}$$

This limit holds also true for $r = \infty$. Namely, in view of the continuity of the Lebesgue and $\mathbb{P}$-measure and since $rB \uparrow \tilde{B} := \bigcup_{r\geq 0} rB$ as $r \uparrow \infty$ we have

$$\lim_{n\to\infty} \lim_{r\to\infty} \hat{p}_{B,n}(r) = \lim_{n\to\infty} \frac{|(Z \oplus (-\tilde{B})) \cap A_n|}{|A_n|} = \mathbb{P}(0 \in Z \oplus (-\tilde{B}))$$

$$= \lim_{r\to\infty} \mathbb{P}(Z \cap rB \neq \emptyset)\ .$$

Further, it is easily checked that $\hat{H}_{B,n}(r)$ is also non-decreasing and right-continuous with $H_B(\infty) := H \leq 1$ and $\hat{H}_{B,n}(\infty) \leq 1$. Therefore, by the dominated convergence theorem,

$$\lim_{n\to\infty} \mathbb{E}\hat{H}_{B,n}(r) = H_B(r) \quad \text{for every} \quad r \geq 0\ .$$

For the passage from (3.2) to the stronger uniform convergence (3.1) the following idea is crucial (similar to the proof of Glivenko-Cantelli's theorem): For every

$N \in \mathbb{N}$, let $0 := r_{0,N} < r_{1,N} < \cdots < r_{N,N} < r_{N+1,N} := \infty$ be a dissection of the positive real axis with $r_{k,N} = \sup\left\{r > r_{k-1,N}\colon H_B(r-0) \le \frac{H}{N} + H_B(r_{k-1,N})\right\}$ for $k = 1, \ldots, N$.

Therefore, for $r \in [r_{k,N}, r_{k+1,N})$, $k = 0, 1, \ldots, N$, we get

$$0 \le H_B(r_{k+1,N} - 0) - H_B(r_{k,N}) \le \frac{H}{N} ,$$

$H_B(r_{k,N}) \le H_B(r) \le H_B(r_{k+1,N} - 0)$ and $\hat{H}_{B,n}(r_{k,N}) \le \hat{H}_{B,n}(r) \le \hat{H}_{B,n}(r_{k+1,N} - 0)$, whence it follows that

$$\hat{H}_{B,n}(r_{k,N}) - H_B(r_{k,N}) - \frac{H}{N} \le \hat{H}_{B,n}(r) - H_B(r)$$

$$\le \hat{H}_{B,n}(r_{k+1,N} - 0) - H_B(r_{k+1,N} - 0) + \frac{H}{N} .$$

Applying the pointwise a.s. convergence (3.2) for $r = r_{k,N+1}$, $k = 0, 1, \ldots, N + 1$, the latter inequality leads to

$$\overline{\lim_{n\to\infty}} \sup_{r\ge 0} |\hat{H}_{B,n}(r) - H_B(r)| \le \frac{1}{N} \quad \mathbb{P}\text{-a.s. for all} \quad N \in \mathbb{N}$$

proving the assertion (3.1) of Lemma 3. □

4 Proof of Lemma 1 and Theorem 1

Proof of Lemma 1: Let $\vartheta \in \Theta$ and $\varepsilon > 0$ arbitrary but fixed. Following an idea of Ibragimov (1967) we find for any $\delta > 0$ a finite number $L = L(\delta)$ of points $\vartheta_1, \ldots, \vartheta_L \in \Theta_\varepsilon := \Theta \backslash b(\vartheta, \varepsilon)$ such that $\bigcup_{j=1}^{L} b(\vartheta_j, \delta)$ covers Θ_ε.

Obviously,

$$\{\|\hat{\vartheta}_n(F \cap A_n) - \vartheta\| \ge \varepsilon\} \subseteq \bigcup_{j=1}^{L} B_j \cup \bigcup_{j=1}^{L} \{\hat{\vartheta}_n(F \cap A_n) \in b(\vartheta_j, \delta)\} \cap \bar{B}_j ,$$

where $B_j = \{U_n(\vartheta_j, F \cap A_n) - U_n(\hat{\vartheta}_n(F \cap A_n), F \cap A_n) \leq \eta\}$ and $\bar{B}_j = \mathscr{F} \setminus B_j$, $j = 1, \ldots, L$, and $\eta > 0$ is given by

$$\eta = \min\{V(\vartheta, \tau) - V(\vartheta, \vartheta) : \tau \in \Theta_\varepsilon\}/4 \ ,$$

where (1.3) and (1.6) guarantee the positivity of η. Further, introduce the events $C_j = \{|U_n(\vartheta_j, F \cap A_n) - V(\vartheta, \vartheta_j)| \leq \eta\}$, $j = 1, \ldots, L$. By $U_n(\hat{\vartheta}_n(F \cap A_n), F \cap A_n) \leq U_n(\vartheta, F \cap A_n)$ and the definition of η it follows that, for $j = 1, \ldots, L$,

$$\begin{aligned} B_j &\subseteq \{U_n(\hat{\vartheta}_n(F \cap A_n), F \cap A_n) \geq U_n(\vartheta_j, F \cap A_n) - \eta\} \cap C_j \cup \bar{C}_j \\ &\subseteq \{U_n(\hat{\vartheta}_n(F \cap A_n), F \cap A_n) \geq V(\vartheta, \vartheta_j) - 2\eta\} \cup \bar{C}_j \\ &\subseteq \{U_n(\vartheta, F \cap A_n) \geq V(\vartheta, \vartheta) + 2\eta\} \cup \{|U_n(\vartheta_j, F \cap A_n) - V(\vartheta, \vartheta_j)| > \eta\} \ . \end{aligned}$$

Hence

$$\begin{aligned} &P_\vartheta\Big(\bigcup_{n \geq m} \{\|\hat{\vartheta}_n(F \cap A_n) - \vartheta\| > \varepsilon\}\Big) \qquad (4.1) \\ &\leq P_\vartheta\Big(\bigcup_{n \geq m} \{|U_n(\vartheta, F \cap A_n) - V(\vartheta, \vartheta)| \geq 2\eta\}\Big) \\ &\quad + \sum_{j=1}^{L} P_\vartheta\Big(\bigcup_{n \geq m} \{|U_n(\vartheta_j, F \cap A_n) - V(\vartheta, \vartheta_j)| > \eta\}\Big) \\ &\quad + P_\vartheta\Big(\bigcup_{j=1}^{L} \bigcup_{n \geq m} \{\|\hat{\vartheta}_n(F \cap A_n) \\ &\quad - \vartheta_j\| < \delta, U_n(\vartheta_j, F \cap A_n) - U_n(\hat{\vartheta}(F \cap A_n), F \cap A_n) > \eta\}\Big) \ . \end{aligned}$$

Taking into account the equivalence

$$\lim_{n \to \infty} X_n = X_0 \quad \mathbb{P}\text{-a.s. iff} \quad \lim_{m \to \infty} \mathbb{P}\Big(\sup_{n \geq m} |X_n - X_0| \geq \varepsilon\Big) = 0 \quad \text{for all} \quad \varepsilon > 0 \ ,$$

we obtain from (1.2) that the first and the second summand on the rhs of (4.1) tend to zero as $m \to \infty$. In view of (1.6) the third term also approaches zero as $m \to \infty$ proving the assertion of Lemma 1. □

Proof of Theorem 1: Let $\vartheta \in \Theta$ be arbitrary but fixed and $P_\vartheta = \mathbb{P} \circ Z^{-1}$. In view of (2.4) and the shape of the function (2.6) it clearly holds

$$0 = V(\vartheta, \vartheta) < V(\vartheta, \tau) \quad \text{for all} \quad \tau \in \Theta, \tau \neq \vartheta \ .$$

We first verify that $\lim_{n\to\infty} U_n(\tau) = V(\vartheta, \tau)$ $\mathbb{P}$-a.s. Using the identity

$$\left(\ln\frac{a}{b}\right)^2 - \left(\ln\frac{a}{c}\right)^2 = \ln\frac{c}{b}\left(\ln\frac{a}{b} + \ln\frac{a}{c}\right)$$

and Schwarz's inequality we obtain

$$|U_n(\tau) - V(\vartheta, \tau)| \leq 2U_n^{1/2}(\vartheta)(U_n(\tau) + V(\vartheta, \tau))^{1/2} \ . \tag{4.2}$$

Because of the estimate

$$U_n(\tau) \leq 2(U_n(\vartheta) + V(\vartheta, \tau)) \tag{4.3}$$

and the boundedness of $V(\vartheta, \tau)$ it suffices therefore to show $\lim_{n\to\infty} U_n(\vartheta) = 0$ $\mathbb{P}$-a.s. Put

$$D_n = \sup_{0\leq r\leq r_0} |\hat{H}_{B,n}(r) - H_B(r, \vartheta)| \quad \text{and} \quad d_n = |\hat{p}_n - p(\vartheta)| \ .$$

The supposed ergodicity of Z and Lemma 3 yield $\mathbb{P}(\lim_{n\to\infty} D_n = \lim_{n\to\infty} d_n = 0) = 1$. Hence, for any fixed $\omega \in \Omega$ (except of an ω-null set) we can choose n so large such that $2D_n \leq 1 - H_B(r_0, \vartheta)$, $2d_n \leq 1 - p(\vartheta)$ and

$$\sup_{0\leq r\leq r_0} \left|\ln\frac{1 - H_B(r, \vartheta)}{1 - \hat{H}_{B,n}(r)}\right| \leq \frac{D_n}{1 - H_B(r_0, \vartheta) - D_n} \quad \text{and}$$

$$\left|\ln\frac{1 - p(\vartheta)}{1 - \hat{p}_n}\right| \leq \frac{d_n}{1 - p(\vartheta) - d_n}$$

proving $U_n(\vartheta) \to 0$ as $n \to \infty$.

In analogy to (4.2) we find that

$$|V(\vartheta, \tau_1) - V(\vartheta, \tau_2)| \leq 2V^{1/2}(\tau_1, \tau_2)(V(\vartheta, \tau_1) + V(\vartheta, \tau_2))^{1/2}$$

and

$$|U_n(\tau_1) - U_n(\tau_2)| \le 2V^{1/2}(\tau_1, \tau_2)(U_n(\tau_1) + U_n(\tau_2))^{1/2} .$$

Condition (2.5) and the dominated convergence theorem ensure $\lim_{\tau_2 \to \tau_1} V(\tau_1, \tau_2) = 0$ which together with $K := \sup_{\vartheta, \tau \in \Theta} V(\vartheta, \tau) < \infty$ gives

$$\sup_{\vartheta \in \Theta} |V(\vartheta, \tau_1) - V(\vartheta, \tau_2)| \to 0 \quad \text{as} \cdot \ \tau_2 \to \tau_1 . \tag{4.4}$$

Moreover the compactness of Θ and $V(\tau_1, \tau_2) = V(\tau_2, \tau_1) \le 2(V(\vartheta, \tau_1) + V(\vartheta, \tau_2))$ for all $\vartheta, \tau_1, \tau_2 \in \Theta$ guarantees that $\omega(\delta) := \sup_{\tau_1, \tau_2 \in \Theta: \|\tau_1 - \tau_2\| \le \delta} V(\tau_1, \tau_2)$ tends to zero as $\delta \downarrow 0$. Combining (4.3) with (4.4) we arrive at

$$\Delta_n(\delta) := \sup_{\tau_1, \tau_2 \in \Theta: \|\tau_1 - \tau_2\| \le \delta} |U_n(\tau_1) - U_n(\tau_2)| \le 4\omega^{1/2}(\delta)(U_n(\vartheta) + K)^{1/2} .$$

For given $\varepsilon > 0$ choose $\delta > 0$ in such a way that $32K\omega(\delta) \le \varepsilon^2$ which leads to

$$\mathbb{P}\left(\sup_{n \ge m} \Delta_n(\delta) \ge \varepsilon\right) \le \mathbb{P}\left(\sup_{n \ge m} U_n(\vartheta) \ge K\right) .$$

Applying the a.s. convergence of $U_n(\vartheta)$ to zero as $n \to \infty$ shows the validity of (1.7). The remaining assertions of Theorem 1 follows from Lemma 1. □

Proof of Corollary 1: Taking into account (1.10) and (2.6) one can reformulate condition (2.4) resp. (1.3) as follows: Setting $\vartheta = (\vartheta_0, \theta)$ and $\vartheta' = (\vartheta_0', \theta')$ the equation

$$\int_0^{r_0} \left[\sum_{k=1}^{d} r^k \binom{d}{k} \frac{W_{d-k}(B)}{\omega_d} (\vartheta_0 w_k(\theta) - \vartheta_0' w_k(\theta'))\right]^2 G(dr)$$

$$+ g_0[\vartheta_0 w_0(\theta) - \vartheta_0' w_0(\theta')]^2 = 0$$

has to imply $(\vartheta_0, \theta) = (\vartheta_0', \theta')$.

In view of (2.8) the latter equation is equivalent to $\vartheta_0 w_0(\theta) = \vartheta_0' w_0(\theta')$ (if $g_0 > 0$) and

$$\sum_{i,j=1}^{d-l} \mu_{i+j} A_i(\vartheta, \vartheta') A_j(\vartheta, \vartheta') = 0 , \tag{4.5}$$

where

$$A_k(\vartheta, \vartheta') = \binom{d}{k} \frac{W_{d-k}(B)}{\omega_d} (\vartheta_0 w_k(\theta) - \vartheta_0' w_k(\theta')) \quad \text{and} \quad \mu_k = \int_0^{r_0} r^k G(dr) \ .$$

The solution of the Stieltjes moment problem (see Shohat & Tamarkin (1963)) reveals that the quadratic form (4.5) is positive definite if the spectrum of the weight measure G consists of at least $d - l$ distinct points which is nothing else but what is required in (2.11). Therefore, (4.5) holds if and only if $A_i(\vartheta, \vartheta') = 0$ for $i = 1, \ldots, d - l$ which coincides with condition (2.9). Hence, the assertion of Corollary 1 is an immediate consequence of Theorem 1. □

5 Proof of Theorem 2

Let us first prove asymptotic normality of the sequence $Y_n = (Y_{n,i})_{i=0}^s$, where

$$Y_{n,i} = \int_0^{r_0} [\hat{H}_{B,n}(r) - H_B(r, \vartheta)] \frac{A_i(r, \vartheta)}{1 - H_B(r, \vartheta)} G(dr)$$

$$+ g_0 \frac{\frac{\partial}{\partial \vartheta_i} \ln(1 - p(\vartheta))}{1 - p(\vartheta)} [\hat{p}_n - p(\vartheta)]$$

for $i = 0, 1, \ldots, s$.

Lemma 4: Under the conditions and with the notation of Theorem 2 we have

$$|A_n|^{1/2} Y_n \xrightarrow{d} \mathcal{N}(0, \Sigma(\vartheta)) \quad \text{as} \quad n \to \infty \ . \tag{5.1}$$

Proof of Lemma 4: Define the random vector $X_n = (X_{n,i})_{i=0}^s$ by

$$X_{n,i} = \int_0^{r_0} [\tilde{p}_{B,n}(r) - p_B(r, \vartheta)] \frac{A_i(r, \vartheta)}{1 - p_B(r, \vartheta)} G(dr) + [\hat{p}_n - p(\vartheta)] \frac{B_i(\vartheta)}{1 - p(\vartheta)} \ ,$$

where

$$\tilde{p}_{B,n}(r) = \frac{|(Z \oplus (-rB)) \cap A_n|}{|A_n|} \quad \text{and} \quad p_B(r, \vartheta) = \mathbb{E}\tilde{p}_{B,n}(r) = \mathbb{P}(0 \in Z \oplus (-rB)) \ .$$

A short calculation using the limits $\lim_{n\to\infty} \hat{p}_n = p(\vartheta)$, $\mathbb{P}$-a.s., and $\lim_{n\to\infty} |A_n| \mathbb{E}(\tilde{p}_{B,n}(r) - \hat{p}_{B,n}(r))^2 = 0$ (uniformly in the interval $[0, r_0]$ which is seen by the help of the below formula (5.2)) together with

$$\hat{H}_{B,n}(r) - H_B(r, \vartheta) = \frac{\hat{p}_{B,n}(r) - p_B(r, \vartheta)}{1 - \hat{p}_n} - \frac{\hat{p}_n - p(\vartheta)}{1 - \hat{p}_n}(1 - H_B(r, \vartheta)) \ ,$$

confirms that $|A_n|^{1/2} Y_n$ and $|A_n|^{1/2} X_n$ have the same limit distribution (if it exists).

To determine this limit distribution we first calculate the asymptotic covariances $\lim_{n\to\infty} |A_n| \mathbb{E} X_{n,i} X_{n,j}$, $0 \le i \le j \le s$, which essentially depend on the covariance function of the standardized empirical process $|A_n|^{1/2}(\tilde{p}_{B,n}(r) - p_B(r, \vartheta))$, $0 \le r \le r_0$.

For brevity write Z^{rB} instead of $Z \oplus (-rB) = \{z - rb: z \in Z, b \in B\}$. Using Fubini's theorem (several times) and taking into account that $Z^{sB} \cup (Z^{tB} - x)$ is again a stationary Boolean model with typical grain $Z_0 \oplus ((-sB) \cup (-tB - x))$ and the original Poisson field $\{X_i, i \in \mathbb{N}\}$ of basic points we get by means of (1.11) the following:

$$|A_n| \mathbb{E}(\tilde{p}_{B,n}(s) - p_B(s, \vartheta))(\tilde{p}_{B,n}(t) - p_B(t, \vartheta))$$

$$= \frac{1}{|A_n|} \int_{\mathbb{R}^d} |A_n \cap (A_n - x)| [\mathbb{P}(0 \in Z^{sB} \cap (Z^{tB} - x)) - \mathbb{P}(0 \in Z^{sB}) \mathbb{P}(x \in Z^{tB})] \, dx$$

$$= \frac{1}{|A_n|} \int_{\mathbb{R}^d} |A_n \cap (A_n - x)| [\mathbb{P}(0 \notin Z^{sB}) \cup (Z^{tB} - x)) - \mathbb{P}(0 \notin Z^{sB}) \mathbb{P}(x \notin Z^{tB})] \, dx$$

$$= \frac{1}{|A_n|} \int_{\mathbb{R}^d} |A_n \cap (A_n - x)| [e^{-\vartheta_0 \mathbb{E}|Z_0^{sB} \cup (Z_0^{tB} - x)|} - e^{-\vartheta_0(\mathbb{E}|Z_0^{sB}| + \mathbb{E}|Z_0^{tB}|)}] \, dx \tag{5.2}$$

$$\le \vartheta_0 \int_{\mathbb{R}^d} \mathbb{E}|Z_0^{sB} \cap (Z_0^{tB} - x)| \, dx \le \vartheta_0 \mathbb{E}|Z_0^{r_0 B}|^2 \ . \tag{5.3}$$

Therefore, by (2.12) and Lebesgue's dominated convergence theorem,

$$\lim_{n\to\infty} |A_n| \mathbb{E}(\tilde{p}_{B,n}(s) - p_B(s, \vartheta))(\tilde{p}_{B,n}(t) - p_B(t, \vartheta))$$

$$= (1 - p_B(s, \vartheta))(1 - p_B(t, \vartheta)) C_B(s, t, \vartheta) \ ,$$

where $C_B(s, t, \vartheta)$ is defined in Theorem 2.

After an elementary computation using the notation of Theorem 2 we arrive at

$$\lim_{n\to\infty} |A_n| \, \mathbb{E} X_{n,i} X_{n,j} = \sigma_{ij}(\vartheta) \ . \tag{5.4}$$

In order to apply the central limit theorem for stationary m-dependent fields we introduce the "truncated" Boolean model

$$Z^{(\rho)} = \bigcup_{i\in\mathbb{N}} (Z_i \cap b^c(0,\rho) + X_i) \ ,$$

where Z_0 and $\{X_i, i \in \mathbb{N}\}$ as in the original model, and define the corresponding random vector $X_n^{(\rho)} = (X_{n,i}^{(\rho)})_{i=0}^s$ by

$$X_{n,i}^{(\rho)} = \int_0^{r_0} [\tilde{p}_{B,n}^{(\rho)}(r) - p_B^{(\rho)}(r,\vartheta)] \frac{A_i(r,\vartheta)}{1 - p_B(r,\vartheta)} G(dr) + [\hat{p}_n^{(\rho)} - p^{(\rho)}(\vartheta)] \frac{B_i(\vartheta)}{1 - p(\vartheta)}$$

for $i = 0, 1, \ldots, s$, where $\tilde{p}_{B,n}^{(\rho)}(r) = |Z^{(\rho)} \oplus (-rB) \cap A_n|/|A_n|$, $p_B^{(\rho)}(r,\vartheta) = \mathbb{E}\tilde{p}_{B,n}^{(\rho)}(r)$ and $\hat{p}_n^{(\rho)} = \tilde{p}_{B,n}^{(\rho)}(0)$, $p^{(\rho)}(\vartheta) = p_B^{(\rho)}(0,\vartheta)$.

Our approximation technique requires to show

$$\lim_{\rho\to\infty} \sup_{n\in\mathbb{N}} |A_n| \, \mathbb{E} \|X_n - X_n^{(\rho)}\|^2 = 0 \tag{5.5}$$

which certainly holds whenever

$$\lim_{\rho\to\infty} \sup_{0\le r\le r_0} \sup_{n\in\mathbb{N}} |A_n| \, \mathbb{D}^2(\tilde{p}_{B,n}(r) - \tilde{p}_{B,n}^{(\rho)}(r)) = 0 \ . \tag{5.6}$$

By a somewhat lengthy calculation quite similar to the above one leading to (5.2) we get

$$\begin{aligned}
&|A_n| \, |\mathbb{E}[\tilde{p}_{B,n}(r) - \tilde{p}_{B,n}^{(\rho)}(r) - (p_B(r,\vartheta) - p_B^{(\rho)}(r,\vartheta))][\tilde{p}_{B,n}(r) - p_B(r,\vartheta)]| \\
&\quad \le \int_{\mathbb{R}^d} |e^{-\vartheta_0 \mathbb{E}|Z_0^{rB} \cup (Z_0^{rB} - x)|} - e^{-2\vartheta_0 \mathbb{E}|Z_0^{rB}|} - e^{-\vartheta_0 \mathbb{E}|Z_0^{rB} \cup ((Z_0 \cap b^c(0,\rho))^{rB} - x)|} \\
&\qquad + e^{-\vartheta_0(\mathbb{E}|Z_0^{rB}| + \mathbb{E}|(Z_0 \cap b^c(0,\rho))^{rB}|)}| \, dx \\
&\quad \le \vartheta_0 \int_{\mathbb{R}^d} (\mathbb{E}|Z_0^{rB}) \cap (Z_0^{rB} - x)| - \mathbb{E}|Z_0^{rB} \cap ((Z_0 \cap b^c(0,\rho))^{rB} - x)|) \, dx
\end{aligned}$$

$$+ \vartheta_0^2 \int_{\mathbb{R}^d} \mathbb{E}|Z_0^{rB} \cap (Z_0^{rB} - x)|(\mathbb{E}|Z_0^{rB}| - \mathbb{E}|(Z_0 \cap b^c(0,\rho))^{rB}|)\, dx$$

$$\leq (c_0^{1/2} + c_0)(\mathbb{E}|(Z_0 \cap \overline{b^c(0,\rho)}) \oplus (-r_0 B)|^2)^{1/2} ,$$

where $\overline{b^c(0,\rho)} = \{x \in \mathbb{R}^d : \|x\| > \rho\}$ and $c_0 = \vartheta_0^2 \mathbb{E}|Z_0^{r_0 B}|^2$.

The same upper bound can be obtained for

$$|A_n|\,|\mathbb{E}[\tilde{p}_{B,n}(r) - \tilde{p}_{B,n}^{(\rho)}(r) - (p_B(r,\vartheta) - p_B^{(\rho)}(r,\vartheta))][\tilde{p}_{B,n}^{(\rho)}(r) - p_B^{(\rho)}(r,\vartheta)]| .$$

Finally, $\mathbb{E}(W_k(Z_0))^2 < \infty$ for $k = 0, 1, \ldots, d-1$, guarantees the validity of (5.6).

In the next step we represent $X_n^{(\rho)}$ as a sum of $m(\rho)$-dependent random vectors. To this end define

$$F_0 = \left[-\frac{1}{2}, \frac{1}{2}\right)^d , \quad V_n = \{z \in \mathbb{Z}^d : F_0 + z \subseteq A_n\} , \quad \partial A_n = A_n \setminus \bigcup_{z \in V_n} (F_0 + z)$$

and the random field $\xi_z^{(\rho)} = (\xi_{z,i}^{(\rho)})_{i=0}^s$, $z \in V_n$, by

$$\xi_{z,i}^{(\rho)} = \int_0^{r_0} (|(Z^{(\rho)} \oplus (-rB)) \cap (F_0 + z)| - p_B^{(\rho)}(r,\vartheta)) \frac{A_i(r,\vartheta)}{1 - p_B(r,\vartheta)} G(dr)$$

$$+ (|Z^{(\rho)} \cap (F_0 + z)| - p^{(\rho)}(\vartheta)) \frac{B_i(\vartheta)}{1 - p_B(\vartheta)} .$$

Since $\lim_{n\to\infty} |\partial A_n|/|A_n| = 0$ and $\# V_n = |A_n \setminus \partial A_n|$ it is not difficult to check that

$$\lim_{n\to\infty} |A_n|^{-1} \mathbb{E}\left(|A_n| X_{n,i}^{(\rho)} - \sum_{z \in V_n} \xi_{z,i}^{(\rho)} \right)^2 = 0 \quad \text{for} \quad i = 0, 1, \ldots, s . \tag{5.7}$$

As a consequence of the Poisson property of the point field $\{X_i, i \in \mathbb{N}\}$ and the independence assumptions imposed on the sequence of grains $\{Z_i, i \in \mathbb{N}\}$ the families of random vectors $\{\xi_z^{(\rho)}, z \in U\}$ and $\{\xi_z^{(\rho)}, z \in V\}$ are stochastically independent for each pair $U, V \subset \mathbb{Z}^d$ of finite sets separated by a distance (= maximum norm) greater than $m = 2(\rho + r_0 d(B))$, where $d(B)$ denotes the diameter of B. This is nothing else but the desired m-dependence of the d-dimensional vector field $\{\xi_z^{(\rho)}, z \in V_n\}$.

Applying the central limit theorem for one-dimensional m-dependent fields (see e.g. Heinrich (1988b)) we obtain

$$(\# V_n)^{-1/2} \sum_{z \in V_n} \sum_{i=0}^{s} a_i \xi_{z,i}^{(\rho)} \xrightarrow{d} \mathcal{N}\left(0, \sum_{i,j=0}^{s} a_i a_j \lim_{n \to \infty} |A_n| \mathbb{E} X_{n,i}^{(\rho)} X_{n,j}^{(\rho)}\right)$$

for every $(a_0, a_1, \ldots, a_s) \in \mathbb{R}^{s+1} \setminus \{(0, 0, \ldots, 0)\}$.

Combining the latter relation with (5.4), (5.6) and (5.7) provides

$$|A_n|^{1/2} \sum_{i=0}^{s} a_i X_{n,i} \xrightarrow{d} \mathcal{N}\left(0, \sum_{i,j=0}^{s} a_i a_j \sigma_{ij}(\vartheta)\right)$$

or equivalently (by employing the method of Cramér-Wold)

$$|A_n|^{1/2} X_n \xrightarrow{d} \mathcal{N}(0, \Sigma(\vartheta))$$

which terminates the proof of Lemma 4. □

We continue to prove Theorem 2. From now on assume that ϑ is an inner point of Θ, that is $b(\vartheta, \delta) \subset \Theta$ for small enough $\delta > 0$. The conditions (2.7) and (2.12) ensure existence and continuity of the partial derivatives $U_n^{(i)}(\vartheta) := \dfrac{\partial}{\partial \vartheta_i} U_n(\vartheta)$ and $U_n^{(i,j)}(\tau) := \dfrac{\partial^2}{\partial \tau_i \partial \tau_j} U_n(\tau)$ for $\tau \in b(\vartheta, \delta)$, $0 \le i, j \le s$.

In view of the consistency of $\hat{\vartheta}_n$ we may assume $\hat{\vartheta}_n \in b(\vartheta, \delta)$ which implies $U_n^{(i)}(\hat{\vartheta}_n) = 0$ for $i = 0, 1, \ldots, s$. Hence, after applying the mean value theorem to the functions $U_n^{(i)}(\tau)$, $0 \le i \le s$ we obtain

$$0 = U_n^{(i)}(\vartheta) + \sum_{j=0}^{s} (\hat{\vartheta}_{j,n} - \vartheta_j) U_n^{(i,j)}(\vartheta_n^{(j)}) , \qquad 0 \le i \le s , \tag{5.8}$$

where $\hat{\vartheta}_n = (\hat{\vartheta}_{i,n})_{i=0}^{s}$, $\vartheta_n^{(j)} = \lambda_j \hat{\vartheta}_n + (1 - \lambda_j)\vartheta \in b(\vartheta, \delta)$ for some $0 \le \lambda_j \le 1$, $0 \le j \le s$.

To accomplish the proof of Theorem 2 we have to determine the asymptotic behaviour of the vector $(U_n^{(i)}(\vartheta))_{i=0}^{s}$ and the random matrix $(U_n^{(i,j)}(\vartheta_n^{(j)})_{i,j=0}^{s}$ as $n \to \infty$.

By virtue of Lemma 3 we can obtain the limiting distribution of $\frac{1}{2}(U_n^{(i)}(\vartheta))_{i=0}^{s}$ on the assumption that $\sup_{0 \le r \le r_0} |\hat{H}_{B,n}(r) - H_B(r, \vartheta)| \le \frac{1}{2}(1 - H_B(r_0, \vartheta))$. Therefore, since $|\ln(1 + z) - z| \le |z|^2$ for $|z| \le \frac{1}{2}$, we have

$$\left|\frac{1}{2}U_n^{(i)}(\vartheta) - Y_{n,i}\right| \le \int_0^{r_0} (\hat{H}_{B,n}(r) - H_B(r,\vartheta))^2 \frac{\left|\frac{\partial}{\partial\vartheta_i}\ln(1 - H_B(r,\vartheta))\right|}{(1 - H_B(r,\vartheta))^2} G(dr)$$

$$+ g_0(\hat{p}_n - p(\vartheta))^2 \frac{\left|\frac{\partial}{\partial\vartheta_i}\ln(1 - p(\vartheta))\right|}{(1 - p(\vartheta))^2} .$$

The rhs of this estimate multiplied by $|A_n|^{1/2}$ tends to zero in probability which results from $H_B(r_0, \vartheta) < 1$ and the inequalities

$$\sup_{0\le r\le r_0} \mathbb{E}(\hat{p}_{B,n}(r) - p(r,\vartheta))^2 \le \frac{\vartheta_0}{|A_n^{r_0 B}|}\mathbb{E}|Z_0 \oplus (-r_0 B)|^2 \qquad \text{and}$$

$$\mathbb{E}(\hat{p}_n - p(\vartheta))^2 \le \frac{\vartheta_0}{|A_n|}\mathbb{E}|Z_0|^2 .$$

derived from (5.2). Hence, by Lemma 4,

$$\frac{1}{2}(U_n^{(i)}(\vartheta))_{i=0}^s \xrightarrow{d} \mathcal{N}(0, \Sigma(\vartheta)) \qquad \text{as } n \to \infty . \tag{5.9}$$

Straightforward differentiation yields

$$\frac{1}{2}U_n^{(i,j)}(\tau) = \int_0^{r_0} A_i(r,\tau)A_j(r,\tau)G(dr) + g_0\frac{\partial}{\partial\tau_i}\ln(1 - p(\tau))\frac{\partial}{\partial\tau_j}\ln(1 - p(\tau))$$

$$+ \int_0^{r_0} \ln\left(\frac{1 - H_B(r,\tau)}{1 - \hat{H}_{B,n}(r)}\right)\frac{\partial^2}{\partial\tau_i\partial\tau_j}\ln(1 - H_B(r,\tau))G(dr)$$

$$+ g_0 \ln\left(\frac{1 - p(\tau)}{1 - \hat{p}_n}\right)\frac{\partial^2}{\partial\tau_i\partial_j}\ln(1 - p(\tau)) .$$

Making use of the $\mathbb{P}$-a.s. limits $\lim_{n\to\infty}\max_{0\le j\le s}\|\vartheta_n^{(j)} - \vartheta\| = 0$ (by Theorem 1) and $\lim_{n\to\infty}\sup_{0\le r\le r_0}\left|\ln\frac{1 - H_B(r,\vartheta)}{1 - \hat{H}_{B,n}(r)}\right| = 0$ (by Lemma 3) and condition (2.12) we get

$$\lim_{n\to\infty}\frac{1}{2}U_n^{(i,j)}(\vartheta_n^{(j)}) = s_{ij}(\vartheta) \quad \mathbb{P}\text{-a.s.} \quad 0 \le i,j \le s .$$

The assumed invertibility of $S(\vartheta)$ enables to treat the system of equations (5.8) under the condition $\det(U_n^{(i,j)}(\vartheta_n^{(j)}))_{i,j=0}^{s} \geq \frac{1}{2} \det S(\vartheta)$.

Therefore,

$$|A_n|^{1/2}(\hat{\vartheta}_n - \vartheta) = -|A_n|^{1/2}(U_n^{(i)}(\vartheta))_{i=0}^{s}((U_n^{(i,j)}(\vartheta_n^{(j)}))_{i,j=0}^{s})^{-1} ,$$

whence together with (5.9) it follows (2.13). This completes the proof of Theorem 2. □

By a special choice of the weight measure G (which may also be signed with bounded variation) one can derive the limit distribution of $|A_n|^{1/2}(\hat{H}_{B,n}(r_i) - H_B(r_i))_{i=1}^{k}$. For this purpose one can apply Lemma 4 for $s = 0$ and

$$G(\cdot) = \sum_{i=1}^{k} c_i \frac{1 - H_B(r_i, \vartheta)}{A_0(r_i, \vartheta)} \delta_{r_i}(\cdot) \quad \text{for} \quad c_1, \ldots, c_k \in \mathbb{R}^1 .$$

Corollary 2: Let Z be a stationary Boolean model in $\mathbb{R}^d$ with intensity parameter $\vartheta_0 \in (0, \infty)$ and convex compact rotation-invariant typical grain Z_0 satisfying $\mathbb{E}(W_k(Z_0))^2 < \infty$ for $k = 0, 1, \ldots, d-1$. Further, assume that B is a convex compact subset of $\mathbb{R}^d$ containing 0. Then, for every k-tuple $0 \leq r_1 < \cdots < r_k < \infty$,

$$|A_n|^{1/2}(\hat{H}_{B,n}(r_i) - H_B(r_i)) \xrightarrow{d} \mathcal{N}(0, \Gamma) \quad \text{as} \quad n \to \infty ,$$

where $\Gamma = (\gamma(r_i, r_j))_{i,j=1}^{k}$ with

$$\gamma(s, t) = \frac{1}{(1-p)^2} \int_{\mathbb{R}^d} [e^{-\vartheta_0 \mathbb{E}|Z_0^{sB} \cup (Z_0^{tB} + x)|} - e^{-\vartheta_0 \mathbb{E}|Z_0^{sB} \cup (Z_0 + x)|}$$

$$- e^{-\vartheta_0 \mathbb{E}|Z_0^{tB} \cup (Z_0 + x)|} + e^{-\vartheta_0 \mathbb{E}|Z_0 \cup (Z_0 + x)|}] \, dx .$$

Quite similar arguments as employed in the proof of Theorem 5 in Heinrich (1988b) and the additional condition

$$\sup_{\substack{0 \leq s \leq t \leq r_0 \\ t-s \leq h}} \mathbb{E}|Z_0^{tB}|^3 |Z_0^{tB} \setminus Z_0^{sB}| \leq c(r_0, B)h \quad \text{for} \quad h > 0$$

enable us to prove that the empirical process $(|A_n|^{1/2}(\hat{H}_{B,n}(r) - H_B(r)))_{0 \leq r \leq r_0}$ converges weakly in the Skorokhod-space $D[0, r_0]$ to an $\mathbb{P}$-a.s. continuous Gaussian process with zero means and covariance function $\gamma(s, t)$.

6 Estimation of the Parameter Vector $(\vartheta_0, \mathbb{E}W_0(Z_0), \ldots, \mathbb{E}W_{d-1}(Z_0))$

We adhere to the notation of Theorem 2 and consider the special case $s = d$, $\theta = (\vartheta_1, \ldots, \vartheta_d)$, where $w_k(\theta) = \vartheta_{k+1}$, $k = 0, 1, \ldots, d-1$, so that

$$U_n(\vartheta) = \int_0^{r_0} \left(\vartheta_0 \left(\sum_{i=1}^{d-1} a_i \vartheta_{i+1} r^i + W_0(B) r^d \right) + \ln(1 - \hat{H}_{B,n}(r)) \right)^2 G(dr)$$

$$+ g_0(\vartheta_0 \vartheta_1 + \ln(1 - \hat{p}_n))^2 \ .$$

The estimations $(\hat{\vartheta}_{i,n})_{i=1}^d$ are the solutions of the equations

$$\frac{\partial}{\partial \vartheta_i} U_n(\vartheta) = 0 \quad \text{for} \quad i = 0, 1, \ldots, d \ . \tag{6.1}$$

To prevent that the system (6.1) is underdetermined we suppose $g_0 > 0$ and $W_0(B) > 0$ (implying $a_j > 0$, $j = 1, \ldots, d-1$ and exclude the trivial solution $\hat{\vartheta}_{0,n} = 0$. Therefore after some rearrangements the equations (6.1) are seen to be equivalent to the system of equations

$$0 = \vartheta_0 \vartheta_1 + \ln(1 - \hat{p}_n) \tag{6.2}$$

$$L_{i,n} = \vartheta_0 \left(\sum_{j=1}^{d-1} a_j \mu_{i+j} \vartheta_{j+1} + \mu_{i+d} W_0(B) \right) \quad \text{for} \quad i = 1, \ldots, d \ , \tag{6.3}$$

where

$$\mu_i = \int_0^{r_0} r^i G(dr) \ , \qquad L_{i,n} = -\int_0^{r_0} r^i \ln(1 - \hat{H}_{B,n}(r)) G(dr) \ , \qquad i \geq 1 \ .$$

Setting $t_j = a_j \vartheta_0 \vartheta_{j+1}$ for $j = 1, \ldots, d-1$ and $t_d = \vartheta_0 W_0(B)$ the system (6.3) takes the form

$$L_{i,n} = \sum_{j=1}^{d} \mu_{i+j} t_j \quad \text{for} \quad i = 1, \ldots, d \ . \tag{6.4}$$

Since the coefficients μ_{i+j} are moments w.r.t. the measure G the matrix $M = (\mu_{i+j})_{i,j=1}^d$ is non-negative definite. Moreover, arguing as in the proof of Corollary

1, the solution of the Stieltjes moment problem (see Shohat & Tamarkin (1963)) yields that $\det M > 0$ holds if the spectrum of the weight measure G consists of at least d distinct points, i.e. (2.10) is fulfilled for $l = 0$.

Exactly this is the condition for a unique solution of (6.4).

Denote by $M_{j,n}$ the matrix which arises from M by replacing its jth column by the column vector $(L_{1,n}, \ldots, L_{d,n})'$.

Therefore, we can express the final solution of the equations (6.2) and (6.3) as follows:

$$\hat{\vartheta}_{0,n} = \frac{1}{W_0(B)} \frac{\det M_{d,n}}{\det M}, \qquad \hat{\vartheta}_{1,n} = -W_0(B) \ln(1 - \hat{p}_n) \frac{\det M}{\det M_{d,n}} \tag{6.5}$$

and

$$\hat{\vartheta}_{j+1,n} = \frac{\omega_d W_0(B)}{\binom{d}{j} W_{d-j}(B)} \frac{\det M_{j,n}}{\det M_{d,n}} \quad \text{for} \quad j = 1, \ldots, d-1 \tag{6.6}$$

To close this example we apply Corollary 1 and Theorem 2 to the just obtained estimates.

Corollary 3: Under the conditions of Corollary 2 and the additional assumptions $g_0 > 0$, $W_0(B) > 0$ *and* (2.10) *for* $l = 0$ *the sequence of mc-estimates* $(\hat{\vartheta}_{i,n})_{i=0}^{d}$ *for the parameter vector* $(\vartheta_0, \mathbb{E}W_0(Z_0), \ldots, \mathbb{E}W_{d-1}(Z_0))$ *given by* (6.5) *and* (6.6) *is strongly consistent and asymptotically normally distributed with covariance matrix* $S^{-1}(\vartheta)\Sigma(\vartheta)S^{-1}(\vartheta)$ *(where* $s = d$ and $\vartheta_k = w_{k-1}(\theta) = \mathbb{E}W_{k-1}(Z_0)$ *for* $k = 1, \ldots, d$*) given in Theorem 2.*

Finally, let us consider the special case of a Boolean model in $\mathbb{R}^2$ with circular grains, where the distribution of the radius R_0 of the typical grain Z_0 is given by the probability density $f_{\gamma,\mu}$ with

$$f_{\gamma,\mu}(x) = \begin{cases} \mu e^{-\mu(x-\gamma)} & \text{if } x \geq \gamma > 0 \\ 0 & \text{otherwise} \end{cases}.$$

It is quickly seen that

$$\mathbb{E}W_0(Z_0) = \pi \mathbb{E}R_0^2 = \pi\left(\left(\frac{1}{\mu} + \gamma\right)^2 + \frac{1}{\mu^2}\right)$$

$$\mathbb{E}W_1(Z_0) = \pi \mathbb{E}R_0 = \pi\left(\frac{1}{\mu} + \gamma\right).$$

Replacing $\mathbb{E}W_0(Z_0)$ by $\hat{\vartheta}_{1,n}$ from (6.5) and $\mathbb{E}W_1(Z_0)$ by $\hat{\vartheta}_{2,n}$ from (6.6) (for $d = 2$) and solving these equations we obtain the estimates

$$\hat{\mu}_n = \left(\frac{\hat{\vartheta}_{1,n}}{\pi} - \left(\frac{\hat{\vartheta}_{2,n}}{\pi}\right)^2\right)^{-1/2} \quad \text{and} \quad \hat{\gamma}_n = \frac{\hat{\vartheta}_{2,n}}{\pi} - \left(\frac{\hat{\vartheta}_{1,n}}{\pi} - \left(\frac{\hat{\vartheta}_{2,n}}{\pi}\right)^2\right)^{1/2} .$$

One can easily confirm that $\hat{\mu}_n$, $\hat{\gamma}_n$ and $\hat{\vartheta}_{0,n}$ from (6.5) coincide with the solution of the equations (6.1) with $\vartheta = (\vartheta_0, \mu, \gamma)$ and

$$U_n(\vartheta) = \int_0^{r_0} \left[2W_1(B)\vartheta_0\left(\frac{1}{\mu} + \gamma\right)r + W_0(B)r^2 + \ln(1 - \hat{H}_{B,n}(r))\right]^2 G(dr)$$

$$+ g_0\left[\pi\vartheta_0\left(\frac{1}{\mu} + \gamma\right)^2 + \frac{\pi\vartheta_0}{\mu^2} + \ln(1 - \hat{p}_n)\right]^2 .$$

To conclude with we make some brief remarks about some desirable improvements and comparisons and possible extensions of the *mc*-estimates studied in the previous sections which should be subject of future investigations.

(i) Estimation variances: It is seen from (6.5) and (6.6) that in general expectation values as well as variances of *mc*-estimates (in the sense of (1.4)) w.r.t. the family of contrast functions (1.12) do not exist so that the obtained estimates could prove rather instable. This unfavourable effect can be reduced upon replacing G by the "randomly truncated" weight measure

$$G_n^{(\delta)}(\cdot) = \int_0^{r_0} 1_{(\cdot)\cap\{s:\, \hat{H}_{B,n}(s) \le 1-\delta\}}(r)G(dr) \quad \text{for a suitable} \quad \delta > 0 .$$

In any case the weight measure G has strong influence on the covariance matrix of the Gaussian limit distribution of $\hat{\vartheta}_n$ given in Theorem 2. So far rules for an "optimal" choice of G are unknown.

(ii) Choice of the contrast function: Let us introduce a rather general contrast (or distance) function:

$$U_n(\tau) = d(\hat{H}_{B,n}(\cdot), H_B(\cdot, \tau)) := \int_0^{r_0} f_1(\hat{H}_{B,n}(r), H_B(r, \tau))G(dr) + g_0 f_2(\hat{p}_n, p(\tau)) .$$

The troubles with the estimation variances could be avoided (at least partly) by replacing the kernel functions $f_1(y, z) = f_2(y, z) = (\ln y - \ln z)^2$ in (1.12), say, by $f_1(y, z) = f_2(y, z) = (y - z)^2$. At this point it arises the question: How to find non-negative kernels $f_i(y, z)$, $i = 1, 2$, which are best in an approximation-

theoretic sense for our *mc*-method. Some results in this direction (in a slightly different context) were obtained by Jones (1989). A further non-trivial problem concerns the comparison of *mc*-procedures working with distance $d(\hat{H}_{B,n}(\cdot), H_B(\cdot, \tau))$ and those based on the distance $d(\hat{C}_n(\cdot), C(\cdot, \tau))$, see (1.9).

(iii) Another estimation method: In case of a planar Boolean model Z with convex, rotation-invariant grains there exists a further procedure (called "method of intensities", for details see Stoyan et al. (1987) and Mecke et al. (1990)) quite different from ours to obtain estimations for ϑ_0, $\vartheta_1 = \mathbb{E}W_0(Z_0)$ and $\vartheta_2 = \mathbb{E}W_1(Z_0)$.

It rests on a system of linear equations establishing relations between mean area, mean boundary length, mean connectivity number of $Z \cap A_n$ and the corresponding densities of the quermassintegrals of Z. Since these densities are expressible in terms of ϑ_0, ϑ_1 and ϑ_2 (see e.g. Mecke et al. (1990), p. 83) one can derive the following estimators:

$$\tilde{\vartheta}_{0,n} = \frac{\hat{\chi}_n^+}{\hat{p}_n}, \qquad \tilde{\vartheta}_{1,n} = -\frac{\hat{p}_n}{\hat{\chi}_n^+} \ln(1 - \hat{p}_n), \qquad \tilde{\vartheta}_{2,n} = \frac{\hat{l}_n}{2\hat{p}_n},$$

where $\hat{p}_n$ is the estimator of the specific area of Z given in (1.8) and

$$\hat{l}_n = (\text{length of } \partial Z \text{ in } A_n)/|A_n|,$$

$$\hat{\chi}_n^+ = (\text{number of positive tangent points w.r.t. the horizontal bottom line})/|A|.$$

One can expect that the limit distribution of $|A_n|^{1/2}((\tilde{\vartheta}_{0,n}, \tilde{\vartheta}_{1,n}, \tilde{\vartheta}_{2,n}) - (\vartheta_0, \vartheta_1, \vartheta_2))$ as $n \to \infty$ is again Gaussian, however an explicit expression of the limiting covariance matrix seems to be hardly to get. In any case a more precise comparison between these estimators and those from (6.5) and (6.6) based on theoretical arguments and simulation studies would be desirable.

(iv) Estimation of more than $d + 1$ parameters: If the parameter vector $\theta \in \Theta^*$ (with the notation of Corollary 1) consists of d (or less) components it is an appropriate and mostly successful way to solve the equations $\hat{\vartheta}_k = w_{k-1}(\theta)$ for $k = 1, \ldots, d$ with $\hat{\vartheta}_1, \ldots, \hat{\vartheta}_d$ from (6.5) and (6.6). When the quantities $w_k(\theta)$, $k = 0, 1, \ldots, d - 1$, depend on more than d parameters some difficulties arise concerning the uniqueness of our estimation problem. A way out of this dilemma could be the following: Compute the estimates $\hat{\vartheta}_{i,n} = \hat{\vartheta}_{i,n}(B_j)$, $i = 1, \ldots, d$, from (6.5) and (6.6) for quite different structuring elements B_j, $j = 1, \ldots, k$, satisfying (2.8). Determine the estimate $\hat{\theta}_n$ such that

$$\sum_{j=1}^{k} \sum_{i=1}^{d} (w_{i-1}(\hat{\theta}_n) - \hat{\vartheta}_{i,n}(B_j))^2 \to \min.$$

Acknowledgement: This paper was written during the author's stay at the SFB 343 *Discrete Structures in Mathematics* of the University of Bielefeld. The author wishes to thank the *Deutsche Forschungsgemeinschaft* for supporting this reseach.

References

Baddeley A (1980) A limit theorem for statistics of spatial data. *Adv. Appl. Prob.* 12:447–461

Daley DJ, Vere-Jones D (1988) *An introduction to the theory of point processes.* Springer, New York

Diggle PJ (1981) Binary mosiacs and the spatial pattern of heather. *Biometrics* 37:531–539

Dupač V (1980) Parameter estimation in the Poisson field of discs. *Biometrika* 67:187–190

Hall P (1988) *Introduction to the theory of coverage processes.* Wiley & Sons, New York

Heinrich L (1988a) Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary Poisson cluster processes. *Statistics* 19:87–106

Heinrich L (1988b) Asymptotic behaviour of an empirical nearest-neighbour distance function for stationary Poisson cluster processes. *Math. Nachr.* 136:131–148

Heinrich L (1992a) Minimum contrast estimates for parameters of spatial ergodic point processes. In: *Transactions of the 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions*, August 27 to 31, 1990, Academia Publishing House, Prague, pp 479–492

Heinrich L (1992b) On existence and mixing properties of germ-grain models. *Statistics*, Statistics 23 (in print)

Ibragimov IA (1967) On maximum likelihood estimation of parameters of the spectral density of stationary time series. *Theor. Appl. Prob.* 12:115–119

Jones LK (1989) Approximation-theoretic derivation of logarithmic entropy principles for inverse problems and unique extension of the maximum entropy method to incorporate prior knowledge. *SIAM J. Appl. Math.* 49:650–661

Mase S (1982) Asymptotic properties of stereological estimators of volume fraction for stationary random sets. *J. Appl. Prob.* 19:111–126

Matheron G (1975) *Random sets and integral geometry.* Wiley & Sons, New York

Mecke J, Schneider R, Stoyan D, Weil W (1990) *Stochastische Geometrie.* DMV-Seminar, Band 16, Birkhäuser, Basel

Mühleis W, Wittwer G (1980) Minimum-Kontrast Schätzungen für stochastische Prozesse. *Math. Op.-forsch. Statist., Ser. Statistics* 11:219–227

Nguyen XX, Zessin H (1979) Ergodic theorems for spatial processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 48:133–158

Pfanzagl J (1969) On the measurability and consistency of minimum contrast estimates. *Metrika* 14:249–272

Stoyan D, Kendall WS, Mecke J (1987) *Stochastic geometry and its applications.* Akademie-Verlag, Berlin

Stoyan D (1992) Statistical estimation of model parameters of planar Neyman-Scott cluster processes. Metrika 39:67–74

Shohat JA, Tamarkin JD (1963) *The problem of moments.* 3rd ed., American Math. Soc., Rhode Island

Walker AM (1964) Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time series. *J. Austr. Math. Soc.* 4:363–384

Received 31.07.1991

Metrika (1993) 40:95–113

Metrika

Economic Designs of Single and Double Screening Procedures

DO SUN BAI AND MIN KOO LEE

Department of Industrial Engineering, Korea Advanced Institute of Science and Technology, Gusung-dong 373-1, Yusung-gu Daejon 305-701, Korea

Summary: Economic designs of single and double screening procedures for improving outgoing product quality based on two screening variables are presented for the case of one-sided specification limit. Two screening variables are observed simultaneously in single screening procedure and are observed sequentially in double screening procedure. It is assumed that the performance variable and the two screening variables are jointly normally distributed. Three quality cost functions – constant, linear, and quadratic – are considered. Cost models are constructed which involve screening inspection cost, and costs of accepted and rejected item. Methods of finding the optimal cutoff values are presented and a numerical example is given.

1 Introduction

Screening procedures are widely used in industries to improve outgoing quality of products. In some situations, however, it would be impossible or not economical to screen items using the performance variable directly. For example, the voltage at an internal point of an electronic device which is difficult to measure directly may be the major quality characteristic of interest. It may, however, be possible to screen the devices by measuring the voltage at an external point that is correlated with the internal voltage. Therefore, we can inspect with the screening variable rather than the performance variable directly. An inspection procedure using the screening variable would be appropriate if the measurement on the performance variable is more difficult or expensive to make than the screening variable. The idea of using correlated variables in a screening procedure has received much attention. Much of the early works in this area is focused on increasing the proportion of units within specifications from the current value γ to a specified higher proportion δ after screening, i.e., improving the outgoing quality; for example, see Owen et al. (1975), Owen and Boddie (1976), Owen and Su (1977), Owen et al. (1981), Madsen (1982) and Menzefricke (1984). Tsai and Moskowitz (1986) considered a single screening procedure in which two screening variables are observed simultaneously to make one of two

0026–1335/93/2/95–113 $2.50

decisions – accept or reject. Moskowitz and Tsai (1988) considered a double screening procedure in which two screening variables are observed sequentially. One variable is used first to make one of three decisions – accept, reject, or undecided. After the first screening, the second variable is employed to screen the undecided items.

Economic designs of screening procedures are also considered by several researchers. Boys and Dunsmore (1986) considered a decision theoretic approach using predictive probability based on the constant and linear loss structures. Tang (1987, 1988a) studied economic designs of one- and two-sided screening procedures with three quality cost functions – constant, linear and quadratic – where all parameters are known. Bai et al. (1990) presented economic designs of one- and two-sided screening procedures for the cases where all parameters are known and some parameters unknown. Kim and Bai (1990) considered economic screening procedures in logistic and normal models in which the performance variable is dichotomous and the screening variable is continuous. Tang (1988b) studied an economic model for two-stage screening where the first-stage is based on the screening variable and the second-stage is based on the performance variable. Tang and Schneider (1990) studied economical effects of inspection imprecision on a screening procedure. Kim and Bai (1992) presented an economic design of one-sided screening procedure for the case where all parameters are unknown.

In this paper, we present economic designs of single and double screening procedures based on two screening variables for the case of one-sided specification limit. The optimal cutoff values of single and double screening procedures are obtained under the three quality cost functions by minimizing the total expected cost which includes three cost components; screening inspection cost and costs per accepted and rejected item. The models are described in Section 2 and methods of finding optimal cutoff values are presented in Section 3. In Section 4, a numerical example is given.

2 The Models

Let Y' be the performance variable and X_1' and X_2' be the two screening variables. Assume that (Y', X_1', X_2') are jointly normally distributed with known means (μ_y, μ_1, μ_2), variances $(\sigma_y^2, \sigma_1^2, \sigma_2^2)$ and correlation coefficients $(\rho_{0i}, \rho_{ij}, i, j = 1, 2)$, where ρ_{0i} and ρ_{ij} denote the correlations between Y and X_i, and between X_i and X_j, $i, j = 1, 2$, respectively. The standardized variables $Y = (Y' - \mu_y)/\sigma_y$, $X_1 = (X_1' - \mu_1)/\sigma_1$, $X_2 = (X_2' - \mu_2)/\sigma_2$ jointly have a standard trivariate normal distribution with density function of the form

$$f(y, x_1, x_2) = \frac{|\Sigma|^{1/2}}{(2\pi)^{3/2}} \exp\left[-\frac{1}{2}(y, x_1, x_2)\Sigma^{-1}(y, x_1, x_2)^T \right], \tag{1}$$

where

$$\Sigma = \begin{bmatrix} 1 & \rho_{01} & \rho_{02} \\ \rho_{01} & 1 & \rho_{12} \\ \rho_{02} & \rho_{12} & 1 \end{bmatrix},$$

and $(y, x_1, x_2)^T$ denotes the transpose of (y, x_1, x_2).

Owen et al. (1975) introduced the procedure based on the linear combination of two screening variables, $Z = \lambda_1 X_1 + \lambda_2 X_2$, where constants λ_1 and λ_2 are computed to maximize the proportion of acceptable product after screening. In this paper, constants λ_1 and λ_2 are selected to maximize the correlation between Y and Z and to minimize the variance of $(Y - Z)$. We then obtain

$$\lambda_1 = \frac{\rho_{01} - \rho_{02}\rho_{12}}{1 - \rho_{12}^2}, \qquad \lambda_2 = \frac{\rho_{02} - \rho_{01}\rho_{12}}{1 - \rho_{12}^2}, \tag{2}$$

see Anderson (1984).

Suppose that the performance variable Y is expensive to measure and that there exists a lower specification limit τ; items with $Y \geq \tau$ are conforming and those with $Y < \tau$ are nonconforming. A conforming item gives a positive profit of p and a nonconforming one cause losses according to the deviation of Y from τ. Since larger value of $(\tau - Y)$ causes more consumer's dissatisfaction which results in losses to the producer, the quality cost function $C(y, \tau)$ may be taken to be a nondecreasing function of $(\tau - y)$. We assume that $C(y, \tau) = -p$ for $y \geq \tau$ and consider three functions for $y < \tau$;

$$C(y, \tau) = a \tag{3a}$$

$$= b(\tau - y) \tag{3b}$$

$$= c(\tau - y)^2, \tag{3c}$$

where a, b and c are positive constants. The cost $C(y, \tau)$ of accepting a nonconforming item may include costs of identifying and handling the nonconforming item, service and replacement costs, and loss of goodwill. These quality cost functions are similar to those cited by Tang (1988b). The cost c_r of a rejected item which includes the costs of repairing, scrapping or selling it at a reduced price and the costs c_i of screening an item with X_i are known constants and $c_r \geq c_i$, $i = 1, 2$.

Single and double screening procedures can be described as follows.

1. Single Screening Procedure

   i) Take measurements X_1 and X_2 for each incoming item, and compute

   $$Z = \lambda_1 X_1 + \lambda_2 X_2 \ .$$

   ii) An item will be

   a) rejected if $Z < \delta_1$,
   b) accepted if $Z \geq \delta_1$,

   where δ_1 is the cutoff value to be determined.

2. Double Screening Procedure

   i) Take measurement X_1 for each incoming item.
   ii) An item will be

   a) rejected if $X_1 < L$,
   b) accepted if $X_1 > U$,
   c) undecided if $L \leq X_1 \leq U$.

   iii) Take measurement X_2 for each undecided item in (ii), and compute

   $$Z = \lambda_1 X_1 + \lambda_2 X_2 \ .$$

   iv) An undecided item will be

   a) rejected if $Z < \delta_2$,
   b) accepted if $Z \geq \delta_2$,

   where L, U and δ_2 are the cutoff values to be determined.

2.1 Single Screening Procedure

The joint distribution of Y and Z is bivariate normal with a density function of the form

$$g(y, z) = \frac{1}{2\pi\theta\sqrt{1-\theta^2}} \exp\left[-\frac{y^2 - 2yz + (z/\theta)^2}{2(1-\theta^2)}\right], \tag{4}$$

where it can be shown that $\theta \equiv \sigma_z = \rho_{yz} = \left[\dfrac{\rho_{01}^2 + \rho_{02}^2 - 2\rho_{01}\rho_{02}\rho_{12}}{1 - \rho_{12}^2}\right]^{1/2}$.

Then the expected cost per accepted item is

$$EC_{s_1} = \int_{\delta_1}^{\infty} \int_{-\infty}^{\infty} C(y, \tau) \frac{1}{\theta\sqrt{1-\theta^2}} \phi\left(\frac{y-z}{\sqrt{1-\theta^2}}\right) \phi\left(\frac{z}{\theta}\right) dy\, dz \ , \tag{5}$$

and the expected cost per rejected item is

$$EC_{s_2} = c_r \Phi\left(\frac{\delta_1}{\theta}\right) \ , \tag{6}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and cumulative distribution functions of the standard normal distribution, respectively. The inspection cost is $c_1 + c_2$. Therefore, the total expected cost per item is

$$ETC_s = EC_{s_1} + EC_{s_2} + c_1 + c_2 \ . \tag{7}$$

The optimal cutoff value δ_1^* can be obtained by minimizing ETC_s.

2.2 *Double Screening Procedure*

The expected cost per accepted item is

$$EC_{d_1} = \int_{U}^{\infty} \int_{-\infty}^{\infty} C(y, \tau) \frac{1}{\sqrt{1-\rho_{01}^2}} \phi\left(\frac{y - x_1\rho_{01}}{\sqrt{1-\rho_{01}^2}}\right) \phi(x_1)\, dy\, dx_1$$

$$+ \int_{L}^{U} \int_{(\delta_2 - \lambda_1 x_1)/\lambda_2}^{\infty} \int_{-\infty}^{\infty} C(y, \tau) f(y, x_1, x_2)\, dy\, dx_2\, dx_1 \ . \tag{8}$$

The expected cost per rejected item is

$$EC_{d_2} = c_r \left[\Phi(L) + \int_{L}^{U} \Phi\left(\frac{\delta_2 - x_1\rho_{01}}{\lambda_2\sqrt{1-\rho_{12}^2}}\right) \phi(x_1)\, dx_1 \right] . \tag{9}$$

The expected cost of inspecting an item is

$$EC_{d_3} = c_1 + c_2(\Phi(U) - \Phi(L)) \ . \tag{10}$$

Therefore, the total expected cost per item is

$$ETC_d = EC_{d_1} + EC_{d_2} + EC_{d_3} \ . \tag{11}$$

3 Optimal Solution Procedures

In this section, methods for finding the optimal cutoff values of single and double screening procedures when all parameters are known will be given for the three quality cost functions.

The following identities will be used for the derivations of equations, where $h(y)$ is the normal density function with mean μ_y and variance σ_y^2 and $\Psi(\cdot, \cdot; \rho)$ is the standardized bivariate normal distribution function with correlation coefficient ρ. See, for example, Tang (1988b) and Moskowitz and Tsai (1988).

i) $$\int_{-\infty}^{\zeta} (\tau - y)h(y)\,dy = \sigma_y\left[\frac{\tau - \mu_y}{\sigma_y}\Phi\left(\frac{\zeta - \mu_y}{\sigma_y}\right) + \phi\left(\frac{\zeta - \mu_y}{\sigma_y}\right)\right] . \tag{12}$$

ii) $$\int_{-\infty}^{\zeta} (\tau - y)^2 h(y)\,dy$$

$$= \sigma_y^2\left[\left(1 + \left(\frac{\tau - \mu_y}{\sigma_y}\right)^2\right)\Phi\left(\frac{\zeta - \mu_y}{\sigma_y}\right) + \frac{2\tau - \mu_y - \zeta}{\sigma_y}\phi\left(\frac{\zeta - \mu_y}{\sigma_y}\right)\right] . \tag{13}$$

iii) $$\Psi(x, y; \rho) = \Psi(y, x; \rho) \ , \quad \text{and} \quad \Psi(x, y; \rho) = \Phi(x) - \Psi(x, -y; -\rho) \ . \tag{14}$$

iv) $$\int_{\zeta}^{\infty} \Phi(\alpha_1 + \beta_1 x)\,d\Phi(\alpha_2 + \beta_2 x)$$

$$= \Psi\left(\frac{\alpha_1\beta_2 - \alpha_2\beta_1}{\sqrt{\beta_1^2 + \beta_2^2}}, -\alpha_2 - \beta_2\zeta; \frac{\beta_1}{\sqrt{\beta_1^2 + \beta_2^2}}\right) . \tag{15}$$

3.1 *Single Screening Procedure*

Letting $\frac{dETC_s}{d\delta_1} = 0$ yields

$$\int_{\zeta}^{\infty} C(y, \tau) \frac{1}{\sqrt{1-\theta^2}} \phi\left(\frac{y-\delta_1}{\sqrt{1-\theta^2}}\right) dy = c_r \ . \tag{16}$$

Case 1: constant quality cost function
From equation (16), we obtain

$$\Phi\left(\frac{\tau-\delta_1}{\sqrt{1-\theta^2}}\right) = \frac{c_r + p}{a + p} \ . \tag{17}$$

The left hand side of equation (17) is a strictly decreasing nonnegative function of δ_1. Hence, equation (17) has a unique solution for $c_r \leq a$. If $c_r \leq a$, the optimal cutoff value δ_1^* is

$$\delta_1^* = \tau - \Phi^{-1}\left(\frac{c_r + p}{a + p}\right)\sqrt{1-\theta^2} \ . \tag{18}$$

Case 2: linear quality cost function
From formula (12) and equation (16), we obtain

$$(k + \eta)\Phi(\eta) + \phi(\eta) = \frac{c_r + p}{b\sqrt{1-\theta^2}} \ , \tag{19}$$

where $k = \frac{p}{b\sqrt{1-\theta^2}}$ and $\eta = \frac{\tau - \delta_1}{\sqrt{1-\theta^2}}$. Let the left hand side of equation (19) be $v(\eta)$. $v(\eta)$ is a strictly increasing nonnegative function of η as it can be seen from the integral representation: $v(\eta) = k\Phi(\eta) + \int_{-\infty}^{\eta} (\eta - x)\phi(x)\, dx$. We have $v(-\infty) = 0 < \frac{c_r + p}{b\sqrt{1-\theta^2}} < \infty = v(\infty)$. Therefore equation (19) has a unique solution. A graph for $v(\cdot)$ for selected value of k is given in Figure 1. The value δ_1^* can then be easily determined for given values of τ and $\frac{c_r + p}{b\sqrt{1-\theta^2}}$ as

$$\delta_1^* = \tau - v^{-1}\left(\frac{c_r + p}{b\sqrt{1-\theta^2}}\right)\sqrt{1-\theta^2} \ . \tag{20}$$

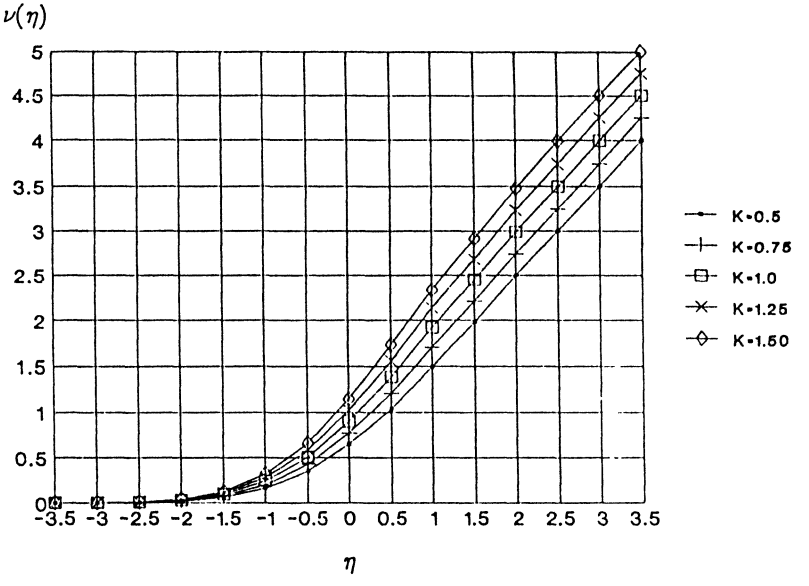

Fig. 1. Graph of $v(\eta)$

Case 3: quadratic quality cost function

Using the formulas (3c) and (13), equation (16) can be rewritten as

$$(1 + m + \eta^2)\Phi(\eta) + \eta\phi(\eta) = \frac{c_r + p}{c(1 - \theta^2)}\ , \tag{21}$$

where $m = \dfrac{p}{c(1 - \theta^2)}$. Let the left hand side of equation (21) be $\chi(\eta)$. Using the integral representation $\chi(\eta) = m\Phi(\eta) + \int_{-\infty}^{\eta}(\eta - x)^2\phi(x)\,dx$, it is easily seen that $\chi(\eta)$ is a strictly increasing nonnegative function of η. We have $\chi(-\infty) = 0 < \dfrac{c_r + p}{c(1 - \theta^2)} < \infty = \chi(\infty)$. Hence, equation (21) has a unique solution. A graph for $\chi(\cdot)$ for selected value of m is given in Figure 2. The value δ_1^* can then be easily determined for given values of τ and $\dfrac{c_r + p}{c(1 - \theta^2)}$ as

$$\delta_1^* = \tau - \chi^{-1}\left(\frac{c_r + p}{c(1 - \theta^2)}\right)\sqrt{1 - \theta^2}\ . \tag{22}$$

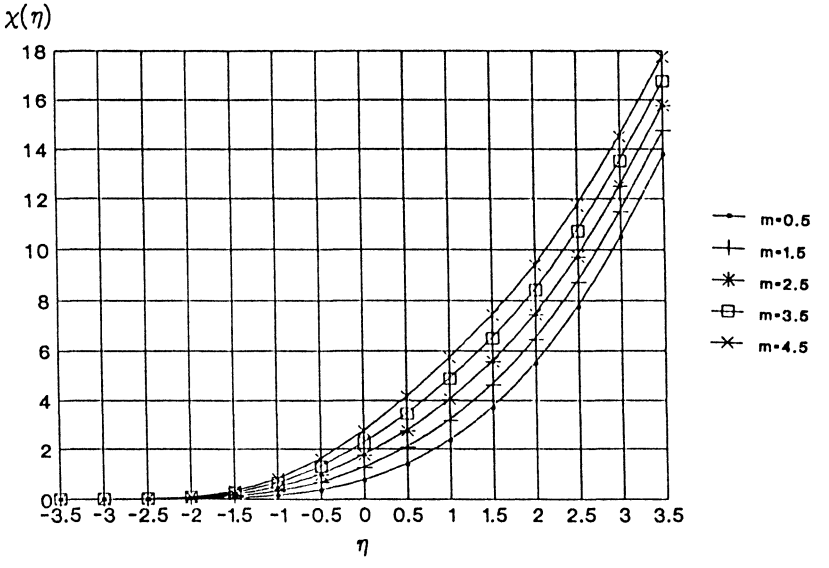

Fig. 2. Graph of $\chi(\eta)$

3.2 *Double Screening Procedure*

Equating the first derivatives of ETC_d with respect to δ_2, L and U zero, respectively, yields

$$\int_{-\infty}^{\infty} C(y,\tau)\frac{1}{\sqrt{1-\theta^2}}\phi\left(\frac{y-\delta_2}{\sqrt{1-\theta^2}}\right)dy = c_r \ , \tag{23}$$

$$\int_{\gamma_1(L)}^{\infty}\int_{-\infty}^{\tau}\frac{C(y,\tau)}{\sqrt{1-\theta^2}\sqrt{1-\rho_{12}^2}}\phi(\gamma_2(y,L))\phi(\gamma_3(L))\,dy\,dx_2$$

$$-\int_{\gamma_1(L)}^{\infty}\int_{\tau}^{\infty}\frac{p}{\sqrt{1-\theta^2}\sqrt{1-\rho_{12}^2}}\phi(\gamma_2(y,L))\phi(\gamma_3(L))\,dy\,dx_2 + c_r\Phi(\gamma_4(L))$$

$$= c_r - c_2 \ , \tag{24}$$

and

$$-\int_{\gamma_1(U)}^{\infty}\int_{-\infty}^{\tau}\frac{C(y,\tau)}{\sqrt{1-\theta^2}\sqrt{1-\rho_{12}^2}}\phi(\gamma_2(y,U))\phi(\gamma_3(U))\,dy\,dx_2$$

$$+\int_{\gamma_1(U)}^{\infty}\int_{\tau}^{\infty}\frac{p}{\sqrt{1-\theta^2}\sqrt{1-\rho_{12}^2}}\phi(\gamma_2(y,U))\phi(\gamma_3(U))\,dy\,dx_2$$

$$+\int_{-\infty}^{\infty}\frac{C(\tau,y)}{\sqrt{1-\rho_{01}^2}}\phi(\gamma_5(y,U))\,dy-c_r\Phi(\gamma_4(U))=c_2\ , \tag{25}$$

where $\gamma_1(L)=\dfrac{\delta_2^*-\lambda_1 L}{\lambda_2}$, $\gamma_2(y,L)=\dfrac{y-\lambda_1 L-\lambda_2 x_2}{\sqrt{1-\theta^2}}$, $\gamma_3(L)=\dfrac{x_2-L\rho_{12}}{\sqrt{1-\rho_{12}^2}}$, $\gamma_4(L)=\dfrac{\delta_2^*-L\rho_{01}}{\lambda_2\sqrt{1-\rho_{12}^2}}$, and $\gamma_5(y,U)=\dfrac{y-U\rho_{01}}{\sqrt{1-\rho_{01}^2}}$. Equation (23) is the same as equation (16). Hence, the optimal cutoff value δ_2^* coincides with δ_1^*.

Case 1: constant quality cost function

Using formula (15), the first term of equation (24) can be rewritten as

$$\int_{\gamma_1(L)}^{\infty}\int_{-\infty}^{\tau}\frac{a}{\sqrt{1-\theta^2}\sqrt{1-\rho_{12}^2}}\phi(\gamma_2(y,L))\phi(\gamma_3(L))\,dy\,dx_2$$

$$=\int_{\gamma_1(L)}^{\infty}\frac{a}{\sqrt{1-\rho_{12}^2}}\Phi(\gamma_2(\tau,L))\phi(\gamma_3(L))\,dx_2$$

$$=\int_{\gamma_1(L)}^{\infty}a\Phi(\gamma_2(\tau,L))d\Phi(\gamma_3(L))$$

$$=a\Psi\left(\frac{\tau-L(\lambda_1+\lambda_2\rho_{12})}{(\lambda_2^2(1-\rho_{12}^2)+1-\theta^2)^{1/2}},-\frac{\delta_2^*-L(\lambda_1+\lambda_2\rho_{12})}{\lambda_2\sqrt{1-\rho_{12}^2}};\right.$$

$$\left.-\frac{\lambda_2(1-\rho_{12}^2)^{1/2}}{(\lambda_2^2(1-\rho_{12}^2)+1-\theta^2)^{1/2}}\right)$$

$$=a\Psi\left(\frac{y-L\rho_{01}}{\sqrt{1-\rho_{01}^2}},-\frac{\delta_2^*-L\rho_{01}}{\lambda_2\sqrt{1-\rho_{12}^2}};\frac{\rho_{01}\rho_{12}-\rho_{02}}{\sqrt{1-\rho_{01}^2}\sqrt{1-\rho_{12}^2}}\right)$$

$$=a\Psi(\gamma_5(\tau,L),-\gamma_4(L);\rho)\ , \tag{26}$$

and using formula (14) the second term of equation (24) as

$$-\int_{\gamma_1(L)}^{\infty}\int_{-\infty}^{\infty}\frac{p}{\sqrt{1-\theta^2}\sqrt{1-\rho_{12}^2}}\phi(\gamma_2(y, L))\phi(\gamma_3(L))\,dy\,dx_2$$

$$+\int_{\gamma_1(L)}^{\infty}\int_{-\infty}^{\tau}\frac{p}{\sqrt{1-\theta^2}\sqrt{1-\rho_{12}^2}}\phi(\gamma_2(y, L))\phi(\gamma_3(L))\,dy\,dx_2$$

$$= -p\{1-\Phi(\gamma_4(L)) - \Psi(\gamma_5(\tau, L), -\gamma_4(L); \rho)\}$$

$$= -p\Psi(-\gamma_5(\tau, L), -\gamma_4(L); -\rho)\ , \tag{27}$$

where $\rho = \dfrac{\rho_{01}\rho_{12}-\rho_{02}}{\sqrt{1-\rho_{01}^2}\sqrt{1-\rho_{12}^2}}$. Equations (24) then simplifies to

$$\Psi(\gamma_5(\tau, L), -\gamma_4(L); \rho) - \frac{p}{a}\Psi(-\gamma_5(\tau, L), -\gamma_4(L); -\rho) + \frac{c_r}{a}\Phi(\gamma_4(L))$$

$$= \frac{1}{a}(c_r - c_2)\ . \tag{28}$$

Equation (25) similarly becomes

$$-\Psi(\gamma_5(\tau, U), -\gamma_4(U); \rho) + \frac{p}{a}\Psi(-\gamma_5(\tau, U), -\gamma_4(U); -\rho) + \Phi(\gamma_5(\tau, U))$$

$$+\frac{p}{a}\Phi(\gamma_5(\tau, U)) - \frac{c_r}{a}\Phi(\gamma_4(U)) = \frac{c_2+p}{a}\ . \tag{29}$$

It is difficult to show analytically that equation (28) or (29) has a unique solution or to find a closed form solution. Let the left hand sides of equations (28) and (29) be $v_1(L)$ and $w_1(U)$, respectively. Numerical studies over a wide range of parameter values of $\left(\frac{c_r}{a}, \frac{p}{a}, \frac{c_2}{a}, \rho_{01}, \rho_{02}, \rho_{12}, \tau\right)$, however, indicate that $v_1(L)$ and $w_1(U)$ are decreasing functions of L and U, respectively. We have $v_1(\infty) = -\dfrac{p}{a} < \dfrac{c_r-c_2}{a} < \dfrac{c_r}{a} = v_1(-\infty)$, and $w_1(\infty) = \dfrac{p}{a}$, $w_1(-\infty) = \dfrac{a+p-c_r}{a}$. Hence, optimal cutoff values L^* can be obtained as unique solution to equation (28) and optimal cutoff values U^* can be obtained as unique solution to equation

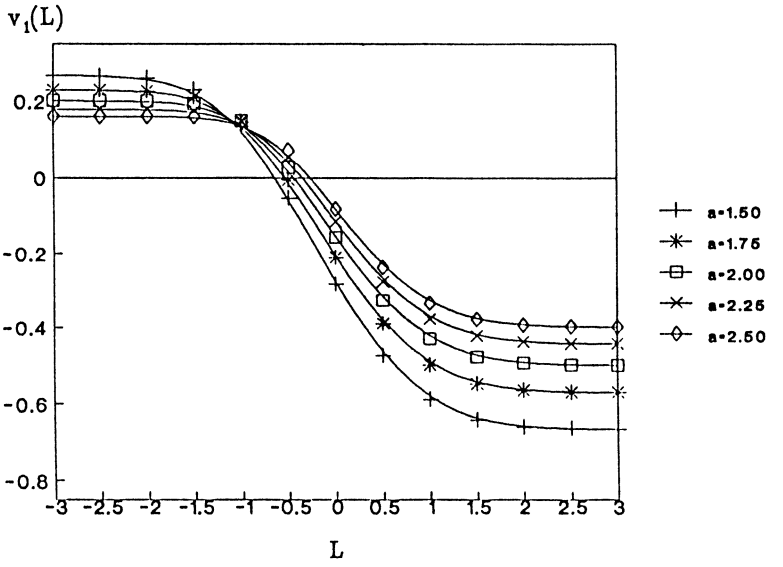

Fig. 3. Graph of $v_1(L)$

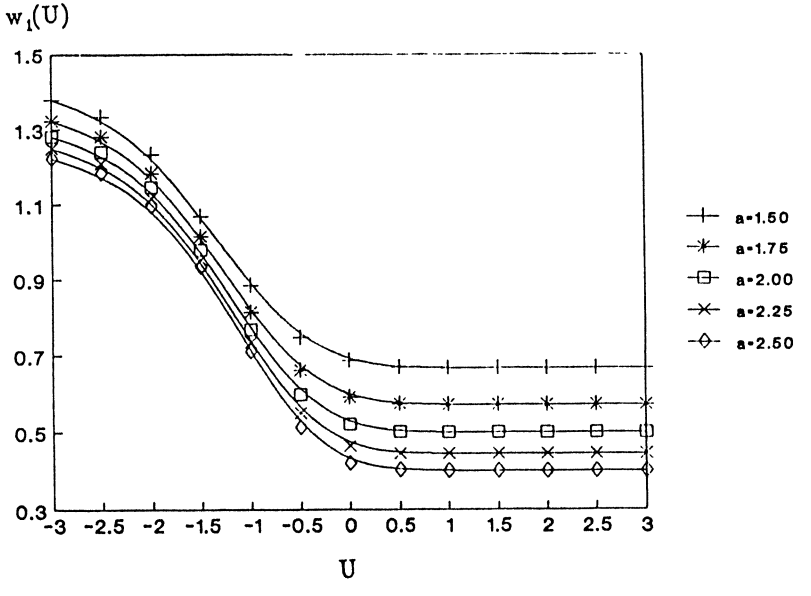

Fig. 4. Graph of $w_1(U)$

(29) for $\frac{c_2 + p}{a} \leq \frac{a + p - c_r}{a}$. Figures 3 and 4 show the graphs of $v_1(L)$ and $w_1(U)$ for $a = 1.5(0.25)2.5$ when $\tau = -0.5$, $p = 1.0$, $c_r = 0.4$, $c_2 = 0.04$, $\rho_{12} = 0.2$, $\rho_{01} = 0.7$, and $\rho_{02} = 0.6$.

Case 2: linear quality cost function

Using formulas (12) and (15), the first term of equation (24) can be rewritten as

$$\int_{\gamma_1(L)}^{\infty} \int_{-\infty}^{\tau} \frac{b(\tau - y)}{\sqrt{1-\theta^2}\sqrt{1-\rho_{12}^2}} \phi(\gamma_2(\tau, L))\phi(\gamma_3(L))\, dy\, dx_2$$

$$= b \int_{\gamma_1(L)}^{\infty} \sqrt{1-\theta^2}\,[\gamma_2(\tau, L)\Phi(\gamma_2(\tau, L)) + \phi(\gamma_2(\tau, L))] \frac{1}{\sqrt{1-\rho_{12}^2}} \phi(\gamma_3(L))\, dx_2$$

$$= b \left\{ \Psi(\gamma_5(\tau, L), -\gamma_4(L); \rho)(\tau - \lambda_1 L) - \int_{\gamma_1(L)}^{\infty} \pi_1(L)\, dx_2 \right\}, \tag{30}$$

where $\pi_1(L) = \{\lambda_2 x_2 \Phi(\gamma_2(\tau, L)) - \sqrt{1-\theta^2}\,\phi(\gamma_2(\tau, L))\} \frac{1}{\sqrt{1-\rho_{12}^2}} \phi(\gamma_3(L))$.

Equation (24) then becomes

$$\Psi(\gamma_5(\tau, L), -\gamma_4(L); \rho)(\tau - \lambda_1 L) - \frac{p}{b}\Psi(-\gamma_5(\tau, L), -\gamma_4(L); -\rho) + \frac{c_r}{b}\Phi(\gamma_4(L))$$

$$- \int_{\gamma_1(L)}^{\infty} \pi_1(L)\, dx_2 = \frac{1}{b}(c_r - c_2) \ . \tag{31}$$

Equation (25) similarly becomes

$$-\Psi(\gamma_5(\tau, U), -\gamma_4(U); \rho)(\tau - \lambda_1 U) + \frac{p}{b}\Psi(-\gamma_5(\tau, U), -\gamma_4(U); -\rho)$$

$$+ \sqrt{1-\rho_{01}^2}\,[\gamma_5(\tau, U)\Phi(\gamma_5(\tau, U)) + \phi(\gamma_5(\tau, U))] + \frac{p}{b}\Phi(\gamma_5(\tau, U))$$

$$- \frac{c_r}{b}\Phi(\gamma_4(U)) + \int_{\gamma_1(U)}^{\infty} \pi_1(U)\, dx_2 = \frac{c_2 + p}{b} \ . \tag{32}$$

Let the left hand sides of equations (31) and (32) be $v_2(L)$ and $w_2(U)$,

respectively. We have $v_2(\infty) = -\frac{p}{b} < \frac{c_r - c_2}{b} < \frac{c_r}{b} = v_2(-\infty)$ and $w_2(\infty) = \frac{p}{b} < \frac{c_2 + p}{b} < \infty = w_2(-\infty)$. As in the previous case, no closed form solutions for equations (31) and (32) can be obtained and a numerical search such as the bisection method can be used to obtain L^* and U^*.

Case 3: quadratic quality cost function

Using formulas (13) and (15), the first term of equation (24) can be rearranged as

$$\int_{\gamma_1(L)}^{\infty} \int_{-\infty}^{\tau} \frac{c(\tau - y)^2}{\sqrt{1-\theta^2}\sqrt{1-\rho_{12}^2}} \phi(\gamma_2(\tau, L))\phi(\gamma_3(L))\, dy\, dx_2$$

$$= c \int_{\gamma_1(L)}^{\infty} (1-\theta^2)[(1 + \gamma_2^2(\tau, L))\Phi(\gamma_2(\tau, L)) + \gamma_2(\tau, L)\phi(\gamma_2(\tau, L))]$$

$$\times \frac{1}{\sqrt{1-\rho_{12}^2}} \phi(\gamma_3(L))\, dx_2$$

$$= c\left\{\Psi(\gamma_5(\tau, L), -\gamma_4(L); \rho)(1 - \theta^2 + (\tau - \lambda_1 L)^2) + \int_{\gamma_1(L)}^{\infty} \pi_2(L)\, dx_2\right\}, \quad (33)$$

where

$$\pi_2(L) = [[(\lambda_2 x_2)^2 - 2(\tau - \lambda_1 L)\lambda_2 x_2]\Phi(\gamma_2(\tau, L))$$

$$+ (1-\theta^2)\gamma_2(\tau, L)\phi(\gamma_2(\tau, L))] \frac{1}{\sqrt{1-\rho_{12}^2}} \phi(\gamma_3(L)) .$$

Equation (24) then becomes

$$\Psi(\gamma_5(\tau, L), -\gamma_4(L); \rho)(1 - \theta^2 + (\tau - \lambda_1 L)^2) - \frac{p}{c}\Psi(-\gamma_5(\tau, L), -\gamma_4(L); -\rho)$$

$$+ \int_{\gamma_1(L)}^{\infty} \pi_2(L)\, dx_2 + \frac{c_r}{c}\Phi(\gamma_4(L)) = \frac{1}{c}(c_r - c_2) , \quad (34)$$

Equation (25) similarly becomes

$$-\Psi(\gamma_5(\tau, U), -\gamma_4(U); \rho)(1-\theta^2+(\tau-\lambda_1 U)^2)+\frac{p}{c}\Psi(-\gamma_5(\tau, U), -\gamma_4(U); -\rho)$$

$$+(1-\rho_{01}^2)\{(1+\gamma_5^2(\tau, U))\Phi(\gamma_5(\tau, U))+\gamma_5(\tau, U)\phi(\gamma_5(\tau, U))\}$$

$$-\int_{\gamma_1(U)}^{\infty}\pi_2(U)\,dx_2+\frac{p}{c}\Phi(\gamma_5(\tau, U))-\frac{c_r}{c}\Phi(\gamma_4(U))=\frac{c_2+p}{c}\,. \tag{35}$$

Let the left hand sides equations (34) and (35) be $v_3(L)$ and $w_3(U)$, respectively. As in the previous cases, we have $v_3(\infty)=-\frac{p}{c}<\frac{c_r-c_2}{c}<\frac{c_r}{c}=v_3(-\infty)$, $w_3(\infty)=\frac{p}{c}<\frac{c_2+p}{c}<\infty=w_3(-\infty)$, and no closed form solutions for equations (34) and (35) can be obtained and L^* and U^* can be found by a numerical search method.

4 Numerical Analysis

In this section, when all parameters are known, a numerical example which originally appeared in Owen et al. (1975) is given to illustrate the optimal screening procedures for the case of known parameters presented in Section 3. Then, based on this example, numerical studies are performed to investiage i) the effects of ρ_{12} on single and double screening procedures, ii) the effects of using a wrong quality cost function on the single and double screening procedures. IMSL (1987) subroutines are used to evaluate statistical distribution functions and integrations.

Example: Suppose the stockpile of some component has an initial proportion 0.7 of acceptable units with regard to some performance variable Y' with a given lower specification limit. A voltage test X_1' and a pressure test X_2' are two screening variables available. The standardized variables Y, X_1 and X_2 jointly have a standard trivariate normal distribution with covariance matrix

$$\Sigma=\begin{bmatrix}1 & \rho_{01} & \rho_{02}\\ \rho_{01} & 1 & \rho_{12}\\ \rho_{02} & \rho_{12} & 1\end{bmatrix}=\begin{bmatrix}1.0 & 0.7 & 0.6\\ 0.7 & 1.0 & 0.2\\ 0.6 & 0.2 & 1.0\end{bmatrix}.$$

Table 1. Optimal ETC_s and ETC_d ($\rho_{12} = 0.2$)

Quality cost function	$\delta_1^* = \delta_2^*$	Single	Double		
		ETC_s	L^*	U^*	ETC_d
Constant	−0.455	−.2400	−1.536	0.252	−.2537
Linear	−0.839	−.4142	−2.037	−0.372	−.4358
Quadratic	−1.460	−.4943	−2.073	−1.002	−.5682

For the three quality cost functions considered in Section 2, it is assumed that $a = 2.0$, $b = 1.56$, $c = 0.73$, $p = 1.0$, $c_r = 0.4$, $c_1 = 0.05$, $c_2 = 0.04$, and the lower specification limit τ for Y is -0.5. The optimal solutions can be obtained from equations (18), (20), or (22) for single screening procedures and equations (23) plus (28)–(29), (31)–(32), or (34)–(35) for double screening procedures, and are summarized in Table 1. Optimal cutoff value δ_1^* ($=\delta_2^*$) can also be obtained using Figure 1: For the case of linear quality cost function, for example, the right hand side of equation (19) is 1.68, and $v^{-1}(1.68) \doteqdot 0.6$ for $k = 1.19$. Hence, δ_1^* is approximately -0.82 from equation (20). One of the major advantages of a double screening procedure is that measurement will not be required on the second screening variable for items accepted or rejected in the first phase of screening. Hence, ETC_d will be less than ETC_s.

i) Effects of ρ_{12}

ETC_s and ETC_d for the above example are given in Table 2 for selected values of ρ_{12}. In Table 2, column B shows the results of employing a double screening procedure, where X_1 is used first and then X_2; column C shows the results of another double screening procedure in which the order of X_1 and X_2 is interchanged. ETC tends to increase as ρ_{12} increases.

ii) Effects of using a wrong quality cost function

Wrong quality cost function will certainly incur economic loss. To study the effects of using a wrong quality cost function, we define the percentage increase of ETC due to use of a wrong quality cost function as

$$PI = \frac{ETC^* - ETC'}{ETC^*} \times 100(\%) ,$$

where ETC^* and ETC' are the expected total cost by using the correct quality cost function and a wrong quality cost function, respectively.

Table 3 shows how sensitive ETC is due to a wrong quality cost function, and indicates that correct use of the quality cost function is very important in designing a screening procedure.

Table 2. Effects of ρ_{12}

(a) Constant cost function

ρ_{12}	$\delta_1^* = \delta_2^*$	A	B			C		
		ETC_s	L^*	U^*	ETC_d	L^*	U^*	ETC_d
.100	−.460	−.2699	−1.732	.439	−.2803	−2.237	.738	−.2770
.150	−.457	−.2539	−1.630	.343	−.2660	−2.134	.642	−.2622
.200	−.455	−.2400	−1.535	.253	−.2537	−2.041	.554	−.2496
.250	−.453	−.2276	−1.447	.168	−.2430	−1.955	.473	−.2385
.300	−.451	−.2165	−1.363	.087	−.2337	−1.875	.397	−.2288

(b) Linear cost function

ρ_{12}	$\delta_1^* = \delta_2^*$	A	B			C		
		ETC_s	L^*	U^*	ETC_d	L^*	U^*	ETC_d
.100	−.842	−.4305	−2.205	−.192	−.4490	−2.778	.001	−.4476
.150	−.841	−.4219	−2.119	−.283	−.4419	−2.693	−.088	−.4406
.200	−.839	−.4142	−2.036	−.371	−.4358	−2.614	−.173	−.4345
.250	−.837	−.4072	−1.954	−.457	−.4303	−2.535	−.256	−.4291
.300	−.834	−.4009	−1.871	−.540	−.4255	−2.456	−.335	−.4242

(c) Quadratic cost function

ρ_{12}	$\delta_1^* = \delta_2^*$	A	B			C		
		ETC_s	L^*	U^*	ETC_d	L^*	U^*	ETC_d
.100	−1.525	−.4994	−2.142	−.883	−.5724	−2.749	−.772	−.5673
.150	−1.490	−.4968	−2.110	−.942	−.5703	−2.727	−.824	−.5659
.200	−1.460	−.4943	−2.073	−1.002	−.5682	−2.698	−.876	−.5644
.250	−1.433	−.4919	−2.031	−1.062	−.5661	−2.663	−.929	−.5629
.300	−1.409	−.4896	−1.984	−1.122	−.5641	−2.623	−.980	−.5614

A: Single Screening Procedure
B: Double Screening Procedure ($X_1 - X_2$ sequence)
C: Double Screening Procedure ($X_2 - X_1$ sequence)

Table 3. Percentage Increase of ETC When Wrong Quality Cost Functions are used

Used / True	Single Screening Procedure			Double Screening Procedure		
	Constant	Linear	Quadratic	Constant	Linear	Quadratic
Constant	0	22.12	79.58	0	53.01	66.89
Linear	10.88	0	12.98	11.15	0	10.05
Quadratic	18.67	1.52	0	21.48	2.09	0

5 Concluding Remarks

We have developed economic designs of single and double screening procedures for one-sided specification limit based on two screening variables. Cost models are constructed under the assumption that the performance variable and the two screening variables are jointly normally distributed, and methods of finding the optimal cutoff values are presented. For the case of single screening procedure, the solutions are shown to be unique. For the case of double screening procedure, it is difficult to show analytically that the solutions are optimal. Numerical studies over wide range of parameter values, however, indicate that ETC_d is indeed unimodal. Numerical results show that ETC tends to increase as ρ_{12} increases, and correct use of quality cost function is very important in designing a screening procedure. A major advantage of a double screening procedure is that it can save inspection cost at the second phase of screening. A possible area of further investigation would be the extension of the model to the cases of two-sided specification limits and to the case of unknown parameters.

References

Anderson TW (1984) An Introduction to Multivariate Statistical Analysis, 2nd Ed., John Wiley & Sons, Inc., New York

Bai DS, Kim SB, Riew MC (1990) Economic Screening Procedures Based on Correlated Variables. Metrika 37:263–280

Boys RJ, Dunsmore IR (1986) Screening in a Normal Model. Journal of Royal Statistical Society – Series B 48:60–69

International Mathematical and Statistical Libraries, Inc. (1987) IMSL Library: Reference manual, Houston

Kim SB, Bai DS (1990) Economic Screening Procedures in Logistic and Normal Models. Naval Research Logistics 37:919–928

Kim SB, Bai DS (1992) Economic Design of One-Sided Screening Procedures Based on a Correlated Variable with All Parameters Unknown. Metrika 39:85–93

Madsen RW (1982) A Selection Procedure Using a Screening Variate. Technometrics 24:301–306

Menzefricke U (1984) A Decision-Theoretic Approach to Some Screening Problems. Annals of Institute of Statistical Mathematics, Part A 36:485–497

Moskowitz H, Tsai HT (1988) A One-Sided Double Screening Procedure Using Individual Unit Misclassification Error. Management Science 34:1139–1153

Owen DB, Boddie JW (1976) A Screening Method for Increasing Acceptable Product with Some Parameters Unknown. Technometrics 18:195–199

Owen DB, Li L, Chou YM (1981) Prediction Intervals for Screening Using a Measured Correlated Variate. Technometrics 23:165–170

Owen DB, McIntire D, Seymour E (1975) Tables Using One or Two Screening Variables to Increase Acceptable Product under One-Sided Specifications. Journal of Quality Technology 7:127–138

Owen DB, Su YH (1977) Screening Based on Normal Variables. Technometrics 19:65–68

Tang K (1987) Economic Design of One-Sided Screening Procedure Using a Correlated Variable. Technometrics 29:477–485

Tang K (1988a) Economic Design of Two-Sided Screening Procedure Using a Correlated Variable. Applied Statistics 37:231–241

Tang K (1988b) Design of Two-Stage Screening Procedure Using Correlated Variables: A Loss Function Approach. Naval Research Logistics 35:513–533

Tang J, Tang K (1989) A Two-Sided Screening Procedure Using Several Correlated Variables. IIE Transactions 21:333–336

Tang K, Schneider H (1990) Cost Effectiveness of Using a Correlated Variable in a Complete Inspection Plan When Inspection Error is Present. Naval Research Logistics 37:893–904

Tsai HL, Moskowitz H (1980) A Single Screening Procedure Using Individual Misclassification Error Under One-Sided Specification. Institute Paper No 896, Krannert Graduate School of Management, Purdue University

Received 27.08.1991
Revised version 21.01.1992

Book Review

T. P. Hutchinson, C. D. Lai: Continuous Bivariate Distributions, Emphasising Applications. Rumsby Scientific Publishing, Adelaide, 1990, AUS$88.00/ UK£38.00

To begin with, this book is not a textbook but a monograph on statistical distributions. Their application is illustrated in a wealth of examples from published research work – certainly the most striking feature.

The text is organised into four parts. The central part II (after an introductory part I) contains an extensive collection of continuous bivariate distributions including the bivariate normal and related distributions and distributions arising in reliability. While formulae and properties are given, theoretical niceties (such as characterisation theorems) are suppressed. Part III presents concepts and measures of association, methods for constructing bivariate distributions and generating random variables, and some issues in reliability. This part ranges from the discussion of basic facts (e.g. properties of correlation coefficients) to highly specialised investigations of reliability classes.

Real data examples are scattered throughout the text. They predominate in part IV. (Together with applications the reader is introduced to probit analysis and signal detection theory.) Research areas as diverse as engineering, meteorology, biology, medicine, economics, and social sciences are covered. Sometimes a fairly detailed account is given – for example when modelling injury severity in road accidents. More often only brief mention is made of a topic. The reader is then referred to the list of references, which is huge.

In short, the book is a valuable source for statistical practitioners and research workers interested in modelling data.

Munich E. Eberle

Metrika (1993) 40:115–120

Metrika

The Lattice Structure of Nonlinear Congruential Pseudorandom Numbers

JÜRGEN EICHENAUER-HERRMANN

Fachbereich Mathematik, Technische Hochschule Darmstadt, Schloßgartenstr. 7, 6100 Darmstadt, Germany

Abstract: Several known deficiencies of the classical linear congruential method for generating uniform pseudorandom numbers led to the development of nonlinear congruential pseudorandom number generators. In the present paper a general class of nonlinear congruential methods with prime power modulus is considered. It is proved that these generators show certain undesirable linear structures, too, which stem from the composite nature of the modulus.

1 Introduction

The linear congruential method for generating uniform pseudorandom numbers in the interval [0, 1) shows a lot of undesirable regularities which stem from the linearity of the underlying recursion. This was the motivation for recent work on nonliner congruential methods·. A review of the development of this area is given in the survey articles of Niederreiter (1991) and of Eichenauer-Herrmann (1992b) and in the excellent monograph of Niederreiter (1992).

In the present paper a general class of nonlinear congruential generators with prime power modulus is analysed. It is shown that tuples of successive pseudorandom numbers generated by any of these nonlinear congruential methods form a superposition of shifted lattices, so-called grids. Hence, these nonlinear congruential pseudorandom numbers still show a rather strong linear structure.

The second section of the present paper contains the precise statement and a short discussion of the main result. Its proof is given in the fourth section. In the third section the result is applied to quadratic and inversive congruential generators which yields known as well as new results on their lattice structure.

0026–1335/93/2/115–120 $2.50

2 Lattice Structure

Let $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ for positive integers n and $m = p^\omega$ for a prime p and an integer $\omega \geq 2$. Let $f\colon \mathbb{Z} \to \mathbb{Z}$ with

$$f(x) = a_s x^s + \cdots + a_1 x + a_0$$

be an integer polynomial. In order to avoid trivial cases it is always assumed that $s \geq 2$ and $\alpha \leq \omega - 1$, where the integer α is defined by $p^\alpha = gcd(m, a_2, \ldots, a_s)$. A *nonlinear congruential sequence* $(x_n)_{n \geq 0}$ of elements of $\mathbb{Z}_m$ is defined by

$$x_{n+1} \equiv f(x_n) \pmod{m} , \qquad n \geq 0 .$$

Nonlinear congruential pseudorandom numbers in the interval $[0, 1)$ are obtained by the normalization x_n/m for $n \geq 0$. In the following it is always assumed that the nonlinear congruential sequence $(x_n)_{n \geq 0}$ is purely periodic. Its period length is denoted by $\Lambda = \min\{n \geq 1 | x_n = x_0\}$. Let $\lambda = \min\{n \geq 1 | x_n \equiv x_0 \pmod{p}\}$ be the period length of the reduced sequence $(x_n \pmod{p})_{n \geq 0}$. It is always assumed that the nonlinear congruential sequence $(x_n)_{n \geq 0}$ has maximal period length in the sense that $\Lambda = \lambda p^{\omega - 1}$. Then the set of generated numbers $X_\omega = \{x_n | 0 \leq n < \Lambda\}$ is given by

$$X_\omega = \{x \in \mathbb{Z}_m | x \pmod{p} \in X_1\} ,$$

where

$$X_1 = \{x_n \pmod{p} \in \mathbb{Z}_p | 0 \leq n < \lambda\}$$

denotes the set of generated residues modulo p. Now, define functions f_n, $g_n\colon \mathbb{Z} \to \mathbb{Z}_m$ by

$$f_1(x) \equiv f(x) \pmod{m} ,$$

$$f_n(x) = f_1(f_{n-1}(x)) , \qquad n \geq 2 ,$$

$$g_1(x) \equiv f'(x) \pmod{m} , \qquad \text{and}$$

$$g_n(x) \equiv g_1(f_{n-1}(x)) g_{n-1}(x) \pmod{m} , \qquad n \geq 2 ,$$

where f' stands for the derivative of the polynomial f. For an integer $d \geq 2$ let

$$V_d = \{(x_n, x_{n+1}, \ldots, x_{n+d-1}) \in \mathbb{Z}_m^d | 0 \leq n < \Lambda\}$$

$$= \{(x, f_1(x), \ldots, f_{d-1}(x)) \in \mathbb{Z}_m^d | x \in X_\omega\}$$

be the set of d-tuples of successive numbers in a nonlinear congruential sequence. Subsets $V_d(z)$ of V_d are defined by

$$V_d(z) = \{(x, f_1(x), \ldots, f_{d-1}(x)) \in \mathbb{Z}_m^d | x \in X_\omega, x \equiv z \pmod{p^\gamma}\}$$

for $z \in X_\gamma = \{x \in Z_{p^\gamma} | x \pmod p \in X_1\}$ with $\gamma = [(\omega - \alpha + 1)/2]$. Let

$$G_d(z) = \{v + mu \in \mathbb{Z}^d | v \in V_d(z), u \in \mathbb{Z}^d\}$$

be the m-periodic continuation of $V_d(z)$ for $z \in X_\gamma$. Now, the main result can be stated precisely.

Theorem: Let $z \in X_\gamma$. Then the set $G_d(z)$ is a grid with shift vector

$$w_0(z) = (z, f_1(z), \ldots, f_{d-1}(z))$$

and basis

$$w_1(z) = p^\gamma(1, g_1(z), \ldots, g_{d-1}(z)) ,$$

$$w_2(z) = (0, m, 0, \ldots, 0) ,$$

$$\vdots$$

$$w_d(z) = (0, 0, \ldots, 0, m) .$$

The m-periodic continuation of V_d, the set

$$G_d = \{v + mu \in \mathbb{Z}^d | v \in V_d, u \in \mathbb{Z}^d\} ,$$

is the disjoint union of the sets $G_d(z)$ for $z \in X_\gamma$. Hence, G_d is a superposition of the $\lambda p^{\gamma-1}$ grids $G_d(z)$ for $z \in X_\gamma$. This undesirable lattice structure of the corresponding nonlinear congruential pseudorandom numbers can render them useless for certain stochastic simulations. The proof of the Theorem in the fourth section shows that these unfavourable regularities are due to the composite nature of the modulus. A substantially better behaviour with respect to linear structures is obtained by nonlinear congruential methods with prime modulus (cf. Eichenauer et al. 1988a, Niederreiter 1988, and Eichenauer-Herrmann 1991).

3 Applications

The first nonlinear congruential method for generating uniform pseudorandom numbers is the *quadratic congruential method* which was suggested by Knuth (1981). This approach fits into the present framework when the polynomial f is defined by $f(x) = ax^2 + bx + c$. The integers a, b, c can be chosen such that the maximal period length $\Lambda = m$ is obtained (cf. Knuth 1981, p. 34, and Eichenauer and Lehn 1987). The conditions for maximal period length imply that $a = 0 \pmod p$ and hence $\alpha \geq 1$. Then the Theorem corresponds to the main result of Eichenauer and Lehn (1987).

The choice $m = 2^\omega$ with $\omega \geq 3$ and $f(x) = ax^\nu + b$ with $\nu = 2^{\omega-2} - 1$ yields the *inversive congruential method with power of two modulus*, since $f(x) \equiv ax^{-1} + b \pmod m$ for all odd integers x. This approach is due to Eichenauer et al. (1988b), where it is shown that the maximal period length $\Lambda = 2^{\omega-1}$ is obtained exactly for $a \equiv 1 \pmod 4$ and $b \equiv 2 \pmod 4$ which implies that $\alpha = 0$ and $\gamma = [(\omega + 1)/2]$. Then the Theorem corresponds to the main result of Eichenauer-Herrmann et al. (1990).

The *inversive congruential method with odd prime power modulus* is obtained for $m = p^\omega$ with $p \geq 3$ and $f(x) = ax^\nu + b$ with $\nu = (p-1)p^{\omega-1} - 1$, since $f(x) \equiv ax^{-1} + b \pmod m$ for all integers $x \not\equiv 0 \pmod p$. This nonlinear congruential method was introduced in Eichenauer-Herrmann and Topuzoğlu (1990). The integers a and b can be chosen such that the maximal period length $\Lambda = \lambda p^{\omega-1}$ with $\lambda \leq (p+1)/2$ is obtained (cf. Eichenauer-Herrmann and Topuzoğlu 1990 and Eichenauer-Herrmann 1992a). The corresponding conditions imply that $a \not\equiv 0 \pmod p$ and hence $\alpha = 0$ and $\gamma = [(\omega + 1)/2]$. Now, the Theorem provides a new result on the lattice structure of these inversive congruential generators.

4 Proof of the Main Result

The proof of the Theorem is based on the following auxiliary result.

Lemma: Let z be an integer. Then

$$f_n(x) \equiv g_n(z)(x - z) + f_n(z) \pmod{m}$$

for all integers $x \equiv z \pmod{p^\gamma}$ and $n \geq 1$.

Proof: The lemma is proved by induction on n. First, let $n = 1$. Any integer $x \equiv z \pmod{p^\gamma}$ can be written in the form $x = yp^\gamma + z$ with some integer y. Since $\alpha + 2\gamma \geq \omega$ and $a_k \equiv 0 \pmod{p^\alpha}$ for $2 \leq k \leq s$, it follows at once that

$$\begin{aligned} f_1(x) &\equiv f(yp^\gamma + z) \equiv \sum_{k=0}^{s} a_k(yp^\gamma + z)^k \\ &\equiv \sum_{k=0}^{s} a_k(kz^{k-1}yp^\gamma + z^k) \\ &\equiv \left(\sum_{k=1}^{s} a_k k z^{k-1}\right)(x - z) + \sum_{k=0}^{s} a_k z^k \\ &\equiv f'(z)(x - z) + f(z) \\ &\equiv g_1(z)(x - z) + f_1(z) \pmod{m} \ . \end{aligned}$$

Now, assume that the assertion is valid for some integer $n \geq 1$. Let $x \equiv z \pmod{p^\gamma}$. Then $f_n(x) \equiv f_n(z) \pmod{p^\gamma}$ and therefore

$$\begin{aligned} f_{n+1}(x) &\equiv f_1(f_n(x)) \\ &\equiv g_1(f_n(z))(f_n(x) - f_n(z)) + f_1(f_n(z)) \\ &\equiv g_1(f_n(z))g_n(z)(x - z) + f_1(f_n(z)) \\ &\equiv g_{n+1}(z)(x - z) + f_{n+1}(z) \pmod{m} \end{aligned}$$

which completes the proof. □

Now, the Theorem follows at once from the Lemma by the same arguments as in the proof of the main result in Eichenauer and Lehn (1987).

References

Eichenauer J, Lehn J (1987) On the structure of quadratic congruential sequences, manuscripta math. 58:129–140

Eichenauer J, Grothe H, Lehn J (1988a) Marsaglia's lattice test and non-linear congruential pseudo random number generators, Metrika 35:241–250

Eichenauer J, Lehn J, Topuzoğlu A (1988b) A nonlinear congruential pseudorandom number generator with power of two modulus, Math. Comp. 51:757–759

Eichenauer-Herrmann J (1991) Inversive congruential pseudorandom numbers avoid the planes, Math. Comp. 56:297–301

Eichenauer-Herrmann J (1992a) Construction of inversive congruential pseudorandom number generators with maximal period length, J. Comp. Appl. Math. 40:345–349

Eichenauer-Herrmann J (1992b) Inversive congruential pseudorandom numbers: a tutorial, Int. Statist. Rev. 60:167–176

Eichenauer-Herrmann J, Topuzoğlu A (1990) On the period length of congruential pseudorandom number sequences generated by inversions, J. Comp. Appl. Math. 31:87–96

Eichenauer-Herrmann J, Grothe H, Niederreiter H, Topuzoğlu A (1990) On the lattice structure of a nonlinear generator with modulus 2^α, J. Comp. Appl. Math. 31:81–85

Knuth DE (1981) The art of computer programming, vol. 2, Addison-Wesley, Reading, Mass., 2nd ed

Niederreiter H (1988) Remarks on nonlinear congruential pseudorandom numbers, Metrika 35: 321–328

Niederreiter H (1991) Recent trends in random number and random vector generation, Ann. Operations Res. 31:323–346

Niederreiter H (1992) Random number generation and quasi-Monte Carlo methods, SIAM, Philadelphia

Received 01.04.1992

Metrika (1993) 40:121–128

Metrika

A Note on Three-Stage Confidence Intervals for the Difference of Locations: The Exponential Case

N. MUKHOPADHYAY

Department of Statistics, U Box 120, University of Connecticut, 196 Auditorium Road, Storrs CT 06269, USA

A. R. PADMANABHAN

Department of Mathematics, Monash University, Clayton, Victoria 3168, Australia

Abstract: Fixed-width confidence intervals for the difference of location parameters of two independent negative exponential distributions are constructed via triple sampling when the scale parameters are unknown and unequal. The present three-stage estimation methodology is put forth because (i) it is operationally more convenient than the existing purely sequential counterpart, and (ii) the three-stage and the purely sequential estimation techniques have fairly similar asymptotic second-order characteristics.

Key Words and Phrases: Fixed-width intervals; Behrens-Fisher situation; two-parameter exponentials; triple sampling; second-order expansions.

1 Introduction

Let $\{X_{i1}, X_{i2}, \ldots\}$, $i = 1, 2$ be two independent sequences of random variables where we assume that $X_{i1}, X_{i2}, \ldots$ are independent and identically distributed (i.i.d.) having the probability density function $f(x; \mu_i, \sigma_i) = \sigma_i^{-1} \exp\{-(x - \mu_i)/\sigma_i\} I(x > \mu_i)$, $0 < \sigma_i < \infty$, $-\infty < \mu_i < \infty$, $i = 1, 2$. Here and elsewhere, $I(\cdot)$ stands for the indicator function of $(\cdot)$. The particular type of distribution considered here has been used widely in many reliability and life testing experiments to describe the failure times of complex equipment, vacuum tubes etc. This distribution has also been suggested as a statistical model in several clinical trials, such as studies of behavior of tumor systems in animals and analysis of survival data in cancer research. One is referred to Zelen (1966). One is also referred to Mukhopadhyay (1988) for other citations.

The parameters μ_1, μ_2, when positive, may be interpreted as the minimum guarantee times or the thresholds of the distributions. The parameters σ_1, σ_2 are known as the scales of the distributions. We assume that all four parameters are

0026–1335/93/2/121–128 $2.50

unknown and our goal is to construct fixed-width confidence intervals for the parameter $\delta = \mu_1 - \mu_2$.

Mukhopadhyay and Hamdy (1984) proposed a two-stage procedure that came with the guarantee of the nominal level for the probability of coverage of δ by means of the constructed fixed-width confidence interval. On the average, however, the two-stage procedures have a tendency to oversample. Hence, in the same paper of Mukhopadhyay and Hamdy (1984), they had also proposed purely sequential procedures in order to achieve "efficiency". All these techniques were proposed and studied when σ_1, σ_2 are unknown and unequal. In the case when $\sigma_1 = \sigma_2 = \sigma$, say, but σ is unknown, Mukhopadhyay and Mauromoustakos (1987) had proposed a three-stage fixed-width confidence interval for δ, motivated by the works of Hall (1981).

In the present note, we design three-stage sampling procedures for constructing fixed-width confidence intervals for δ when σ_1, σ_2 are unknown and unequal. Section 2 gives some of the preliminaries and the main result (Theorem 1) that sets the rate of convergence of the difference between the achieved confidence level and the target value. The proof of Theorem 1 is included in Section 3.

2 Formulation and Preliminaries

Having recorded $X_{i1}, \ldots, X_{in_i}$ we write $X_{in_i(1)} = \min\{X_{i1}, \ldots, X_{in_i}\}$ and $U_{in_i} = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (X_{ij} - X_{in_i(1)})$ for $n_i \geq 2$ which respectively estimates μ_i and σ_i, $i = 1, 2$. Let $\underset{\sim}{n} = (n_1, n_2)$ and given a preassigned number $d(> 0)$ we propose the fixed-width confidence interval

$$J(\underset{\sim}{n}) = [X_{1n_1(1)} - X_{2n_2(1)} \pm d] \tag{2.1}$$

for the parameter $\delta(= \mu_1 - \mu_2)$. We are also given a preassigned number $\alpha \in (0, 1)$ and we wish to conclude that $P\{\delta \in J(\underset{\sim}{n})\} \geq 1 - \alpha$ for all σ_1, σ_2. Now, with $a = \ln(1/\alpha)$, $n_i \geq a\sigma_i/d = C_i$, say, $P\{\delta \in J(\underset{\sim}{n})\}$ can be shown to be at least $(1 - \alpha)$. But C_1, C_2 are unknown and hence certain multistage procedures are in order. Mukhopadhyay and Hamdy (1984) considered "efficient" purely sequential sampling techniques. At present, we consider proposing three-stage estimation procedures so that these will be operationally more convenient than the purely sequential counterpart, and yet these three-stage methodologies and the purely sequential ones will have very similar asymptotic characteristics. Throughout, asymptotics are carried out as $d \to 0$.

2.1 *A Three-Stage Methodology*

One starts with $X_{i1}, \ldots, X_{im}$ where the starting sample size $m(\geq 2)$ is such that $m = O(d^{-1/r})$ for some $r > 1$. We also choose and fix two numbers, $0 < \rho_1, \rho_2 < 1$. Now, let

$$T_i = T_i(d) = \max\{m, \langle a\rho_i U_{im} d^{-1} \rangle + 1\} \ , \tag{2.2}$$

$$N_i = N_i(d) = \max\{T_i, \langle aU_{iT_i} d^{-1} + \varepsilon_i \rangle + 1\} \ , \tag{2.3}$$

where $\langle x \rangle$ stands for the largest integer $< x$ and $\varepsilon_1, \varepsilon_2$ are *known* real numbers which are to be defined precisely in the sequel, $i = 1, 2$. Let us write $\underset{\sim}{N} = (N_1, N_2)$.

The three-stage methodology (2.2)–(2.3) is implemented as follows. Based on $X_{i1}, \ldots, X_{im}$, one determines T_i which estimates $\rho_i C_i$, a fraction of C_i. If $T_i = m$, then one does not take any more samples from X_i in the second stage, however if $T_i > m$, then one samples the difference $(T_i - m)$ from X_i in the second stage arriving at $X_{i1}, \ldots, X_{iT_i}$. Now, one determines N_i based on $X_{i1}, \ldots, X_{iT_i}$. Notice that N_i estimates C_i. If $N_i = T_i$, then one does not take any more samples from X_i in the third stage, however if $N_i > T_i$, then one samples the difference $(N_i - T_i)$ from X_i in the third stage, finally arriving at $X_{i1}, \ldots, X_{im}, \ldots, X_{iT_i}, \ldots, X_{iN_i}$. We then propose the fixed-width confidence interval $J(\underset{\sim}{N})$, defined in (2.1), for δ.

2.2 *Preliminary Derivations*

Notice that $I(\underset{\sim}{N} = \underset{\sim}{n})$ is independent of $(X_{1n_1(1)}, X_{2n_2(1)})$ for all fixed $\underset{\sim}{n}$, and hence

$$P\{\delta \in J(\underset{\sim}{N})\} = E[g(N_1/C_1, N_2/C_2)] \tag{2.4}$$

where

$$g(x, y) = 1 - [x^{-1} + y^{-1}]^{-1}[x^{-1}e^{-ax} + y^{-1}e^{-ay}] \ , \tag{2.5}$$

for $x > 0$, $y > 0$. It follows easily that $N_i \to \infty$ a.s., $N_i/C_i \to 1$ a.s. and hence $P\{\delta \in J(\underset{\sim}{N})\} \to 1 - e^{-a} = 1 - \alpha$ as $d \to 0$, that is, the three-stage fixed-width confidence interval procedure is asymptotically consistent in the Chow-Robbins (1965) sense. Our aim is to show that the difference, $P\{\delta \in J(\underset{\sim}{N})\} - (1 - \alpha)$, can be made $o(d)$ if one chooses the "fine-tuning" factors ε_1 and ε_2 in (2.3) appropriately.

The three-stage procedure (2.2)–(2.3), without the "fine-tuning" factors ε_i, is of the form considered in general in (2.1)–(2.2) of Mukopadhyay (1990) where $\gamma = \rho_i$, $\lambda = a/d$, $\theta = \sigma_i$ and Y_{ij}'s are i.i.d. $\frac{1}{2}\sigma_i\chi^2_2$. Hence Mukhopadhyay's (1990) Theorem 2 implies

$$E(N_i) = C_i + (\eta_i + \varepsilon_i) + o(d) \ , \tag{2.6}$$

where $\eta_i = \frac{1}{2} - \rho_i^{-1}$, $i = 1, 2$. Also, Theorem 3(i) of Mukhopadhyay (1990) shows that

$$P(N_i \leq \xi C_i) = O(C_i^{-Lm}) \ , \tag{2.7}$$

for arbitrarily large, but fixed, $L(> 0)$. Let us write $N_i^* = C_i^{-1/2}(N_i - C_i)$, and then Mukhopadhyay's (1990) Theorem 3(ii) shows that

$$N_i^* \xrightarrow{\mathscr{L}} N(0, p_i) \ , \tag{2.8}$$

where $p_i = \rho_i^{-1}$. Also, one will observe that

$$|N_i^*|^s \text{ is uniformly integrable, } s = 1, 2, 3 \ . \tag{2.9}$$

2.3 *Main Result*

Theorem 1: In the three-stage estimation methodology (2.2)–(2.3), *suppose that one chooses* $\varepsilon_i = \frac{1}{2}(a + 3 - \rho_i)/\rho_i$, $i = 1, 2$. *Then, we have as* $d \to 0$,

$$P\{\delta \in J(\underset{\sim}{N})\} = (1 - \alpha) + o(d) \ .$$

In other words, $P\{\delta \in J(\underset{\sim}{N})\}$ is approximately $(1 - \alpha)$ up to the order $o(d)$. Note that (2.6) already gives the asymptotic second-order expansion of $E(N_i - C_i)$.

3 Proof of Theorem 1

We expand $g(x, y)$ given in (2.5) around $(x, y) = (1, 1)$ and hence for some random variables W_i between 1 and N_i/C_i, $i = 1, 2$, one obtains from (2.4),

$$
\begin{aligned}
P\{\delta \in J(\underset{\sim}{N})\} &= (1 - e^{-a}) + \frac{1}{2}de^{-a}\sum_{i=1}^{2}\sigma_i^{-1}E(N_i - C_i) \\
&- \frac{1}{4}de^{-a}(1 + a)\sum_{i=1}^{2}\sigma_i^{-1}E(N_i^{*2}) + \frac{1}{2}de^{-a}(\sigma_1\sigma_2)^{-1}E(N_1^*N_2^*) \\
&+ E\left[\frac{1}{6}d^{3/2}a^{-3/2}\sum_{i=1}^{2}\sigma_i^{-3/2}e^{-aW_i}(a^{-1}V^{-4}W_i^{-4} + 4a^{-2}V^{-4}W_i^{-5}\right. \\
&\left. - 4a^{-3}V^{-5}W_i^{-6})N_i^{*3}\right] + E\left[\frac{1}{6}d^{3/2}a^{-3/2}\sum_{1\le i\ne j\le 2}\sigma_i^{-3/2}e^{-aW_i}\right. \\
&\{V^{-3}(2a^{-2}W_i^{-3}W_j^{-2} + 6a^{-3}W_i^{-4}W_j^{-2} - 6a^{-3}W_i^{-2}W_j^{-4}) \\
&+ V^{-4}(3a^{-1}W_i^{-3}W_j^{-1} + 3a^{-1}W_i^{-2}W_j^{-2} + a^{-1}W_i^{-1}W_j^{-3} \\
&+ 11a^{-2}W_i^{-4}W_j^{-1} + 3a^{-2}W_i^{-2}W_j^{-3} + 10a^{-2}W_i^{-3}W_j^{-2} \\
&+ 8a^{-3}W_i^{-5}W_j^{-1} + 4a^{-3}W_i^{-3}W_j^{-3} + 12a^{-1}W_i^{-4}W_j^{-2} \\
&+ 6a^{-4}W_i^{-2}W_j^{-5} - 6a^{-4}W_i^{-5}W_j^{-2}) + V^{-5}(12a^{-3}W_i^{-5}W_j^{-1} \\
&+ 12a^{-3}W_i^{-4}W_j^{-2} + 4a^{-3}W_i^{-3}W_j^{-3} + 8a^{-4}W_i^{-6}W_j^{-1} \\
&\left. + 8a^{-4}W_i^{-4}W_j^{-3} + 16a^{-4}W_i^{-5}W_j^{-2})\}N_i^{*3}\right] \\
&+ E\left[\frac{1}{6}d^{3/2}a^{-3/2}\sum_{1\le i\ne j\le 2}\sigma_i^{-1}\sigma_j^{-1/2}e^{-aW_j}(6a^{-4}V^{-4}W_i^{-3}W_j^{-4}\right. \\
&\left. - 2a^{-2}V^{-3}W_i^{-3}W_j^{-2} - 4a^{-3}V^{-3}W_i^{-3}W_j^{-3})N_i^{*2}N_j^*\right] \\
&+ E\left[\frac{1}{6}d^{3/2}a^{-3/2}\sum_{1\le i\ne j\le 2}\sigma_i^{-1}\sigma_j^{-1/2}e^{-aW_i}\{V^{-4}(3a^{-2}W_i^{-3}W_j^{-2}\right. \\
&+ 6a^{-2}W_i^{-3}W_j^{-3} + 3a^{-2}W_i^{-1}W_j^{-4} + 2a^{-3}W_i^{-4}W_j^{-2} + 6a^{-3}W_i^{-1}W_j^{-4}
\end{aligned}
$$

$$+ 8a^{-3}W_i^{-3}W_j^{-3} - 6a^{-4}W_i^{-3}W_j^{-4}) + 4a^{-1}V^{-3}W_i^{-2}W_j^{-3}$$

$$- V^{-5}(4a^{-3}W_i^{-4}W_j^{-2} + 12a^{-3}W_i^{-3}W_j^{-3} + 12a^{-3}W_i^{-2}W_j^{-4}$$

$$+ 4a^{-3}W_i^{-1}W_j^{-5} + 8a^{-4}W_i^{-4}W_j^{-3} + 8a^{-4}W_i^{-2}W_j^{-5}$$

$$+ 4a^{-4}W_i^{-3}W_j^{-4})\}N_i^{*2}N_j^{*}\Big]$$

$$= (1-\alpha) + E(R_1 + R_2) - E(R_3 + R_4) + E(R_5) + E(R_6 + R_7)$$

$$+ E\Big[\sum_{1\le i\ne j\le 2} R_{8,i,j}\Big] + E\Big[\sum_{1\le i\ne j\le 2} R_{9,i,j}\Big] + E\Big[\sum_{1\le i\ne j\le 2} R_{10,i,j}\Big], \quad \text{say,} \tag{3.1}$$

where

$$V = (W_1^{-1} + W_2^{-1})/a \ .$$

From (2.6), one gets

$$E(R_i) = \tfrac{1}{2}de^{-a}\sigma_i^{-1}(\eta_i + \varepsilon_i) + o(d) \ , \qquad i = 1, 2 \tag{3.2}$$

and from (2.8)–(2.9) one has

$$E(R_i) = \tfrac{1}{4}de^{-a}(1 + a)\sigma_i^{-1}p_i + o(d) \ , \qquad i = 3, 4 \ . \tag{3.3}$$

Then, note that

$$E(R_5) = \tfrac{1}{2}de^{-a}(\sigma_1\sigma_2)^{-1}E(N_1^{*})E(N_2^{*})$$

$$= \tfrac{1}{2}de^{-a}(\sigma_1\sigma_2)^{-1}(\eta_1 + \varepsilon_1 + o(1))(\eta_2 + \varepsilon_2 + o(1))(C_1C_2)^{-1/2}$$

$$= O(d^2) + o(d^2) = o(d) \ , \tag{3.4}$$

in view of (2.6).

The terms R_6, R_7, $R_{8,i,j}$ are handled in similar fashions and note that each is a linear combination of expressions of which a typical one can be written as

$$L_{ij} = Bd^{3/2}e^{-aW_i}V^{-i_1}W_i^{-i_2}W_j^{-i_3}N_i^{*3} \tag{3.5}$$

where i_1, i_2, i_3 are fixed non-negative integers. Here and elsewhere we write B as a positive generic constant, independent of d. The aim is to verify that $E(L_{ij}) = o(d)$ for all appropriate i, j. For brevity, we only show how to handle $E(L_{12})$. For $0 < \varepsilon < 1$, define $A_1 = [N_1 > \varepsilon C_1] \cap [N_2 > \varepsilon C_2]$, $A_2 = [N_1 > \varepsilon C_1] \cap [N_2 \le \varepsilon C_2]$, $A_3 = [N_1 \le \varepsilon C_1] \cap [N_2 > \varepsilon C_2]$, and $A_4 = [N_1 \le \varepsilon C_1] \cap [N_2 \le \varepsilon C_2]$. Now,

$$|L_{12}|I(A_1) \le Bd^{3/2}\varepsilon^{-(i_2+i_3)}|N_1^*|^3 \text{ and hence}$$

$$|E(L_{12}I(A_1))| = O(d^{3/2}) = o(d) \tag{3.6}$$

in view of (2.9). Also,

$$|L_{12}I(A_2)| \le Bd^{3/2}(C_2/N_2)^{i_3}|N_1^*|^3 I(A_2) , \quad \text{and thus}$$

$$|E(L_{12}I(A_2))| \le Bd^{(3/2)-i_3}E[|N_1^*|^3 I(N_1 > \varepsilon C_1)]P(N_2 \le \varepsilon C_2)$$

$$= O(d^{Lm+(3/2)-i_3}) = o(d) , \tag{3.7}$$

in view of (2.7) and since "L" can be chosen large. Again,

$$|E(L_{12}I(A_3))| \le Bd^{3/2}E[|N_1^*|^3(C_1/N_1)^{i_3}I(A_3)]$$

$$\le Bd^{-i_2}\int_{A_3}\left(1 - \frac{N_1}{C_1}\right)^3 dP \le Bd^{-i_2}P(A_3)$$

$$= O(d^{Lm-i_2}) = o(d) , \tag{3.8}$$

along the lines of (3.7). Similarly,

$$|E(L_{12}I(A_4))| \le Bd^{-i_2-i_3}P(N_1 \le \varepsilon C_1)P(N_2 \le \varepsilon C_2)$$

$$= O(d^{2Lm-i_2-i_3}) = o(d) , \tag{3.9}$$

along the lines of (3.7). Now, combining (3.6)–(3.9), we claim that $E(L_{12}) = o(d)$. In fact, the terms $R_{9,i,j}$ and $R_{10,i,j}$ also can be evaluated in similar fashions and shown to be $o(d)$, since N_i^* and N_j^* are indeed independent random variables.

Now, combining all these with (3.1)–(3.9), one obtains

$$P\{\delta \in J(\underset{\sim}{N})\} = (1 - \alpha) + \frac{1}{2}de^{-a}\sum_{i=1}^{2}\left\{(\eta_i + \varepsilon_i) - \frac{1}{2}(1 + a)p_i\right\}\sigma_i^{-1} + o(d)$$

which simplifies to $(1 - \alpha) + o(d)$ if one chooses

$$\varepsilon_i = \frac{1}{2}(1 + a)p_i - \eta_i = \frac{1}{2}(3 + a - \rho_i)/\rho_i \ , \qquad i = 1, 2 \ .$$

Remark 1: Suppose now one chooses ε_1, ε_2 as given in Theorem 1 and obtains $X_{i1}, \ldots, X_{iN_i}$, $i = 1, 2$ by implementing the three-stage methodology (2.2)–(2.3). Let us then consider separate confidence intervals for μ_1, μ_2 and propose the natural fixed-width confidence interval $J_i(N_i) = [X_{iN_i(1)} - d, X_{iN_i(1)}]$ for the location parameter μ_i, $i = 1, 2$. From Theorem 2 of Mukhopadhyay and Mauromoustakos (1987), we can immediately conclude that $P\{\mu_i \in J_i(N_i)\} \geq (1 - \alpha) + o(d)$.

Acknowledgement: The first author thanks the Statistics Section of the Department of Mathematics, Monash University, for its kind support given during his sabbatic leave from the University of Connecticut. We also thank Dr R.C. Griffiths of Monash University for showing encouragement.

References

Chow YS, Robbins H (1965) On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* 36:457–462

Hall P (1981) Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.* 9:1229–1238

Mukhopadhyay N (1988) Sequential estimation problems for negative exponential populations. *Commun. Statist., Theory and Methods (Reviews Section)* 17:2471–2506

Mukhopadhyay N (1990) Some properties of a three-stage procedure with applications in sequential anslysis. *Sankhya, Ser. A* 51:218–231

Mukhopadhyay N, Hamdy HI (1984) On estimating the difference of location parameters of two negative exponential distributions. *Canad. J. Statist.* 12:67–76

Mukhopadhyay N, Mauromoustakos A (1987) Three-stage estimation procedures for the negative exponential distributions. *Metrika* 34:83–93

Zelen M (1966) Application of exponential models to problems in cancer research. *J. Roy. Statist. Soc.* 129:368–398

Received 19.05.1992

Book Reviews

'Student' A statistical biography of William Sealy Gosset. Based on writings by E. S. Pearson. Edited and augmented by R. L. Plackett, with the assistance of G. A. Barnard. Clarendon Press, Oxford 1990, pp. viii + 142, £ 15.00

When E. S. Pearson died in 1980, he left a bulky but unfinished set of autobiographical notes which included much material relating to 'Student' (the pseudonym which W. S. Gosset used in his published papers because of his conditions of employment with Arthur Guinness, Son & Co., Ltd.). The editors, R. L. Plackett and G. A. Barnard, together with the Clarendon Press, Oxford, are to be congratulated on producing a very readable and valuable biography of 'Student' by the collation and augmentation of this material. The manner of presentation, which sandwiches chapters on Gosset, Karl Pearson, R. A. Fisher and E. S. Pearson between a chapter on background and a concluding general commentary, is inevitably a bit arbitrary; but nevertheless is successful in providing the reader with a fascinating insight into the professional activities of these four important contributors to the development of the British school of statistical science in the first third of the twentieth century, with Gosset often acting as a catalyst among the other three. The editors were fortunate in the amount of original correspondence available, this including not only that between both Pearsons and Gosset, but also the so-called 'Guinness collection' of letters between Gosset and Fisher circulated by L. McMullen in 1962.

Obviously all the detail available in the book must be savoured by the reader. The reviewer can only mention a few miscellaneous fragments which appealed to him:

(i) In Chapter 2 there is a useful sketch of the beginnings of biometry in which Francis Galton and W. F. R. Weldon were also key figures. Weldon, who was three years younger than Karl Pearson, was, like him, a professor at University College, London, and a close friendship developed between them, so that Pearson was deeply shocked when Weldon suddenly died in 1906, only five years after the founding of the journal *Biometrika.*

(ii) E. S. Beaven, a friend of Gosset, had links with agricultural science at Cambridge, where Fisher's tutor was the astronomer F. J. M. Stratton, whose interest in the combination of observations led via Beaven to an acquaintance with Gosset; a copy of Fisher's first paper on estimation, written while he was still an undergraduate, was sent on Stratton's advice to Gosset.

(iii) Gosset's *Biometrika* paper in 1908 had inferred the sampling distribution of t, where the sample mean is scaled by the sample standard deviation, and was highly esteemed by Fisher. Fisher's use in 1912 of sample space to provide a geometrical derivation of the distribution apparently baffled Karl Pearson when Gosset showed Pearson a letter Fisher had written to him (this in reply to a query from Gosset to Fisher about the appropriate divisor n or n-1 in estimating the

standard deviation from the sample sum of squares of deviations from the mean). It was not until 1915 when Fisher published his derivation of the sample correlation coefficient distribution that his geometrical methods appeared in *Biometrika*, and, as a by-product, confirmed Gosset's inferred *t*-distribution.

(iv) Gosset did not meet Fisher until 1992, when Gosset visited Rothamsted. However, his close relation with Fisher is exemplified by Fisher's asking him to read the proofs of the first edition of *Statistical Methods for Research Workers*, which was published in 1925.

(v) In a quoted letter to Gosset in 1923 from Fisher, there is a clear statement about the subdivision of the total sum of squares into independent components with appropriate degrees of freedom – in effect, the 'analysis of variance'.

(vi) In the chapter on E. S. Pearson, reference is made to his exchange in 1926 with Gosset on the interpretation of statistical tests, this leading in due course to the Neyman-Pearson theory of testing statistical hypotheses.

(vii) An illuminating comment by E. S. Pearson quoted in the concluding General commentary refers to Fisher asking E. S. Pearson to act as a go-between with his father in seeking to put up Gosset for Fellowship of the Royal Society of London. However, E. S. Pearson did not feel it was up to him to take any such action, having in mind the prickly relations which existed at the time (1933 or 1934) even between the two Pearsons. It is sad to think that it was their difficult personalities that prevented Karl Pearson and R. A. Fisher getting together to sponsor Gosset, whom, in spite of his occasional differences with each of them, they both held in such high regard.

Exmouth *M. S. Bartlett*

Brandt A, Franken P, Lisek B: *Stationary Stochastic Models.* John Wiley & Sons 1990, p. 344, £39.95

In recent years fundamental results concerning stochastic equations of the type

$$X_{n+1} = f(X_n, U_n) \ , \quad n \in \mathbb{Z} \tag{1}$$

have been obtained. The interest in (1) was stimulated by the desire to understand the evolution of stochastic models arising in econometrics (see de Haan, Resnick, Rootzen and de Vries (1989)), physics (Chamayou (1973)), environmental processes (Rachev and Todorovic (1990)), sociology (Cavalli-Sforza and Feldman (1973)). The book "Stationary Stochastic Models" is an excellent collection of results on the equation (1) mainly paying attention to its applications to queueing models.

Chapter 1 deals with the steady state behaviour of a stochastic system which time evolution is described by (1) with U_n being a stationary sequence, and U_n

and X_n taking values in complete metric spaces. The aim is to provide stationary weak (or strong) solutions of (1). The qualitative stability (continuity) of the solutions of (1) is also studied here, while the problem of qualitative stability is only stated. Chapter 2 deals with stationary solutions of (1) for i.i.d. U_n. The main result is the Markov property of the weak solution provided that $f(x,u)$ is nondecreasing and left-continuous in x. Chapter 3 contains a short review on marked point processes (MPP) with emphasis on Markov renewal processes and ergodic MPP. The continuous analogue of (1)

$$X(t) = \bar{f}(X(u),\ t-u,\ \theta_u\bar{\varphi},\ u\ ,\quad t\in\mathbb{R}\ ,\quad u<t\ ,$$

with $\bar{\varphi}$ being a MPP was introduced in chapter 4.

For some basic examples of queueing models (as $G|G|\infty, G|G|m|0, G|M|m|r$) the imput flow U_n in (1) has the form (A_n, Y_n) where A_n is treated as the interarrival time between the uth and $(n+1)$th arrival, and Y_n is a mark comprising the characteristics of the nth customer. Chaper 5 deals with the weak stationary solution of (1) exhibiting the behaviour of the above queueing models as the time tends to infinity. Chapter 6 contains examples of relations between the input- and time-stationary characteristics of queueing models. Functional analogue of the Little's formula is discussed in the framework of MPP's. Chapter 7 presents generalizations to queueing models with batch arrivals; having geometric, constant or some aging distribution. Chapter 8 deals with model qualitative stability of the output characteristics of $G|G|\infty$, $G|G|m|0$, $GI|GI|m|0$, $G|G|1|\infty$, $G|GI|m|\infty$. The quantitative stability, so natural for the problems considered and the methods being used, is not here included, cf. the review Rachev (1989). Chapter 9 is a nice illustration of the applicability of the methods developed in the monograph. It deals with the stochastic difference equation

$$X_{n+1} = B_nX_n + C_n\ ,\quad n\in\mathbb{Z}\ , \tag{2}$$

where the sequence (B_n, C_n) is assumed to be stationary and ergodic. The equation arises in economics, physics, nuclear technology, environmental models, biology and sociology, cf. Vervaat (1979). As an example, take the first order ARCH process defined by $\xi_n = Y_n(\beta+\lambda\,\xi_{n-1}^2)^{1/2}$. The results in this section generalized the results of Vervaat (1979) for (B_n, C_n) being i.i.d.

Some final comments: (a) the book is designed for specialists in mathematical queueing theory, applied stochastic processes and applied analysis; (b) there are no exercises, and the monograph is written for readers familiar with measure-theoretic probability and point processes; (c) the vehicle chosen for the exposition is the stochastic differential equation (1), its weak solution and stability. Related topics and techniques in the modern queueing theory are presented in the monographs Borovkov (1984), Franken, König, Arndt and Schmidt (1982), Kalashnikov and Rachev (1990) and Rachev (1991).

References

1. Borovkov AA (1984) Asymptotic Methods in Queueing Theory. Wiley, Chichester
2. Cavalli-Sforza L, Feldman MW (1973) Models for Cultural Inheritance I. Group Mean and Within Group Variation. Theoret Popn Biol 4:42–55
3. Chamayou JMF (1973) Volterra's Functional Integral Equations of the Statistical Theory of Damage. J Computational Phys 13:70–93
4. de Haan L, Resnick SI, Rootzen H, de Vries CG (1989) Extremal Behavior of Solutions to a Stochastic Difference Equation with Applications to ARCH Processes. Processes and Appl 32:213–224
5. Franken P, König D, Arndt H, Schmidt V (1982) Queues and Point Processes. Wiley, Chichester
6. Kalashnikov VV, Rachev ST (1990) Mathematical Methods for Construction of Queueing Models. Wadsworth & Brooks/Cole, Pacific Grove, California
7. Rachev ST (1989) The Problem of Stability in Queueing Theory. Queueing Systems and Applications 4:287–318
8. Rachev ST (1991) Probability Metrics and Stability of Stochastic Models. Wiley, Chichester
9. Rachev ST, Todorovic P (1990) On the Rate of Convergence of Some Functionals of a Stochastic Process. J Appl Prob 28:805–814
10. Vervaat W (1979) On a Stochastic Difference Equation and a Representation of Non-Negative Infinitely Divisible Random Variables. Adv Appl Prob 11:750–783

Santa Barbara *S. T. Rachev*

Wentzell AD: *Limit Theorems on Large Deviations for Markov Stochastic Processes.* Kluwer Acad, Publishers, Mathematics and its Applications (Soviet Series). Dordrecht 1990, p. xv+176, Dfl. 150.00, US $ 79.00, UK £ 53.00

This book is the English translation of the Russian edition, which appeared in 1986. In some sense it can be viewed as a first attempt to give an unified treatment of the various large deviation results, which have been discovered during the last two decades since the famous classical theorem of Cramér in 1938.

Let me first give the rough idea of a large deviation problem as it is presented in the book. Suppose that $(\xi^\alpha)_{\alpha\in A}$ is a net of stochastic processes defined on probability spaces $(\Omega^\alpha, F^\alpha, P^\alpha)$ and taking values in a metric space (X, d). The index set T of the ξ^α may be discrete or continuous. Let X be a space of functions $\varphi: T\to X$ such that every ξ^α has its paths in X. X is assumed to have a metric ϱ inherited from the metric d on X (normally the sup-norm). Suppose that $S: X\to\mathbb{R}_+$ and $k: A\to\mathbb{R}_+$ (with $\varliminf_\alpha k(\alpha)=\infty$) are two functions. Then it is said that a large deviation theorem holds for (ξ^α) with the *action functional* $I_\alpha(\varphi)=k(\alpha)S(\varphi)$, if essentially the following two conditions hold for any $\delta>0$, $\gamma>0$, $t>0$ and sufficiently large α:

$$P^\alpha[\varrho(\xi^\alpha,\varphi)<\delta]\geqslant\exp[-k(\alpha)(S(\varphi)+\gamma)] \tag{I}$$

for every $\varphi\in\Phi(t):=\{\psi\in X: S(\psi)\leqslant t\}$, and

$$P^{\alpha}[\varrho(\xi^{\alpha}, \Phi(s)) \geqslant \delta] \leqslant \exp[-k(\alpha)(s-\gamma)] \tag{II}$$

for every $s \leqslant t$.

The functions S and k are called the *normalized action functional* and the *normalizing coefficient* respectively.

It is mentioned in the introduction (with a reference to the proof) that these two conditions are equivalent to the conditions due to Varadhan:

$$\varliminf_{\alpha} k(\alpha)^{-1} \log P^{\alpha}[\xi^{\alpha} \in G] \geqslant -\inf\{S(\varphi) : \varphi \in G\} \text{ , and} \tag{I'}$$

$$\varliminf_{\alpha} k(\alpha)^{-1} \log P^{\alpha}[\xi^{\alpha} \in F] \leqslant -\inf\{S(\varphi) : \varphi \in F\} \tag{II'}$$

for all open subsets G and all closed subsets F of X. Since the conditions of Varadhan are probably the more familiar conditions for large deviations, it surely would have been better to include the proof of the equivalence of (I), (II) and (I'), (II').

The book is split into two parts. The first part (Chaps. 1–3) contains (beside the necessary prerequisites at the beginning) the basic and very general theorems on large deviations for Markov processes – both for the discrete and continuous parameter case. The idea is to deduce from these general results the special results presented in the second part of the book (Chaps. 4–6).

In Chapter 2 two fundamental inequalities are proved, which are essentially the conditions (I) and (II) for a single Markov process ξ satisfying an analogue of Cramér's condition of finiteness of exponential moments. In the case of a continuous time parameter, ξ is in addition assumed to be locally infinitely divisible. The action functional $I(\varphi)$ for ξ is given by an integral of the Legendre transform of the so-called cumulant G of ξ. The estimate from below for the probability $P[\varrho(\xi, \varphi) < \delta]$ is obtained by a measure substitution, called the generalized Cramér's transformation. The estimate from above for $P[\varrho(\xi, \Phi(t)) \geqslant \delta]$ is obtained by approximating ξ by a suitable polygon $l(\xi)$.

From the basic inequalities in Chapter 2 the main large deviation results for the discrete and continuous parameter case are derived in Chapter 3. The essential assumption on the family (ξ^{α}) is made on the corresponding family (G^{α}) of cumulants, e.g. convergence after norming by the normalizing coefficient:

$$\varliminf_{\alpha} k(\alpha)^{-1} G^{\alpha}(t, x, k(\alpha) z) = G_0(t, x, z) \ .$$

The action functional is now the integral of the Legendre transform of G_0.

In Chapter 4 the general results are applied to a number of examples, especially to the case of locally infinitely divisible processes without jumps, i.e. to diffusions. The possible large deviation results are subclassified into three different cases: "not very large", "very large", and "super-large deviations". For the classical problem of asymptotics of the probabilities:

$$P\left[\frac{\xi_1+\ldots+\xi_n}{n}>x\right]$$

for an i.i.d. sequence (ξ_k), these three cases correspond to the cases $x=0$ $(\sqrt{n})$, x of the same order as $\sqrt{n}$, and x tending to infinity faster than $\sqrt{n}$.

In the fifth chapter precise limit theorems on large deviations for Markov processes are presented. This means that for certain smooth functions F the asymptotic behaviour of the expectations

$$E^{\alpha}\exp\left(k(\alpha)F(\xi^{\alpha})\right)$$

is described.

The last chapter deals with large deviation theorems of a different type than those considered before. Such theorems occur e.g. in the case of Markov processes with large jumps, or in the classical case when the i.i.d. sequence (ξ_k) is attracted to a stable law. In this situation the asymptotic behaviour of the probabilities is no longer exponential but of power type.

The presentation especially of the main results in Chapters 2 and 3 is not very transparent and the proofs would have benefited by more detailed arguments. For a first entry to the subject of large deviations this book alone is surely not suitable, and one should first cast a good look into some of the main original papers (e.g. those of Varadhan). With such prerequisites then the book will give the reader a good introduction into the general theory of large deviations of Markov processes.

Tübingen *E. Dettweiler*

Gittins JC: *Multi-armed Bandit Allocation Indices.* John Wiley & Sons, Chichester, UK, 252+xii pages. £29.95

Table of Chapters:

This book is intended for use by "researchers in chemometrics, combinatorics, economics, numerical analysis, operational research, probability theory and statistics" (author's preface). It may also serve as a text in a graduate course in applied probability. With the latter in mind the author has included exercises at the end of each chapter.

A discrete-time multiarmed bandit problem involves selecting one from a number of stochastic processes (arms) at each of a possibly infinite number of stages. Only one arm may be selected at each stage. There is a reward at each stage that depends on the result obtained at the stage. The results of the previous selections are known when these selections are to be made. Results from arms not selected are not observed. The problem is sequential: Select an arm for observation at each stage depending on what is known at that time. Rewards accumulate from one stage to the next and the objective is to maximize expected discounted reward. The problem is difficult because immediate reward must be weighed against the possibility of gaining information about an arm that may help in making better selections later.

This book considers the important case in which the available arms are independent a priori and the number of stages is infinite with future observations discounted geometrically-exponentially when time is continuous. Further, the discount factor is assumed to be completely known. In this case "index policies" are optimal – a most important and impressive result due to the author. But index policies are not optimal when the arms are dependent or for other types of discounting. In particular, they are not optimal for geometric discounting when the discount factor is unknown. While no book should try to do everything, a discussion of the consequences of using index strategies when they are not optimal would have been a nice addition.

There is another omission that will be important for some users. The book takes a Bayesian approach in which unknown parameters are random variables. How is the user to come up with a prior probability distribution for these parameters? While not important for the theoretician, a discussion of this issue is essential for the practitioner. For example, the book includes tables of allocation indices for Bernoulli arms for various beta prior distributions of the Bernoulli parameter. How is the user to decide which beta distribution to use, or whether a beta distribution is appropriate? What if none is?

The exercises in the book range from difficult to trivial. An example of the latter: Show that the mean of a nonnegative random with distribution function F is

$$\int_0^\infty (1-F(x))\,dx\ .$$

Most of the exercises are well thought out and will serve as valuable learning aids for the diligent reader.

The author is a brilliant mathematician. However, his writing style is not always easy for the uninitiated to penetrate. For example, it is difficult to read the

later chapters in this book without having read the earlier ones. Consider a reader who wants to find out about search theory. For the first problem in Chapter 8, the author indicates that a "related SFABP in fact turns out to have an identical cost structure, thereby providing a solution in terms of an index" (Page 187). (Acronyms such as SFABP are not included in the book's index; nor for that matter is its longer version: simple family of alternative bandit processes".) This SFABP has "a cost structure of the type covered by Corollary 3.10" (Page 187). This corollary is: "Under Condition C the generalized EWFT criterion for a SFABP is minimized by the index policy obtained by putting $\gamma = 0$ in the general expression for the index" (Page 56). The serious reader is then off to find out about "generalized EWFT" (this acronym means expected weighted flow-time) and the "general expression for the index", again with no help from the book's index.

In view of the larger issues, this is nitpicking. The book is an important and useful research tool. The quality of the "stuff" below the surface makes scratching for it worthwhile. I recommend it highly to its intended audience.

Durham *D. A. Berry*

W. Haerdle (Ed.): Applied Nonparametric Regression, Cambridge University Press, 1991, 348 pp., $44.50/£30,00

This book offers a detailed overview on recent developments in nonparametric regression. After a short introduction on different smoothing techniques the discussion concentrates on kernel smoothing. Main themes of current research on kernel smoothing and nonparametric curve estimation are addressed: bandwidth choice, curve estimation in the presence of outliers and for correlated data, nonparametric inference on the qualitative shape of a curve, smoothing in high dimensions. These topics are discussed in full detail and many illustrative examples are included.

The emphasis of the book lies more in discussing statistical motivations and implications of presented results than in a rigorous mathematical argumentation. More technical points (of proofs) are typically only sketched in separate sections, where references are also given for more details. This style allows the author to include a lot of material in the book without getting lost in technicalities. Therefore the book can be read on a relatively nonmathematical level, which makes it particularly useful for people interested in applications of smoothing techniques, rather than in theoretical considerations. The presentation of the book is very clear and the reader gets a good impression about trends and possible future developments in nonparametric curve estimation. The book is recommendable for practitioners as well as for theoreticians.

Heidelberg E. Mammen

Metrika (1993) 40:137–147

Metrika

Combined Unbiased Estimators Based on Q-Reduced Model

KWAN R. LEE

Eastman Chemical Company, Kingsport, TN 37662, USA

C. H. KAPADIA

Department of Statistical Science, Southern Methodist University, Dallas, TX 75275-032, USA

Abstract: In the variance component model $\underline{Y} \sim \left(X\underline{\beta}, \sum_{j=1}^{c} \sigma_j^2 v_j\right)$, Pukelsheim (1981, *Anal. of Statist.*) proved that the non-negative and unbiased estimation of the variance component $\sigma_j^2, j = 1, 2, \ldots, c$ entails a transformation of the original model to $Q\underline{Y}$ (called the Q-reduced model). Lee and Kapadia (1988, *Metrika*), considered the maximum likelihood approach based on the likelihood of the QY (denoted by $Q - ML$) and applied to an incomplete block design model. In this note, the results given in their paper are used to show that the variance of the combined estimator of treatment contrast proposed by Graybill and Deal (1959, *Biometrics*) can be reduced if the $Q - ML$ estimators of the variance components are used instead of AOV estimators as is done in Graybill and Deal's paper.

Key Words: Variance components, $Q - ML$ estimators, Balanced Incomplete Block

1 Model

Consider a mixed linear model of the form,

$$\underline{Y}_j = A_j\underline{\theta} + b_j\underline{1} + \underline{e}_j \tag{1.1}$$

In this model, $\underline{Y}_j = (Y_{1j}, \ldots, Y_{Kj})$ is a $K \times 1$ vector (or block of observations, and A_j is a $K \times I$ matrix of fixed elements. The quantity $\underline{\theta} = (\theta_1, \ldots, \theta_I)'$ is a $I \times 1$ vector of regression coefficients, b_j is the j^{th} block (random) effect having mean zero and variance σ_b^2, $\underline{1}$ is a $K \times 1$ vector of ones, and $\underline{e}_j = (e_{1j}, \ldots, e_{Kj})'$ is a $K \times 1$ vector of random errors. In this model the vector $A_j\underline{\theta}$ is a fixed component of the vector of observations $\underline{Y}_j$ and the vectors $b_j\underline{1}$ and $\underline{e}_j$ are two distinct random components. For an incomplete block design, $\underline{\theta}$ is the $I \times 1$ vector of treatment means and A_j consists of 1's and 0's, each row of which contains a

0026–1335/93/3–4/137–147 $2.50

single 1 indicating which treatment is included in the j^{th} block. This model includes a balanced incomplete block design (BIBD) as a special case. (See Box and Tiao 1973.)

We assume that b_j and the element $\underline{e}_j, j = 1, 2, \ldots, J$ are independent and that $b_j \sim N(0, \sigma_b^2)$, $\underline{e}_j \sim N(0, \sigma_e^2 I)$. Let $\sigma_{be}^2 = \sigma_e^2 + K\sigma_b^2$.

In a BIBD, t treatments are examined in b blocks, each of equal size K. Each treatment appears r times and each pair of treatments appears exactly λ times in different blocks. For any such designs, we must have

$$b \geq t \ , \qquad bK = tr \ , \qquad \lambda(t-1) = r(K-1) \ . \tag{1.2}$$

In terms of the general model in (1.1), A_j is a $K \times t$ matrix of 1's and 0's each row of which contains a single 1 in the column corresponding to the treatment applied. Clearly A_j consists of othogonal rows. The $t \times b$ matrix

$$N = [A_1' \underline{1}_K | \ldots | A_b' \underline{1}_K] \tag{1.3}$$

is known as the incidence matrix whose elements consist of 1's and 0's indicating the presence and absence of the t treatments in the b blocks. Since $A_j, j = 1, \ldots, b$ have patterned forms whose rows are orthogonal, the following relationships hold,

$$\sum_{j=1}^{b} A_j' A_j = rI_t \ , \tag{1.4}$$

$$\sum_{j=1}^{b} A_j' R A_j = \left(\frac{r-\lambda}{K}\right) I_t + \frac{\lambda}{K} \underline{1}_t \underline{1}_t' \ , \tag{1.5}$$

$$\left(\sum_{j=1}^{b} A_j' R A_j\right)^{-1} = \left(\frac{K}{r-\lambda}\right)\left(I_t - \frac{\lambda}{rK} \underline{1}_t \underline{1}_t'\right) \ , \qquad r > \lambda \tag{1.6}$$

and

$$\sum_{j=1}^{b} A_j'(I-R)A_j = \frac{\lambda t}{K}(I_t - t^{-1} \underline{1}_t \underline{1}_t') \ , \tag{1.7}$$

where I_t is a $t \times t$ identity matrix $\underline{1}_t$ is a $t \times 1$ vector of 1's.

Let $\underline{B}' = (B_1, \ldots, B_t)$ and B_i is the average of the block means $\bar{y}_{.j}$ for those blocks j where the i^{th} treatment appears and $\underline{T}' = (T_1, \ldots, T_t)$ where T_i is the average for the i^{th} treatment over the blocks where the treatment appears and $\bar{y}_{..}$ is the grand average, i.e.,

$$\underline{B} = \frac{1}{r}\sum_{j=1}^{b} A_j' \underline{1}_K y_{.j} = \frac{1}{r} N \underline{Y}_b \ ,$$

$$\underline{T} = \frac{1}{r}\sum_{j=1}^{b} A_j' \underline{Y}_j \qquad \text{and}$$

$$\bar{y}_{..} = \frac{1}{bk}\sum_{j=1}^{b}\sum_{k=1}^{K} Y_{kj} = \frac{1}{tr}\sum_{j=1}^{b}\sum_{k=1}^{K} Y_{kj}$$

or

$$\sum_{j=1}^{b} A_j' \underline{1}_K y_{.j} = r\underline{B} \ , \tag{1.8}$$

$$\sum_{j=1}^{b} A_j'(I - R)\underline{Y}_j = r(\underline{T} - \underline{B}) \qquad \text{and} \tag{1.9}$$

$$\underline{1}_t' \underline{T} = 1_t' = \underline{B} = t\bar{y}_{..} \ . \tag{1.10}$$

From (1.5) and (1.7), it follows that

$$q_b = \operatorname{rank}\left[\sum_{j=1}^{b} A_j' R A_j\right] = t \ ,$$

$$q_e = \operatorname{rank}\left[\sum_{j=1}^{b} A_j'(I - R)A_j\right] = t - 1 \ .$$

2 Estimator of the Treatment Contrasts Inter and Intra Block Model and *QML* Estimators

The estimable inter and intral block treatment contrasts are given in the following table. For more details see Box and Tiao (1973).

Table 1. Estimators of Treatment Contrasts in BIBD

Parameter	Estimator	Variance
$u_i = \theta_i - \bar{\theta}$	inter $\tilde{u}_i = \frac{Kr}{r-\lambda}(B_i - y_{..})$ (2.1)	$\frac{K}{r-\lambda}\left(\frac{t-1}{t}\right)\sigma_{be}^2$ (2.2)
$i = 1, \ldots, t$	intra $\hat{u}_i = \frac{rK}{\lambda t}(T_i - B_i)$ (2.3)	$\frac{K}{\lambda t}\left(\frac{t-1}{t}\right)\sigma_e^2$ (2.4)
$v_i = \frac{1}{\sqrt{2}}(\theta_i - \theta_{i+1})$	inter $\tilde{v}_i = \frac{1}{\sqrt{2}}\frac{Kr}{r-\lambda}(B_i - B_{i+1})$ (2.5)	$\frac{K}{r-\lambda}\sigma_{be}^2$ (2.6)
$i = 1, 2, \ldots, t-1$	intra $\hat{v}_i = \frac{1}{\sqrt{2}}\frac{rK}{\lambda t}(T_i + B_i - T_{i+1} - B_{i+1})$ (2.7)	$\frac{K}{\lambda t}\sigma_e^2$ (2.8)

Finally, from Kapadia and Lee (1988),

$$Q - ML(\sigma_e^2) = \min\left[\frac{S_e}{bK-b-t+1}, \frac{S_b+S_e}{bK-2t-1}\right],$$

$$Q - ML(\sigma_b^2) = \frac{1}{K}\left[\frac{S_b}{b-t} - \frac{S_e}{bK-b-t+1}\right]^+, \tag{2.9}$$

and

$$Q - ML(\sigma_{be}^2) = \max\left[\frac{S_b}{b-t}, \frac{S_e+S_b}{bK-2t-1}\right],$$

where $a^+ = \max(a, 0)$, $S_e = \sum_{j=1}^{b}\sum_{k=1}^{K} Y_{kj}^2 - K\sum_{j=1}^{b} y_{.j}^2 \frac{r^2K}{\lambda t}\sum_{i=1}^{t}(T_i - B_i)^2$

and $S_b = K\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2 - \frac{r^2K}{r-\lambda}\sum_{i=1}^{t}(B_i - \bar{y}_{..})^2$.

3 Estimation of the Treatment Contrasts in the Inter and Intra Block Model

The combined estimator of $\underline{u}$ denoted by $\bar{\underline{u}}$ is

$$\bar{\underline{u}} = \frac{(1-e)\tilde{\underline{u}} + e\cdot\omega\hat{\underline{u}}}{(1-e)+e\cdot\omega} \quad \text{or} \quad \bar{u}_i = \frac{(1-e)\tilde{u}_i + e\cdot\omega\hat{u}_i}{(1-e)+e\cdot\omega} \tag{3.1}$$

and that of $\underline{v}$, denoted by $\overline{\underline{v}}$, is

$$\overline{\underline{v}} = \frac{(1-e)\tilde{\underline{v}} + e\cdot\omega\hat{\underline{v}}}{(1-e)+e\cdot\omega} \quad \text{or} \quad \overline{v}_i = \frac{(1-e)\tilde{v}_i + e\omega\hat{v}_i}{(1-e)+e\cdot\omega} \tag{3.2}$$

where $e = \dfrac{\lambda t}{rK}$, $\omega = \sigma_{be}^2/\sigma_e^2$ u_i's and v_i's are given in Table 1.

In the above results, $\omega = \sigma_{be}^2/\sigma_e^2$ is unknown and hence we have to estimate ω. $\overline{v}_i = \overline{v}_i(\omega)$ can be written in the following form

$$\overline{v}_i(\omega) = \hat{v}_i + \frac{(r-\lambda)}{(r-\lambda)\lambda t\cdot\omega}(\tilde{v}_i - \hat{v}_i) \ . \tag{3.2a}$$

Graybill and Deal (1959) use the AOV estimators of σ_b^2 and σ_e^2 to estimate ω, i.e., they substituted the individual estimators σ_b^2 and σ_e^2 in ω, which is

$$AOV(\omega) = \frac{AOV(\sigma_e^2) + K\,AOV(\sigma_b^2)}{AOV(\sigma_e^2)} \ . \tag{3.3}$$

where

$$AOV(\sigma_e^2) = \frac{S_e}{\nu_1} \quad \text{and} \quad AOV(\sigma_b^2) = \frac{1}{K}\left[\frac{S_b}{\nu_2} - \frac{S_e}{\nu_1}\right] , \tag{3.4}$$

where

$$\nu_1 = bK - b - t + 1 \ , \qquad \nu_2 = b - t \ .$$

Substituting (3.3) into (3.2) gives

$$AOV(\omega) = \frac{\nu_1}{\nu_2}\frac{S_b}{S_e} \ . \tag{3.5}$$

Hence the resultant estimator of v_i is

$$\overline{v}_i(AOV(\omega)) = \hat{v}_i + \frac{r-\lambda}{(r-\lambda) + \lambda t\cdot\dfrac{\nu_1}{\nu_2}\dfrac{S_b}{S_e}}(\tilde{v}_i - \hat{v}_i) \ . \tag{3.6}$$

Graybill and Deal (1959) show that $\bar{v}_i(AOV(\omega))$ is unbiased for v_i and is uniformly better (smaller variance) than either $\hat{v}_i$ or $\tilde{v}_i$ if the following is true.

i) $rt - b - t + 1 \geq 18$ and $b - t = 9$
or ii) $b - t \geq 10$.

If the $Q - ML$ estimators of σ_b^2 and σ_e^2 (and hence σ_{be}^2 and σ_b^2) are used instead of the AOV estimators then the resultant estimator of ω is given as

$$
\begin{aligned}
Q - ML(\omega) &= \frac{Q - ML(\sigma_{be}^2)}{Q - ML(\sigma_e^2)} \\
&= \frac{\max\left[\dfrac{S_b}{\nu_2}, \dfrac{S_e + S}{\nu_1 + \nu_2}\right]}{\min\left[\dfrac{S_e}{\nu_1}, \dfrac{S_e + S_b}{\nu_1 + \nu_2}\right]} \\
&= \max\left[1, \frac{\nu_1}{\nu_2}\frac{S_b}{S_e}\right] .
\end{aligned}
\tag{3.7}
$$

The last equality follows from the fact that

$$
\frac{S_b}{\nu_2} \geq \frac{S_e + S_b}{\nu_1 + \nu_2} \quad \text{is equivalent to} \quad \frac{S_e}{\nu_1} \leq \frac{S_e + S_b}{\nu_1 + \nu_2} ,
$$

which is in turn equivalent to

$$
1 \leq \frac{\nu_1}{\nu_2}\frac{S_b}{S_e} .
$$

If the $Q - ML$ estimator of ω in (3.7) is used in (3.2a) the result is

$$
\bar{v}_i(Q - ML(\omega)) = \hat{v}_i + \frac{(r - \lambda)}{(r - \lambda) + \lambda t \cdot Q - ML(\omega)}(\tilde{v}_i - \hat{v}_i) . \tag{3.8}
$$

Note that in (3.6), the condition

$$
\frac{\nu_1}{\nu_2}\frac{S_b}{S_e} > 1
$$

is equivalent to

$$
\hat{\sigma}_b^2 \equiv AOV(\sigma_b^2) > 0 .
$$

Hence (3.8) can be written as:

$$\text{If} \quad \sigma_b^2 > 0 \ , \qquad \bar{v}_i(Q - ML(\omega)) = \hat{v}_i + S(\tilde{v}_i - \hat{v}_i)$$

$$\text{where} \quad S = (r - \lambda) \Big/ \left\{(r - \lambda) + \lambda t \frac{v_1}{v_2} \frac{S_b}{S_e}\right\} \tag{3.9}$$

$$\text{If} \quad \hat{\sigma}_b^2 \leq 0 \ , \qquad \bar{v}_i(Q - ML(\omega)) = \hat{v}_i + \frac{r - \lambda}{rK}(\tilde{v}_i - \hat{v}_i) \ .$$

or

$$\text{If} \quad \hat{\sigma}_b^2 > 0 \ , \qquad \bar{v}_i(Q - ML(\omega)) = \bar{v}_i(AOV(\omega)) \ .$$

$$\text{If} \quad \hat{\sigma}_b^2 \leq 0 \ , \qquad \bar{v}_i(Q - ML(\omega)) = \hat{v}_i + \frac{r - \lambda}{rK}(\tilde{v}_i - \hat{v}_i) \tag{3.10}$$

from (3.6) and (3.9).

$$\text{Let} \quad I(\sigma_b^2) = \begin{cases} 0 \ , & \text{if} \quad \sigma_b^2 > 0 \\ 1 \ , & \text{if} \quad \sigma_b^2 \leq 0 \ . \end{cases} \tag{3.11}$$

To show that $\bar{v}_i(Q - ML(\omega))$ is unbiased for v_i, rewrite (3.10) using (3.11), i.e.,

$$\bar{v}_i(Q - ML(\omega)) = \left\{1 - I(\hat{\sigma}_b^2)\right\} \bar{v}_i(AOV(\omega)) + I(\sigma_b^2) \left\{\hat{v}_i + \frac{r - \lambda}{rK}(\tilde{v}_i - \hat{v}_i)\right\}$$

or

$$\bar{v}_i(Q - ML(\omega)) = \bar{v}_i(AOV(\omega)) + I(\hat{\sigma}_b^2) \left\{\left(\frac{r - \lambda}{rK} - S\right)(\tilde{v}_i - \hat{v}_i)\right\} \ . \tag{3.12}$$

Since $\bar{v}_i(AOV(\omega))$ is unbiased for v_i, in order to show $\bar{v}_i(Q - ML(\omega))$ is unbiased, it is sufficient to show that

$$E\left[I(\hat{\sigma}_b^2) \left\{\left(\frac{r - \lambda}{rK} - S\right)(\tilde{v}_i - \hat{v}_i)\right\}\right] = 0 \ , \tag{3.13}$$

where

$$S = \frac{r - \lambda}{(r - \lambda) + \lambda t \cdot \dfrac{v_1}{v_2} \dfrac{S_b}{S_e}} .$$

Note that $\hat{\sigma}_b^2$ and S depend on the data only through S_b and S_e and since $\tilde{v}_i$, $\hat{v}_i$, S_b and S_e are mutually independent, the left-hand side of (3.13) can be factored as

$$E\left\{I(\hat{\sigma}_b^2)\left(\frac{r - \lambda}{rK} - S\right)\right\} E(\tilde{v}_i - \hat{v}_i)$$

and $E(\tilde{v}_i - \hat{v}_i) = 0$ since $\tilde{v}_i$ and $\hat{v}_i$ are unbiased for v_i. Hence $\bar{v}_i(Q - ML(\omega))$ is unbiased for v_i. Since both $\bar{v}_i(AOV(\omega))$ and $\bar{v}_i(Q - ML(\omega))$ are unbiased for v_i, we wish to compare their variances.

$$\begin{aligned}
&\mathrm{Var}\{\bar{v}_i(Q - ML(\omega))\} \\
&= E\left[\{\bar{v}_i(AOV(\omega)) - v_i\} + I(\hat{\sigma}_b^2)\left\{\left(\frac{r - \lambda}{rK} - S\right)(\tilde{v}_i - \hat{v}_i)\right\}\right]^2 \\
&= \mathrm{Var}\{\bar{v}_i(AOV(\omega))\} \\
&\quad + 2E\left[\{\bar{v}_i(AOV(\omega)) - v_i\} I(\hat{\sigma}_b^2)\left\{\left(\frac{r - \lambda}{rK} - S\right)(\tilde{v}_i - \hat{v}_i)\right\}\right] \\
&\quad + E\left\{I(\hat{\sigma}_b^2)\left(\frac{r - \lambda}{rK} - S\right)^2 (\tilde{v}_i - \hat{v}_i)^2\right\} .
\end{aligned}$$

From (3.6)

$$\begin{aligned}
&\mathrm{Var}\{\bar{v}_i(AOV(\omega))\} - \mathrm{Var}\{\bar{v}_i(Q - ML(\omega))\} \\
&= -2E\{(\tilde{v}_i - \hat{v}_i)(\hat{v}_i - v_i)\} E\left\{I(\hat{\sigma}_b^2)\left(\frac{r - \lambda}{rK} - S\right)\right\} \\
&\quad - E(\tilde{v}_i - \hat{v}_i)^2 E\left\{I(\hat{\sigma}_b^2)\left(\left(\frac{r - \lambda}{rK}\right)^2 - S\right)\right\} ,
\end{aligned} \tag{3.14}$$

where the last equality follows from the mutual independence of

$\underline{\tilde{v}}, \underline{\hat{v}}, S_b$ and S_e .

Let $w_i = \hat{v}_i$ and $Z_i = \tilde{v}_i - \hat{v}_i$, then we have

$$\begin{pmatrix} w_i \\ Z_i \end{pmatrix} \sim N\left[\begin{pmatrix} v_i \\ 0 \end{pmatrix}, \begin{pmatrix} K\sigma_e^2/\lambda t & -K\sigma_e^2/\lambda t \\ -K\sigma_e^2/\lambda t & K^2(r\sigma_e^2 + \lambda t\sigma_b^2)/\lambda t(r-\lambda) \end{pmatrix}\right] \tag{3.15}$$

Hence from (3.14),

$$\begin{aligned} &\mathrm{Var}\{\bar{v}_i(AOV(\omega))\} - \mathrm{Var}\{\bar{v}_i(Q - ML(\omega))\} \\ &= \frac{2K\sigma_e^2}{\lambda t} E\left\{\frac{r-\lambda}{rK} - S \,\middle|\, \hat{\sigma}_b^2 \le 0\right\} \cdot Pr(\hat{\sigma}_b^2 \le 0) \\ &\quad - \frac{K^2(r\sigma_e^2 + \lambda t\sigma_b^2)}{\lambda t(r-\lambda)} E\left\{\left(\frac{r-\lambda}{rK}\right)^2 - S^2 \,\middle|\, \hat{\sigma}_b^2 \le 0\right\} \cdot Pr(\hat{\sigma}_b^2 \le 0) \ . \end{aligned} \tag{3.16}$$

Let $\theta = \dfrac{r\sigma_e^2}{r\sigma_e^2 + \lambda t\sigma_b^2}$ and $p = Pr(\hat{\sigma}_b^2 \le 0)$ and note that

$$\begin{aligned} S &= \frac{r-\lambda}{rK} \frac{rS_e}{rS_e + \lambda t v_1 \hat{\sigma}_b^2} \\ &= \frac{r-\lambda}{rK} y \ , \qquad \text{where} \qquad y = \frac{rS_e}{rS_e + \lambda t v_1 \hat{\sigma}_b^2} \end{aligned}$$

and $y \ge 1$ whenever $\hat{\sigma}_b^2 \le 0$. Hence from (3.16)

$$\begin{aligned} &\mathrm{Var}\{\bar{v}_i(AOV(\omega))\} - \mathrm{Var}\{\bar{v}_i(Q - ML(\omega))\} \\ &= \frac{r-\lambda}{\lambda t r^2}(r\sigma_e^2 + \lambda t\sigma_b^2) E\{(y-1)(y+1-2\theta) | y \ge 1\} \cdot p \ . \end{aligned} \tag{3.17}$$

Here $r - \lambda \ge 0$, $0 < \theta \le 1$ and hence

$$(y-1)(y+1-2\theta) \ge 0 \qquad \text{whenever} \quad y \ge 1 \ .$$

Therefore, (3.17) yields

$$\mathrm{Var}\{\bar{v}_i(AOV(\omega))\} \ge \mathrm{Var}\{\bar{v}_i(Q - ML(\omega))\} \ ,$$

i.e., a better (smaller variance) estimator of treatment contrasts can be obtained by using the $Q - ML$ estimators of the variance components than the AOV estimators. Note that the reduction of variance is porportional to the probability of $\hat{\sigma}_b^2 = AOV(\sigma_b^2)$ being negative. Since the sampling distribution of y is unknown, further reduction of (3.17) cannot be made at this stage.

4 An Example

To illustrate the theory the following data set is generated by simulation where $t = 3$ treatments, $b = 15$ blocks of size $k = 2$, each treatment replicated $r = 10$ times; thus $\lambda = 5$. The data were generated from BIBD model with $\theta_1 = 5, \theta_2 = 7$, $\theta_3 = 9$, $\sigma_e^2 = 2.25$, $\sigma_b^2 = 1.0$.

Treatment.

Block	1	2	3	$y_{.j}$
1	4.66	5.99		5.32
2	5.86	7.30		6.58
3	4.65	9.22		6.94
4	5.74	8.02		6.88
5	4.49	7.73		6.11
6	5.21		10.52	7.87
7	4.02		9.79	6.91
8	5.60		8.87	7.24
9	7.46		9.15	8.31
10	4.24		10.25	7.24
11		8.06	5.81	6.94
12		8.25	10.65	9.45
13		8.78	10.88	9.83
14		9.50	9.20	9.35
15		6.53	10.01	8.27
T_i	5.19	7.94	9.51	$y_{..} = 7.547$
B_i	6.94	7.57	8.14	$S_e = 21.9011$
				$S_b = 17.1995$

Hence from (3.4),

$$AOV(\sigma_e^2) = 1.6847 \ , \qquad AOV(\sigma_b^2) = -0.1257 \ .$$

Also from (2.9),

$$Q - ML(\sigma_e^2) = 1.5640 \ , \quad Q - ML(\sigma_b^2) = 0 \qquad \text{and hence} \qquad \omega = 1 \ .$$

References

Box GEP, Tiao GC (1973) Bayesian to Inference in Statistical Analysis. Reading, Mass: Addison-Wesley

Graybill FA, Deal RB (1959) Combining unbiased estimators, *Biometrics* 15:543–550

Lee KR, Kapadia CH (1988) Maximum likelihood estimators of the variance components based on the Q-reduced model, *Metrika* 35:178–189

Pukelsheim F (1981) On the existence of unbiased non-negative estimates of variance covariance components. *Ann. Statist.* 9:293–299

Received 29.05.1992

Book Review

S. M. Veres: Structure Selection of Stochastic Dynamic Systems. The Information Criterion Approach, Gordon and Breach Sc. Publ., 1991, XIII/342 pp., Hardcover, £28.00/$50.00

This book gives a reliable review on structure selection of stochastic dynamic systems using information criteria and stochastic complexity. Selection of model structure for a dynamic system is part of system identification. The use of information criteria seems to be useful if one has little a priori information about the possible models. After a short introduction in Chap. 2, the most important results of the literature are summarized. In the rest of the chapters the author presents his own results. The information criteria AIC, BIC and ϕ, are reported and their advantages and disadvantages are pointed out. In Chap. 3, the asymptotic theory of discrete-type models is investigated in Bayesian framework. The strong consistency of the complexity criterion for ARMA processes is proved. In Chap. 4, the model validation is concerned. The author concentrates on the invariance principle of likelihood ratios for ARMA processes. After theoretical investigations many simulations are included to show finite-sample behaviour of the model-structure estimators which illustrate the effectiveness and the limitations of the methods. In Chap. 5, the small-sample behavior of several methods is studied. Recursive least squares, extended least squares, on-line two-stage, recursive maximum likelihood, on-line three-stage, off-line two-stage, off-line three-stage and off-line maximum likelihood are studied by simulation. Chapter 6 is concerned with the estimation of continuous-time processes.

The reader can gain his own experience on the working of many methods using the demonstration disk. Finally, in Appendix A results on martingale convergence are given, in Appendix B some PASCAL routines are listed and Appendix C is a description of the attached simulation and identification programs.

This book will be helpful to researchers interested in applying methods of model-structure selection in time series analysis, in control or in signal processing.

Debrecen M. Arató

Metrika (1993) 40:149–172

Metrika

Two-Sample Rank Tests with Truncated Populations

C. H. LIN

Ming Chuan College, Taipei, ROC

SHASHIKALA SUKHATME

Department of Statistics, Iowa State University, Ames, IA 50011

Abstract: Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent samples from continuous distributions F and G respectively. Using a Hoeffding (1951) type theorem, we obtain the distributions of the vector $\mathbf{S} = (S_{(1)}, \ldots, S_{(n)})$, where $S_{(j)} = \#(X_i\text{'s} \leq Y_{(j)})$ and $Y_{(j)}$ is the j-th order statistic of Y sample, under three truncation models: (a) G is a left truncation of F or G is a right truncation of F, (b) F is a right truncation of H and G is a left truncation of H, where H is some continuous distribution function, (c) G is a two tail truncation of F. Exploiting the relation between $\mathbf{S}$ and the vector $\mathbf{R}$ of the ranks of the order statistics of the Y-sample in the pooled sample, we can obtain exact distributions of many rank tests. We use these to compare powers of the Hajek test (Hajek 1967), the Sidak Vondracek test (1957) and the Mann-Whitney-Wilcoxon test.

We derive some order relations between the values of the probagility-functions under each model. Hence find that the tests based on $S_{(1)}$ and $S_{(n)}$ are the UMP rank tests for the alternative (a). We also find LMP rank tests under the alternatives (b) and (c).

Key Words: Hoeffding type theorem, UMP rank test, LMP rank test

1 Introduction

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent samples from continuous distributions F and G respectively. We obtain the exact distributions of $\mathbf{S}$, defined in (3), when the samples are from truncated populations. We consider the cases when one distribution is truncated by the other or the two distributions are different types of truncation of some other distribution. Truncation does occur in many practical problems, for example, see Deeley et al. (1969) and Nelson (1990).

The vector $\mathbf{S}$ is obviously related to $\mathbf{R}$, the vector of ranks of the order statistics of Y sample in the pooled sample $\{X_1, \ldots, X_m, Y_1, \ldots, Y_n\}$. Hence we find the distribution of $\mathbf{R}$ and the distributions of rank statistics when the observations are truncated. The basic tool is a Hoeffding (1951) type theorem given in the

0026–1335/93/3–4/149–172 $2.50

appendix. We give power comparison of the Hajek (1967) test, the Sidak and Vondracek (1957) test and the Mann-Whitney-Wilcoxon test.

We derive some order relations between the values of the probability functions under each model, prove that the tests based on $S_{(1)}$ and $S_{(n)}$ are the UMP rank tests for the model (a), and find LMP rank tests for the models (b) and (c).

Several two-sample location tests based on exceedances are for the truncation alternatives where in the distributions do not have the same support, e.g., see Haga (1960), Hajek and Sidak (1967, p. 89), and Sidak (1977). Also, the two-sample scale tests, studied by Klotz (1962), are for the alternatives under consideration. The following are some examples of truncation models:

(a) In a two-sample location shift problem for exponential distributions, one distribution is just a truncation of the other. (b) Let F be the d.f. of $U(a, b)$ and G be the d.f. of $U(a + \Delta, b + \Delta)$ for $\Delta > 0$. Then F is the right truncated distribution of $U(a, b + \Delta)$ and G is the left truncated distribution of $U(a, b + \Delta)$. (c) Let F be the d.f. of $U(-0.5, 0.5)$ and G be the d.f. of $U(-0.5\eta, 0.5\eta)$ for $\eta > 0$. If $\eta > 1$ then F is the two-sided (left and right) truncation of the distribution function G.

Though some of the exceedance tests, appropriate for truncation alternatives, were introduced over thirty years ago, apart from some results based on Monte Carlo studies and some for special shift alternatives relatively very little is known about their exact powers. Sidak and Hojek (1977) give Monte Carlo comparison of some rank tests for uniform distributions, Sukhatme (1986) uses the idea of expressing a rank statistic as a function $S_{(i)}$'s and/or $S^*_{(j)}$'s defined in (3) and the properties of order statistics for computing exact powers of rank tests, and illustrates the method by computing the exact powers of the Mathisen (1943) and Rosenbaum (1953) tests for some Lehmann (1953) alternatives, alternatives with shift in location of a uniform distribution. Callfas and Mohanty (1986) develop an algorithm to derive the distributions of rank tests under Lehmann alternatives and derive the exact joint distribution of the Wilcoxon statistic and the Kolmogorov- Smirnov statistic. Katzenbeisser (1989) uses a method similar to Sukhatme (1986) to find exact powers of two-sample location tests for shift alternatives, some of his results are in Sukhatme (1986). Lin and Sukhatme (1990) compute exact powers of several two-sample rank tests.

In Section 2 we derive the distribution of **S** under the alternative H_a and prove that the exceedance test is the UMP rank test for the alternative.

In Section 3 we derive the distribution of **S** under the alternative H_b, present a method of ranking the values of the probability function of **S** in many situations, and use the ranking to search for LMP tests.

In Section 4 we give some rank tests appropriate for H_b. Hajek and Sidak (1967, Problem II.13) have shown that the Sidak and Vondracek (1957) test is the LMP rank test at the right end for testing the shift to right in location of an uniform distribution for a particular choice of α. They also proved the Hajek test is the LMP rank test at the left end for testing the same problem. We find that these tests, however, are not the locally most powerful tests for a general choice of α.

In Section 5 we briefly discuss rank tests for the alternative H_c.

Appendix **A** is for a Hoeffding (1951) type theorem, used to obtain distributions of **R** under various alternatives.

Definitions and Notations:
Define

$$S_{(k)} = mF_m(Y_{(k)}) = \#\{X_i s \le Y_{(k)}\} , \qquad k = 1, \ldots, n \tag{1}$$

$$S^*_{(j)} = nG_n(X_{(j)}) = \#\{Y_k s \le X_{(j)}\} , \qquad j = 1, \ldots, m \tag{2}$$

and

$$\mathbf{S} = (S_{(1)}, \ldots, S_{(n)}) , \qquad \mathbf{S}^* = (S^*_{(1)}, \ldots, S^*_{(m)}) \tag{3}$$

where $\#\mathbf{A}$ denotes the number of elements of the set **A**, $Y_{(k)}$ is the k-th order statistic of the Y sample and $X_{(j)}$ is the j-th order statistic of the X sample, $F_m(x)$ and $G_n(x)$ are the empirical distribution functions of F and G respectively. The sample spaces of **S** and **S*** are, respectively,

$$\mathbf{A} = \{(i_1, \ldots, i_n): 0 \le i_1 \le \cdots \le i_n \le m\} \tag{4}$$

and

$$\mathbf{A}^* = \{(j_1, \ldots, j_m): 0 \le j_1 \le \cdots \le j_m \le n\} . \tag{5}$$

Note that $R_k = S_{(k)} + k$ and $R^*_j = S^*_{(j)} + j$ are the ranks of $Y_{(k)}$ and $X_{(j)}$ in the pooled sample, respectively. Therefore, **S** and **S*** are entirely dependent; that is, for each $\mathbf{i} = (i_1, \ldots, i_n) \in A$ there exists a unique $\mathbf{j} = (j_1, \ldots, j_m) \in \mathbf{A}^*$ so that

$$\{\mathbf{S} = \mathbf{i}\} = \{\mathbf{S}^* = \mathbf{j}\}$$

and

$$\{i_1 + 1, \ldots, i_n + n, j_1 + 1, \ldots, j_m + m\} = \{1, \ldots, m + n\} . \tag{6}$$

Therefore, for each $\mathbf{i} \in \mathbf{A}$, we call $\mathbf{j} \in \mathbf{A}^*$ the dual vector of **i** if the above relation is satisfied. Every rank statistic can be expressed in terms of $S_{(k)}$s and $S^*_{(j)}$s.

For each $N = 1, 2, \ldots$, let $U_{1:N} \leq \cdots \leq U_{N:N}$ be the order statistics of a random sample of size N from $U(0, 1)$, the uniform distribution on $(0, 1)$. For each $r(1 \leq r \leq N)$, $U_{r:N}$ has the Beta distribution, $\text{Be}(r, N - r + 1)$. Also, define the incomplete beta function with parameters, (α, β), by

$$B(t; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \int_0^t u^{\alpha-1}(1-u)^{\beta-1}\, du \tag{7}$$

where $B(\alpha, \beta)$ is the usual beta function.

We denote by $I_A(x)$ and $I(A)$ the usual indicator function and the usual indicator random variable, respectively.

Define three basic truncation transformations, T_L, T_R and T_2 as follows:

1. $G = [(F - \delta)/(1 - \delta)]I_{(F > \delta)}$ is the left truncation of F with truncation rate $\delta \in [0, 1)$. The left truncation transformation $T_L(\cdot; \delta)$ for $t \in [0, 1]$ is then,

$$T_L(t; \delta) = [(t - \delta)/(1 - \delta)]I_{(t > \delta)} \,. \tag{8}$$

2. $G = [F/(1 - \delta)]I_{(F < 1-\delta)}$ is the right truncation of F with the truncation rate $\delta \in [0, 1)$. The right truncation transformation $T_R(\cdot; \delta)$ for $t \in [0, 1]$ is then,

$$T_R(t; \delta) = [t/(1 - \delta)]I_{(t < 1-\delta)} \,. \tag{9}$$

3. $G = [(F - \delta_L)/(1 - \delta_L - \delta_R)]I_{(\delta_L < F < 1-\delta_R)}$ is the two-sided truncation of F with the left and right truncation rates, $\delta_L, \delta_R \geq 0$, $\delta_L + \delta_R < 1$. The two-sided truncation transformation $T_2(\cdot; \delta_L, \delta_R)$ for $t \in [0, 1]$ is,

$$T_2(t; \delta_L, \delta_R) = [(t - \delta_L)/(1 - \delta_L - \delta_R)]I_{(\delta_L < t < 1-\delta_R)} \,. \tag{10}$$

Note that T_L, T_R and T_2 are absolutely continuous distribution functions on $[0, 1]$.

We consider three alternative hypotheses as follows:

A. H_a: $G = T_L(F; \delta)$ (or H_a': $G = T_R(F; \delta)$) for $0 < \delta < 1$.
B. H_b: $F = T_R(H; \delta_R)$ and $G = T_L(H; \delta_L)$ for $\delta_R, \delta_L > 0$ and $\delta_R + \delta_L < 1$ where H is some continuous distribution function.
C. H_c: $G = T_2(F; \delta_L, \delta_R)$ for $\delta_R, \delta_L > 0$ and $\delta_R + \delta_L < 1$.

2 The Distribution of S Under the Alternative H_a

Corollary A.2 in Appendix A is used to obtain the distribution of **S** in the following theorem. From the definitions of T_L and T_R we have for $t, \delta \in (0, 1)$,

$$T_L'(t; \delta) = (1 - \delta)^{-1} I_{(t > \delta)} \tag{11}$$

$$T_R'(t; \delta) = (1 - \delta)^{-1} I_{(t < 1 - \delta)} \tag{12}$$

Theorem 2.1: For each $\mathbf{i} \in \mathbf{A}$ and the corresponding dual vector $\mathbf{j} \in \mathbf{A}^*$ and for each $\delta \in [0, 1)$, we have

1. $$\mathbf{P}\{\mathbf{S} = \mathbf{i} | G = T_L(F; \delta)\} = \frac{m!n!}{(m+n)!(1-\delta)^n} \mathbf{P}\{U_{i_1+1:m+n} > \delta\}$$

$$= \frac{(n-1)!(m-i_1)!}{(m+n-i_1-1)!} \mathbf{P}\{S_{(1)} = i_1 | G = T_L(F; \delta)\} \ ,$$

2. $$\mathbf{P}\{\mathbf{S} = \mathbf{i} | G = T_R(F; \delta)\} = \frac{m!n!}{(m+n)!(1-\delta)^n} \mathbf{P}\{U_{i_n+n:m+n} < 1 - \delta\}$$

$$= \frac{(n-1)!(i_n)!}{(i_n+n-1)!} \mathbf{P}\{S_{(n)} = i_n | G = T_R(F; \delta)\}$$

3. $$\mathbf{P}\{\mathbf{S} = \mathbf{i} | F = T_L(G; \delta)\} = \frac{m!n!}{(m+n)!(1-\delta)^m} \mathbf{P}\{U_{j_1+1:m+n} > \delta\}$$

$$= \frac{(m-1)!(n-j_1)!}{(m+n-j_1-1)!} \mathbf{P}\{S^*_{(1)} = j_1 | F = T_L(G; \delta)\} \ ,$$

4. $$\mathbf{P}\{\mathbf{S} = \mathbf{i} | F = T_R(G; \delta)\} = \frac{m!n!}{(m+n)!(1-\delta)^m} \mathbf{P}[U_{j_m+m:m+n} < 1 - \delta\}$$

$$= \frac{(m-1)!(j_m)!}{(j_m+m-1)!} \mathbf{P}\{S^*_{(m)} = j_m | F = T_R(G; \delta)\} \ ,$$

Proof: We have, using Corollary A.2 and (11),

$$\mathbf{P}\{\mathbf{S} = \mathbf{i} | G = T_L(F; \delta)\}$$

$$= \frac{m!n!}{(m+n)!}\mathbf{E}\left\{\prod_{k=1}^{n} T_L'(U_{i_k+k:m+n}; \delta)\right\}$$

$$= \frac{m!n!}{(m+n)!}\mathbf{E}\left\{\prod_{k=1}^{n} [(1-\delta)^{-1} I_{(U_{i_k+k:m+n} > \delta)}]\right\}$$

$$= \frac{m!n!}{(m+n)!(1-\delta)^n}\mathbf{P}\{U_{i_1+1:m+n} > \delta\} \ .$$

Therefore,

$$P\{S_{(1)} = i_1 | G = T_L(F; \delta)\}$$

$$= \frac{(m+n-i_1-1)!m!n!}{(m+n)!(n-1)!(m-i_1)!(1-\delta)^n}\mathbf{P}\{U_{i_1+1:m+n} > \delta\}$$

since the number of elements of $\{\mathbf{i} \in \mathbf{A}$: for given $i_1\}$ is $\binom{m+n-i_1-1}{n-1}$. Hence the proof of part 1 of the theorem is completed. The proof of part 2 is similar. The proofs of parts 3 and 4 follow easily by symmetry and by replacing $n \to m$, $i_1 \to j_1$, $i_n \to j_m$ and $S_{(i)} \to S^*_{(j)}$.

2.1 *UMP Rank Test Against* H_a

Theorem 2.1 (part 1) shows that the probability function of **S**, under the alternative H_a: $G = T_L(F; \delta)$, $1 \geq \delta > 0$, depends on i_1, the number of X's less than $Y_{(1)}$. The probability function is increasing in i_1 for a fixed δ since

$$U_{r:m+n} > U_{r':m+n} \qquad \text{for all } m+n \geq r > r' \geq 1 \ .$$

Therefore, a UMP rank test under H_a exists and is based on $S_{(1)}$.

Similarly, the second part of Theorem 2.1 shows that the probability function of **S**, under the alternative H_a: $G = T_R(F; \delta)$, $1 \geq \delta > 0$, depends on i_n only and is decreasing in i_n for a fixed δ. Therefore, the UMP rank test in this case exists and is based on $S_{(n)}$.

3 Distribution of S Under the Alternative H_b

Since there is no functional relationship between F and G in H_b, using Corollary A.2 and (11) and (12), we have

$$\mathbf{P}\{\mathbf{S}=\mathbf{i}\} = \frac{m!n!}{(m+n)!} E\left\{\prod_{k=1}^{m} T_R'(U_{i_k+k};\,\delta_R)\prod_{k=1}^{n} T_L'(U_{j_k+k};\,\delta_L)\right\}$$

$$= c_0 \mathbf{E}\{I_{(\delta_L < U_{i_1+1:\,m+n})} I_{(U_{j_m+m:\,m+n}) < 1-\delta_R)}\} \tag{13}$$

where $c_0 = \left[\binom{m+n}{m}(1-\delta_R)^m(1-\delta_L)^n\right]^{-1}$.

Theorem 3.1: Let $F = T_R(H;\,\delta_R)$ and $G = T_L(H;\,\delta_L)$ where H is a continuous distribution function, $\delta_R,\ \delta_L \geq 0$ and $\delta_R + \delta_L < 1$.

1. For $\mathbf{i} = (m, \ldots, m)$,

$$\mathbf{P}\{\mathbf{S}=\mathbf{i}\} = P\{S_{(1)} = m,\ S^*_{(m)} = 0\}$$

$$= c_0[\mathbf{P}\{U_{m+1:\,m+n} > \delta_L\} - \mathbf{P}\{U_{m:\,m+n} > 1-\delta_R\}]$$

where $c_0 = \left[\binom{m+n}{m}(1-\delta_R)^m(1-\delta_L)^n\right]^{-1}$.

2. For each $\mathbf{i} \in \mathbf{A}$ such that $i_1 < m$,

$$\mathbf{P}\{\mathbf{S}=\mathbf{i}\} = \frac{(j_m-1)!(m-i_1-1)!}{(j_m+m-i_1-2)!}\mathbf{P}\{S_{(1)} = i_1,\ S^*_{(m)} = j_m\}$$

$$= c_0 \mathbf{P}\{\delta_L < U_{i_1+1:\,m+n} < U_{j_m+m:\,m+n} < 1-\delta_R\}$$

$$= c_1 \sum_{k=0}^{n-j_m}\left[\binom{m+n}{n-j_m-k}\left(\frac{\delta_R}{1-\delta_R}\right)^{n-j_m-k} * \mathbf{P}\left\{U_{i_1+1:\,j_m+m+k} > \frac{\delta_L}{1-\delta_R}\right\}\right]$$

where $\mathbf{j} \in \mathbf{A}^*$ is the dual vector of $\mathbf{i}$, c_0 is defined above in the first part and

$$c_1 = \left(\frac{1-\delta_R}{1-\delta_L}\right)^n \Big/ \binom{m+n}{m}.$$

Proof 1: If all the X observations are less than $Y_{(1)}$, $\mathbf{i} = (m, \ldots, m)$ and its dual vector $\mathbf{j} = (0, \ldots, 0)$. From (13), we have

$$\mathbf{P}\{\mathbf{S} = \mathbf{i}\} = c_0 \mathbf{P}\{\delta_L < U_{m+1:\, m+n},\ U_{m:\, m+n} < 1 - \delta_R\} \ .$$

and the first part of the theorem follows immediately.

Proof 2: As $i_1 < m$, there is at least one X observation larger than $Y_{(1)}$, and $\mathbf{j} \in \mathbf{A}^*$, the dual vector of $\mathbf{i}$ is such that $j_m > 1$ and $j_m + m > i_1 + 1$. Hence, from (13),

$$\mathbf{P}\{\mathbf{S} = \mathbf{i}\} = c_0 \mathbf{P}\{\delta_L < U_{i_1+1:\, m+n} < U_{j_m+m:\, m+n} < 1 - \delta_R\} \ ,$$

and the proof is completed using the following proposition.

Proposition 3.1: For $0 < a < b < 1$ and $1 \leq i < j \leq N$.

$$\mathbf{P}\{a < U_{i:N} < U_{j:N} < b\}$$

$$= b^N \sum_{k=0}^{N-j} \binom{N}{N-j-k} \left(\frac{1-b}{b}\right)^{N-j-k} \mathbf{P}\left(\frac{a}{b} < U_{i:\, j+k}\right)$$

$$= b^N \sum_{k=0}^{N-j} \binom{N}{N-j-k} \left(\frac{1-b}{b}\right)^{N-j-k} B\left(1 - \frac{a}{b}; j+k-i+1, i\right) \ .$$

Proof: It is known that $U_{i:N}/U_{j:N}$ and $U_{j:N}$ are independent (for $i < j$) and the conditional distribution of $U_{i:N}/u$, given $U_{j:N} = u$, is the same as that of $U_{i:\, j-1}$, David (1981). Now setting $U_{i:\, j-1}$ and $U_{j:N}$ to be independent, we can write

$$P\{a < U_{i:N} < U_{j:N} < b\}$$

$$= E[P\{a < U_{i:N} | U_{j:N}\} I_{(a < U_{j:N} < b)}]$$

$$= E\left[P\left\{\frac{a}{U_{j:N}} < U_{i:\, j-1}\right\} I_{(a < U_{j:N} < b)}\right]$$

$$= \int_a^b \left[\int_{a/u}^1 \frac{v^{i-1}(1-v)^{j-i-1}}{B(i, j-i)}\, dv\right] \frac{u^{j-1}(1-u)^{N-j}}{B(j, N-j+1)}\, du \ .$$

Let $w = u/b$ in the above integration, and then $1 - u = b[1 - w + (1 - b)/b]$ and the result follows after integration.

When $\delta_L = \delta_R = \delta$, i.e., the truncation rate is the same on each side, $\mathbf{P}\{\mathbf{S} = \mathbf{i}\}$ takes a simple form given by the next corollary.

Corollary 3.1: Suppose $\delta_L = \delta_R = \delta$ and $\delta \in [0; 0.5)$, then

$$\mathbf{P}\{\mathbf{S} = \mathbf{i}\} = \sum_{k=0}^{n-j_m} \binom{m+n}{n-j_m-k} \Delta^{n-j_m-k} \mathbf{P}\{U_{i_1+1:\, j_m+m+k} > \Delta\} \Big/ \binom{m+n}{m},$$

where $\Delta = \delta/(1 - \delta)$, $U_{m+1:\, m} \equiv 1$ and $\mathbf{P}(U_{m+1:\, m} > \Delta) = 1$

Proof: For $\mathbf{i} \in \mathbf{A}$ with $i_1 < m$, the result of the theorem is obvious because of

Theorem 3.1: Assume $i_1 = m$. Then from part 1 of Theorem 3.1,

$$\mathbf{P}\{\mathbf{S} = i\} = c_0[\mathbf{P}\{U_{m+1:\, m+n} > \delta\} - \mathbf{P}\{U_{m:\, m+n} > 1 - \delta\}]$$

where $c_0 = \left[\binom{m+n}{m}(1 - \delta)^{m+n}\right]^{-1}$. Since

$$\mathbf{P}\{U_{m:\, m+n} > 1 - \delta\} = \int_{1-\delta}^{1} \frac{u^{m-1}(1-u)^n}{B(m, n+1)}\, du$$

$$= -\delta^n(1 - \delta)^m \binom{m+n}{m} + P\{U_{m+1:\, m+n} > 1 - \delta\},$$

we have

$$\mathbf{P}\{\mathbf{S} = \mathbf{i}\} = \left(\frac{\delta}{1-\delta}\right)^n + c_0 \int_{\delta}^{1-\delta} \frac{u^m(1-u)^{n-1}}{B(m+1, n)}\, du$$

$$= \Delta^n + \int_{\Delta}^{1} \frac{x^m[(1-x) + \Delta]^{n-1}}{B(m+1, n)}\, dx \Big/ \binom{m+n}{m}.$$

Letting $x = u/(1 - \delta)$ and $\Delta = \delta/(1 - \delta)$, and integrating we see

$$\mathbf{P}\{\mathbf{S} = \mathbf{i}\} = \Delta^n + \sum_{k=1}^{n} \binom{m+n}{n-k} \Delta^{n-k} P\{U_{m+1:m+k} > \Delta\} \Big/ \binom{m+n}{m}$$

$$= \sum_{k=0}^{n} \binom{m+n}{n-k} \Delta^{n-k} P\{U_{m+1:m+k} > \Delta\} \Big/ \binom{m+n}{m} .$$

where we set $U_{m+1:m} = 1$, $P(U_{m+1:m} > \Delta) = 1$.
Since $i_1 = m$ implies $j_m = 0$, the result follows.
The above expression is used to compute exact powers later on.

The sample space of $(S_{(1)}, S^*_{(j_m)})$ is

$$\mathbf{B} = \{(m, 0)\} \cup \{(i, j)\colon i = 0, \ldots, m-1 \text{ and } j = 1, \ldots, n\} . \tag{14}$$

Therefore, from Theorem 3.1, the probability function of $\mathbf{S}$ under the alternative H_b has at most $mn + 1$ different values. In the next section we give the ordering of the $mn + 1$ probability values for some cases, that will be useful for defining some LMP rank tests in Section 4.

3.1 Ranking the Probability Values of S Under H_b

It was relatively easy to find a UMP test under H_a. However, under H_b and H_c we consider the following basic method of finding a LMP rank test. For example, the method is described for testing H_0: $\delta = 0$ versus H_b: $\delta \in (0, \delta_0)$. Under H_0 each rank configuration is equally likely. Hence a critical region, for an $\alpha = k \Big/ \binom{m+n}{m}$ level test, consists of k vectors of the ranks of $(Y_{(1)}, \ldots, Y_{(n)})$. To make such a test most powerful against a particular $\delta > 0$, the critical region should consist of k rank configurations with the largest probabilities for the particular δ. Since the most powerful test for $\delta > 0$ must have this property for all its natural α-level tests the order in which it puts the rank configurations in the critical region must coincide with ordering the configurations by their probabilities under the alternative $\delta > 0$. Thus the rank configuration with the largest probability under H_a must be placed *first* in the critical region and so on. Thus ranking of the probability values is useful in searching for LMP tests.

Theorem 3.2: Define for each $(i_1, j_m) \in \mathbf{B}$,

$$f_{i_1, j_m}(\delta_L, \delta_R) = \mathbf{P}\{\delta_L < U_{i_1+1:m+n}, U_{j_m+m:m+n} < 1 - \delta_R) \tag{15}$$

where **B** is defined in (14). Then we have the following results.

1. For all $\delta_L, \delta_R \geq 0$, $\delta_L + \delta_R < 1$,

$$f_{m,0}(\delta_L, \delta_R) \geq f_{m-1,1}(\delta_L, \delta_R)$$

and the equality holds if and only if $\delta_L = \delta_R = 0$.

2. For $i_1 < m, j_m > 0$ and for all $\delta_L, \delta_R \geq 0$, $\delta_L + \delta_R < 1$.

   a) $f_{i_1,j_m}(\delta_L, \delta_R) \geq f_{i_1-1,j_m}(\delta_L, \delta_R)$ provided $i_1 > 0$ and the equality holds if and only if $\delta_L = 0$;
   b) $f_{i_1,j_m}(\delta_L, \delta_R) \geq f_{i_1,j_m+1}(\delta_L, \delta_R)$ provided $j_m < n$ and the equality holds if and only if $\delta_R = 0$.

3. For each l, h, h' satisfying $0 \leq l \leq \min\{m-2, n-2\}$, $1 \leq h + l$, $h' + l \leq \min\{m-1, n-1\}$, and $h, h' > 0$, we have

   a) $f_{l,n-l-h}(\delta_L, \delta_R) = f_{l+h,n-l}(\delta_L, \delta_R)$ if $\delta_L = \delta_R$,
   b) $f_{l,n-l-h}(\delta_L, \delta_R) < f_{l+h',n-l}(\delta_L, \delta_R)$ if $\delta_L < \delta_0$, $\delta_R = c\delta_L$ for some $0 < c < 1$ and $\delta_0 > 0$,
   c) $f_{l,n-l-h}(\delta_L, \delta_R) > f_{l+h',n-l}(\delta_L, \delta_R)$ if $\delta_L < \delta_0$, $\delta_R = c\delta_L$ for some $c > 1$ and $\delta_0 > 0$.

4. For each l with $0 < l \leq \min\{n-1, m-1\}$, we have

$$f_{l,n-l}(\delta_L, \delta_R) > f_{m-1,n-l+1}(\delta_L, \delta_R)$$

and

$$f_{l,n-l}(\delta_L, \delta_R) > f_{l-1,1}(\delta_L, \delta_R)$$

if $\delta_L < \delta_0$, $\delta_R = c\delta_L$ for some $c \neq 0$ and $\delta_0 > 0$.

5. For each h, k satisfying $0 \leq h < m + n - 1$ and $h - n < k \leq \min\{h, m-1\}$, we have

$$\lim_{\delta_0 \to .5^+} \frac{f_{k,n-h+k}(\delta_0, \delta_0)}{(.5 - \delta_0)^{m+n-h}} = c_1 \binom{h}{k}$$

where c_1 is positive and independent of k.

In general, for each $c > 0$,

$$\lim_{\delta_L \to (1/(1+c))^+} \frac{f_{k,n-h+k}(\delta_L, c\delta_L)}{[(1/(c+1)) - \delta_L]^{m+n-h}} = c_1 c^{h-k} \binom{h}{k}$$

where c_1 is positive and independent of k.

Proof:

1. $f_{m,0}(\delta_L, \delta_R) = \mathbf{P}\{U_{m+1:m+n} > \delta_L, U_{m:m+n} < 1 - \delta_R\}$

   $f_{m-1,1}(\delta_L, \delta_R) = \mathbf{P}\{\delta_L < U_{m:m+n} < U_{m+1:m+n} < 1 - \delta_R\}$.

   Since

   $$\{\delta_L < U_{m:m+n} < U_{m+1:m+n} < 1 - \delta_R\}$$
   $$\subset \{\delta_L < U_{m+1:m+n}, U_{m:m+n} < 1 - \delta_R\} ,$$

   part 1 of Theorem 3.2 follows.
2. For $i_1 < m$ we have using (15)

   $$f_{i_1,j_m}(\delta_L, \delta_R) = \mathbf{P}\{\delta_L < U_{i_1+1:m+n} < U_{j_m+m:m+n} < 1 - \delta_R\} .$$

   Similarly, we consider expressions for $f_{i_1-1,j_m}(\delta_L, \delta_R)$ and $f_{i_1,j_m+1}(\delta_L, \delta_R)$. Since

   $$\{\delta_L < U_{i_1:m+n} < U_{j_m+m:m+n} < 1 - \delta_R\}$$
   $$\subset \{\delta_L < U_{i_1+1:m+n} < U_{j_m+m:m+n} < 1 - \delta_R\} ,$$

   and

   $$\{\delta_L < U_{i_1+1:m+n} < U_{j_m+m+1:m+n} < 1 - \delta_R\}$$
   $$\subset \{\delta_L < U_{i_1+1:m+n} < U_{j_m+m:m+n} < 1 - \delta_R\} ,$$

   part 2 of Theorem 3.2 follows.

3. $f_{l,n-l-h}(\delta_L, \delta_R) = \mathbf{P}\{\delta_L < U_{l+1:m+n} < U_{m+n-l-h:m+n} < 1 - \delta_R\}$

$$= \mathbf{P}\{\delta_R < U_{l+h+1:m+n} < U_{m+n-l:m+n} < 1 - \delta_L\}$$

$$= f_{l+h,n-l}(\delta_R, \delta_L) \ ,$$

the result (a) of this part follows. To prove (b) and (c), we need the following proposition.

Proposition 3.2: For $1 \leq r < s \leq t$,

(i) $\delta^{-r} P\{U_{r:t} < \delta\} \to \binom{t}{r}$, as $\delta \to 0^+$;

(ii) $\delta_L^{-r}\delta_R^{-(t-s+1)}\mathbf{P}\{U_{r:t} < \delta_L, U_{s:t} > 1 - \delta_R\} \to \dfrac{t!}{r!(s-r-1)!(t-s+1)!}$ as $\delta_L, \delta_R \to 0^+$.

Therefore, for $\delta_R = c\delta_L$, $c \neq 0$, we have

$$f_{l,n-l-h}(\delta_L, c\delta_L) = 1 - \mathbf{P}\{U_{l+1:m+n} < \delta_L\} - O(\delta_L^{l+h+1}) \ ,$$

$$f_{l+h',n-l}(\delta_L, c\delta_L) = 1 - \mathbf{P}\{U_{l+1:m+n} < c\delta_L\} - O(\delta_L^{l+h'+1}) \ .$$

Hence, the last two results of part 3 of Theorem 3.2 follow.

4. For $\delta_R = c\delta_L$, $c \neq 0$, using (15) and Proposition 3.1 we see,

$$f_{l,n-l}(\delta_L, c\delta_L) = 1 - O(\delta_L^{l+1})$$

$$f_{l-1,1}(\delta_L, c\delta_L) = 1 - \mathbf{P}\{U_{l:m+n} < \delta_L\} - O(\delta_L^n) = 1 - O(\delta_L^l)$$

$$f_{m-1,n-l+1}(\delta_L, c\delta_L) = 1 - P\{U_{l:m+n} < c\delta_L\} - O(\delta_L^m) = 1 - O(\delta_L^l) \ .$$

Then part 4 of the theorem follows.

5. Let $\rho = [1/(c+1)] - \delta_L$. Then using (15) and l'Hospital's rule we have:

$$\lim_{\delta_L \to (1/1+c)^+} \frac{f_{k,n-h+k}(\delta_L, c\delta_L)}{[(1/(c+1)) - \delta_L]^{m+n-h}}$$

$$= \lim_{\rho \to 0^+} \frac{\mathbf{P}\left\{\dfrac{1}{1+c} - \rho < U_{k+1:m+n} < U_{m+n-h+k:m+n} < \dfrac{1}{1+c} + c\rho\right\}}{\rho^{m+n-h}}$$

$$= \lim_{\rho \to 0^+} \frac{\int_{(1/1+c)-\rho}^{(1/1+c)+c\rho} v^k \left(\frac{1}{1+c} + c\rho - v\right)^{m+n-h-2} \left(1 - \frac{1}{1+c} - c\rho\right)^{h-k} dv}{(m+n-h-2)!k!(h-k)!(m+n-h)\rho^{m+n-h-1}} (m+n)!$$

$$+ \lim_{\rho \to 0^+} \frac{\int_{(1/1+c)-\rho}^{(1/1+c)+c\rho} \left(\frac{1}{1+c} - \rho\right)^k \left(u - \frac{1}{1+c} + \rho\right)^{m+n-h-2} (1-u)^{h-k} du}{(m+n-h-2)!k!(h-k)!(m+n-h)\rho^{m+n-h-1}} (m+n)!$$

$$= 2 \lim_{\rho \to 0^+} \frac{\left(\frac{1}{1+c} - \rho\right)^k [(1+c)\rho]^{m+n-h-2} \left(1 - \frac{1}{1+c} - c\rho\right)^{h-k}}{(m+n-h)!k!(h-k)!\rho^{m+n-h-2}} (m+n)!$$

$$= 2\binom{m+n}{h}\binom{h}{k}(c+1)^{m+n-2h-2}c^{h-k}$$

Therefore, the results of part 5 follow and the proof of Theorem 3.2 is completed.

Proof of Proposition 3.2:
To prove (i), we note that

$$\delta^{-r}\mathbf{P}(U_{r:t} < \delta) = \frac{t!}{(r-1)!(t-r)!} \sum_{k=0}^{t-r} \binom{t-r}{k} (-1)^k \frac{\delta^k}{(k+r)}$$

To prove (ii) we see

$$\delta^{-r}\delta_R^{-(t-s+1)}\mathbf{P}(U_{r:t} < \delta_L, U_{s:t} > 1 - \delta_R)$$

$$= \frac{t!}{(r-1)!(s-r-1)!(t-s)!} \sum_{k=0}^{s-r-1} \binom{s-r-1}{k} (-1)^k$$

$$\times \sum_{i=1}^{k} \binom{k}{i}^k \frac{\delta_L^{k-i}\delta_R^i}{(r+k-i)(t-s+i+1)}$$

and take the limit as $\delta_L \to 0^+$ and $\delta_R \to 0^+$.

Theorem 3.2 gives a method of ranking the probability values of **S** in various cases. It is useful to define rank (score) $V(i,j)$ of $f_{i,j}(\delta,\delta)$ in $\{f_{i_1,j_m}(\delta,\delta)$: $(i_1,j_m) \in \mathbf{B}\}$. The following are two examples where we assume $\delta_L = \delta_R = \delta$ and the scores satisfy the conditions of Theorem 3.2.

Case A: Assume $\delta_L = \delta_R = \delta$, $\delta \in (0, \delta_0)$ where δ_0 is sufficiently small. For each $\delta \in (0, \delta_0)$, define $V(i, j)$ such that

a) $V(0, n) = 1$;
b) for $m \geq n$ and $j = 1, \ldots, n$,

$$V(i + 1, j) = 1 + V(i, j) \qquad \text{if } n - j \geq i < m - 1 \ ,$$

$$V(n - j + 1, j - 1) = 1 + V(m - 1, j) \qquad \text{provided } j > 1 \ ;$$

c) for $m < n$ and $i = 0, \ldots, m - 1$,

$$V(i, j - 1) = 1 + V(i, j) \qquad \text{for } 1 < j \leq n - i \ ,$$

$$V(i + 1, n - i - 1) = 1 + V(i, 1) \qquad \text{provided } i < m - 1 \ ;$$

d) $V(i, j) = V(n - j, n - i)$ for $i, j \leq \min\{m, n\}$;
e) $V(m, 0) = 1 + V(m - 1, 1)$.

Example 3.1: For given $m = 4$ and $n = 5$, assign the $V(i, j)$ as follows:

		j					
		0	1	2	3	4	5
i	0		5	4	3	2	1
	1		9	8	7	6	2
	2		12	11	10	7	3
	3		14	13	11	8	4
	4	15					

Case B: Given some $\delta_0 < .5$ which is close enough to .5, for each $\delta \in (\delta_0, .5)$, define the rank $V(i, j)$ of $f_{i,j}(\delta, \delta)$ in $\{f_{i_1, j_m}(\delta, \delta)\colon (i_1, j_m) \in \mathbf{B}\}$ such that

a) $V(0, n) = 1$;
b) for $m \geq n$ and for $0 < i < m$,

$$V(i, n) = 1 + V(l, n + l - i + 1)$$

where l is the least integer greater than or equal to $\max\{(i - 1)/2, (i - n)\}$;

c) for $m \geq n$ and for $0 < j < n$,

$$V(m-1, j) = 1 + V(l, l + j - m + 2)$$

where l is the least integer greater than or equal to $\max\{(m + n - j)/2 - 1, (m - j - 1)\}$;

d) for $m \geq n$, $1 < h < m + n - 3$ and $\max\{(h/2) + 1, (h + 2 - n)\} \leq i \leq \min\{h, m - 1\}$,

$$V(i - 1, n + i - 1 - h) = 1 + V(i, n + i - h) \ ;$$

e) for $m < n$ and for $0 < j < n$,

$$V(0, j) = 1 + V(l, j + l + 1)$$

where l is the greatest integer less than or equal to $\min\{(n - j - 1)/2, (m - 1)\}$;

f) for $m < n$ and for $0 < i < m$,

$$V(i, 1) = 1 + V(l, l - i + 2)$$

where l is the greatest integer less than or equal to $\min\{(n + i)/2 - 1, (m - 1)\}$;

g) for $m < n$, $1 < h < m + n - 3$ and $0 \leq i \leq \min\{(h/2) - 1, m - 1\}$;

$$V(i + 1, n + i + 1 - h) = 1 + V(i, m + i - h) \ ;$$

h) $V(i, j) = V(n - j, n - i)$ for $i, j \leq \min\{m, n\}$;

i) $V(m, 0) = 1 + V(m - 1, 1)$.

Example 3.2: For given $m = 4$ and $n = 5$, assign the $V(i, j)$ as follows:

i \ j	0	1	2	3	4	5
0		7	5	3	2	1
1		10	8	6	4	2
2		12	11	9	6	3
3		14	13	11	8	5
4	15					

4 Some Rank Tests for the Alternative H_b

From Theorem 3.2, it is obvious that for $m, n \geq 2$, there does not always exist a UMP rank test for testing H_0 against H_b. However, the results in part 1 and part 2 in Theorem 3.2 indicate that a rank test should have the following property: If for some (i_1, j_m), $\{(S_{(1)}, S^*_{(m)}) = (i_1, j_m)\}$ is in the rejection region of the test, then for any $i'_1 \geq i_1$, $j'_m \geq j_m$, $\{(S_{(1)}, S^*_{(m)}) = (i'_1, j'_m)\}$ is also in the rejection region. Therefore, if we define some score function $V(i, j)$, $(i, j) \in \mathbf{B}$, the sample space of $(S_{(1)}, S^*_{(m)})$, so that

i) $V(m, 0) \geq V(m-1, 1)$,
ii) $V(i, j) \geq V(i-1, j)$ if $0 < i < m$ and $0 < j \leq n$,
iii) $V(i, j) \geq V(i, j+1)$ if $0 \leq i < m$ and $0 < j < n$,

then $V(S_{(1)}, S^*_{(m)})$ would provide a reasonable rank test. The following are some score functions with these properties.

1. $V_1(i, j) = i$ for each (i, j).
2. $V_2(i, j) = n - j$ for each (i, j).
3. $V_3(i, j) = \max\{i, n - j\}$ for each (i, j).
4. $V_4(i, j) = \min\{i, n - j\}$ for each (i, j).
5. $V_5(i, j) = i + n - j$ for each (i, j).
6. $V_6(i, j) = V_4(i, j) + V_5(i, j)$ for each (i, j).
7. Let $V_7(i, j)$ be the rank $V(i, j)$ of $f_{i,j}$ defined in Case A in Section 3.1.
8. Let $V_8(i, j)$ be the rank $V(i, j)$ of $f_{i,j}$ defined in Case B in Section 3.1.

V_1 and V_2 were first considered by Rosenbaum (1953). We like to note that $V_4(S_{(1)}, S^*_{(m)})$ and $V_5(S_{(1)}, S^*_{(m)})$ are, respectively, the Hajek test statistic and the Sidak and Vondracek (1957) test statistic for testing H_0: $F = G$ against H^*_b: $G(x) = F(x - \Delta)$, $\Delta \in (0, 1)$ where $F \sim U(0, 1)$. Also note that if H is the d.f. of $U(0, 1 + \Delta)$, then the alternative is equivalent to H_a: $F = T_R(H; \delta)$, $G = T_L(H; \delta)$ where $\delta = \Delta/(1 + \Delta)$. It has been indicated that the Hajek test and Sidak and Vondracek test are the LMP rank tests, respectively, in some neighborhoods of $\Delta = 0$ and $\Delta = 1$, for a particular choice of α. However, the subsequent examples will show that the two tests are not the LMP tests for a general choice of α.

4.1 LMP Rank Test for H^*_b: $F = T_R(H; \delta, \delta)$, $G = T_L(H; \delta, \delta)$, $\delta \in (0, .5)$

Let us consider the alternative, H_a: $F = T_R(H; \delta, \delta)$, $G = T_L(H; \delta, \delta)$ for $\delta \in (0, .5)$. From the results of Cases A and B in Section 3.1, it is obvious that $V_7(S_{(1)}, S^*_{(m)})$ and $V_8(S_{(1)}, S^*_{(m)})$ generate LMP rank tests in the neighborhoods of $\delta = 0$ and $\delta = .5$, respectively.

Example 4.1: For $m = 4$, $n = 5$ and $\alpha = 7/126$, from Example 3.1, the LMP rank test in the neighborhood of $\delta = 0$ rejects H_0 if $V_7(S_{(1)}, S^*_{(m)}) \geq 11$ and from Example 3.2, the LMP rank test in the neighborhood of $\delta = .5$ rejects H_0 if $V_8(S_{(1)}, S^*_{(m)}) \geq 11$. In the case, the two LMP rank tests are the same and they reject H_0 if $(S_{(1)}, S_{(m)}) \in \{(4, 0), (3, 1), (3, 2), (3, 3), (2, 1), (2, 2)\}$. Actually, the LMP tests in the case (for $m = 4$, $n = 5$ and $\alpha = 7/126$) are the UMP rank tests.

For $m = 4$ and $n = 5$, the score values of $V_4(i, j)$ and $V_5(i, j)$ are as follows:

Score values of $V_4(i,j)$

i \ j	0	1	2	3	4	5
0		0	0	0	0	0
1		1	1	1	1	0
2		2	2	2	1	0
3		3	3	2	1	0
4	4					

Score values of $V_5(i,j)$

i \ j	0	1	2	3	4	5
0		4	3	2	1	0
1		5	4	3	2	1
2		6	5	4	3	2
3		7	6	5	4	3
4	8					

For the given significance level $\alpha = 7/126$, the Hajek test and the Sidak and Vondracek test are defined, respectively, by

$$\phi_H = \begin{cases} 1 & \text{if } V_4(S_{(1)}, S^*_{(m)}) > 2 \\ 4/7 & \text{if } V_4(S_{(1)}, S^*_{(m)}) = 2 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\phi_{SV} = \begin{cases} 1 & \text{if } V_5(S_{(1)}, S^*_{(m)}) > 5 \\ 3/4 & \text{if } V_5(S_{(1)}, S^*_{(m)}) = 5 \\ 0 & \text{otherwise} \end{cases}$$

Table 1 shows the power functions of those tests and the Mann-Whitney-Wilcoxon test (MWW test). As we are considering $\delta_L = \delta_R$ we use corollary 3.1 to compute these powers. The table shows that the power of the MWW test is quite comparable with that of others even for small values of Δ.

Table 1. Power Comparison of the Hajek test, the Sedak-Vondracek test, LMP test and the Mann-Whitney-Wilcoxen test for different values of $\Delta = \delta/(1 + \delta)$

	$\Delta = .03$	$\Delta = .05$	$\Delta = .10$	$\Delta = .20$	$\Delta = .40$	$\Delta = .60$	$\Delta = .80$
Hajek	.07238	.08564	.12603	.23595	.53045	.81154	.97018
SV	.07221	.08525	.12513	.23744	.55987	.85951	.98776
LMP	.07241	.08574	.12719	.24484	.57775	.87341	.99027
MWW	.07215	.08508	.12444	.23497	.55391	.85488	.98693

5 The Distribution of S Under the Alternative H_c

Let $G = T_2(F; \delta_L, \delta_R)$ where $\delta_R, \delta_L \geq 0$ and $\delta_R + \delta_L < 1$. Then from Corollary A.2, we have for $n = 1$ and for each $i = 0, \ldots, m$,

$$P\{S_{(1)} = i\} = [(1 - \delta_L - \delta_R)(m + 1)]^{-1} P\{\delta_L < U_{i+1:m+1} < 1 - \delta_R\}$$

and for $n > 1$, and for each $\mathbf{i} \in \mathbf{A}$,

$$\mathbf{P}\{\mathbf{S} = \mathbf{i}\} = c_0 \mathbf{P}\{\delta_L < U_{i_1+1:m+n} < U_{i_n+n:m+n} < 1 - \delta_R\}$$

where

$$c_0 = \left[\binom{m+n}{m}(1 - \delta_R - \delta_L)^n\right]^{-1} .$$

Therefore, for $n > 1$, we can show

$$\mathbf{P}\{\mathbf{S} = \mathbf{i}\} = \mathbf{P}\{S_{(1)} = i_1, S_{(n)} = i_n\} \Big/ \binom{i_n - i_1 + n - 2}{n - 2}$$

$$= c_1 \sum_{k=0}^{m-i_n} \binom{m+n}{m - i_n - k} \left(\frac{\delta_R}{1 - \delta_R}\right)^{m-i} n^{-k} \mathbf{P}\left\{U_{i_1+1:i_n+n+k} > \frac{\delta_L}{1 - \delta_R}\right\} ,$$

where

$$c_1 = c_0(1 - \delta_R)^{m+n} .$$

Note that $\mathbf{P}(\mathbf{S} = \mathbf{i})$ depends on i_1 and i_n only.
Define

$$f_{i_1,i_n}(\delta_L, \delta_R) = \mathbf{P}\{\mathbf{S} = \mathbf{i} | G = T_2(F; \delta_L, \delta_R)\} \ , \tag{16}$$

Following the methods discussed for the alternative H_b, we can consider ranking of

$$f_{i_1,i_n}(\delta_L, \delta_L) = P(\mathbf{S} = \mathbf{i} | G = T_2; \delta_L \delta_R)$$

and define some score functions. Details of these are given in Lin and Sukhatme (1991). An UMP rank test does not always exist. We can find LMP tests under the condition $\delta_L = \delta_R = \delta$. For example:

(1) Let $V_1(i_1, i_n)$ be the rank of $f_{i_1,i_n}(\delta, \delta)$ in the neighborhood of $\delta = 0$.
(2) Let $V_2(i_1, i_n)$ be the rank of $f_{i_1,i_n}(\delta, \delta)$ in the neighborhood of $\delta = .5$.

Now V_1 and V_2 generate the LMP rank tests in the neighborhoods of $\delta = 0$ and $\delta = .5$ respectively.

Acknowledgements: The authors wish to thank the referees for their helpful comments on a previous version of this paper.

Appendix A: Hoeffding Type Theorem

Theorem A.1: Let $X = (X_{11}, \ldots, X_{1n_1}; X_{21}, \ldots, X_{2n_2}; \ldots; X_{k1}, \ldots, X_{kn_k})$ be a random vector, with the joint d.f. $W(H(x_{11}), \ldots, H(x_{kn_k}))$ where $H(x)$ is a continuous distribution function and $W(t_{11}, \ldots, t_{kn_k})$ is an absolutely continuous distribution function on $\{(t_{11}, \ldots, t_{kn_k}): 0 < t_{ij} < 1, \ i = 1, \ldots, k, \ j = 1, \ldots, n_i\}$ and W is invariant under all permutations within $(t_{i1}, \ldots, t_{in_i})$, $(i = 1, \ldots, k)$. For each (i, j), let $R_{i(j)}$ be the rank of $X_{i(j)}$ in the pooled sample $\{X_{11}, \ldots, X_{kn_k}\}$, where $X_{i(j)}$ is the j-th order statistic (the j-th smallest value) in $\{X_{i1}, \ldots, X_{in_i}\}$. Denote $\mathbf{R} = (R_{1(1)}, \ldots, R_{k(n_k)})$, $N = \sum_{i=1}^{k} n_i$ and R is the set of

$$\mathbf{r} = (r_{11}, \ldots, r_{1n_1}; \ldots; r_{k1}, \ldots, r_{kn_k}) \ ,$$

a permutation of $(1, \ldots, N)$ satisfying $r_{11} < \cdots < r_{1n_1}; \ldots; r_{k1} < \cdots < r_{kn_k}$.

Then for each $\mathbf{r} \in R$, we have

$$P\{R = r\} = \frac{n_1! \dots n_k!}{N!} E\{w(U_{r_{11}:N}, \dots, U_{r_{kn_k}:N})\} \tag{A.1}$$

where w is the density function of W and $U_{1:N}, \dots, U_{N:N}$ are the order statistics of a random sample of size N from $U(0, 1)$, the uniform distribution on $(0, 1)$.

Proof: Let

$$A_r = \{(x_{11}, \dots, x_{kn_k}): \text{the rank of } x_{i(j)} \text{ is } r_{ij}, \text{ for each } i, j\} ,$$

$$B_r = \{(x_{1(1)}, \dots, x_{1(n_1)}, \dots, x_{k(1)}, \dots, x_{k(n_k)}):$$

$$x_{1(1)} < \cdots < x_{1(n_1)}; \dots; x_{k(1)} < \cdots < x_{k(n_k)}\} ,$$

$$B = \{(t_{(1)}, \dots, t_{(N)}): 0 < t_{(1)} < \cdots < t_{(N)} < 1\} .$$

Then we have

$$\begin{aligned} P\{R = r\} &= \int_{A_r} dW(H(x_{11}), \dots, H(x_{kn_k})) \\ &= \int_{B_r} \left\{\prod_{i=1}^{k} n_i!\right\} dW(H(x_{1(1)}), \dots, H(x_{k(n_k)})) \end{aligned} \tag{A.2}$$

Since $W(t_{11}, \dots, t_{kn_k})$ is invariant under all permutations within $(t_{i1}, \dots, t_{in_i})$, for each $i = 1, \dots, k$, we write (A.2) as

$$\begin{aligned} P\{R = r\} &= \int_{B} \left\{\prod_{i=1}^{k} n_i!\right\} dW(t_{(r_{11})}, \dots, t_{(r_{kn_k})}) \\ &= \int_{B} \left\{\prod_{i=1}^{k} n_i!\right\} w(t_{(r_{11})}, \dots, t_{(r_{kn_k})}) \left\{\prod_{i=1}^{k} \prod_{j=1}^{n_i} dt_{(r_{ij})}\right\} \\ &= \frac{n_1! \dots n_k!}{N!} \int_{B} N! w(t_{(r_{11})}, \dots, t_{(r_{kn_k})}) \left\{\prod_{i=1}^{N} dt_{(i)}\right\} . \end{aligned}$$

Then the theorem follows.

In Theorem A.1, the k samples, $\{X_{i1}, \ldots, X_{in_i}\}$, $i = 1, \ldots, k$ may not be independent. However, if the k samples are independent, we have the following result.

Corollary A.1: Let $X_i = (X_{i1}, \ldots, X_{in_i})$, $i = 1, \ldots, k$ be independent random vectors and for each i, X_i be distributed as $W_i(H(x_1), \ldots, H(x_{n_i}))$ where $W_i(t_1, \ldots, t_{n_i})$ is an absolutely continuous and exchangeable distribution function on $\{(t_1, \ldots, t_{n_i}): 0 < t_j < 1, j = 1, \ldots, n_i\}$ and $H(x)$ is a continuous distribution function. For each (i, j), let $R_{i(j)}$ be the rank of $X_{i(j)}$ in the pooled sample $\{X_{11}, \ldots, X_{kn_k}\}$, where $X_{i(j)}$ is the j-th order statistic (the j-th smallest value) in $\{X_{i1}, \ldots, X_{in_i}\}$. Denote $\mathbf{R} = (R_{1(1)}, \ldots, R_{k(n_k)})$, $N = \sum_{i=1}^{k} n_i$ and R is the set of

$$\mathbf{r} = (r_{11}, \ldots, r_{1n_1}, \ldots, r_{k1}, \ldots, r_{kn_k}) ,$$

a permutation of $(1, \ldots, N)$ satisfying $r_{11} < \cdots < r_{1n_1}; \ldots; r_{k1} < \cdots < r_{kn_k}$.

Then for each $r \in R$, we have

$$P\{R = r\} = \frac{n_1! \ldots n_k!}{N!} E\left\{\prod_{i=1}^{k} w_i(U_{r_{i1}:N}, \ldots, U_{r_{in_i}:N})\right\} \tag{A.3}$$

where w_i is the density function of W_i and $U_{1:N}, \ldots, U_{N:N}$ are the order statistics of a random sample of size N from $U(0, 1)$, the uniform distribution on $(0, 1)$.

Proof: The result follows from Theorem A.1 since the joint distribution of $X = (X_1, \ldots, X_k)$ is

$$W(H(x_{11}), \ldots, H(x_{kn_k})) = \prod_{i=1}^{k} W_i(H(x_1), \ldots, H(x_{n_i}))$$

We consider the k independent random samples and obtain the distribution of R under the Lehmann alternatives.

Corollary A.2: Let $\{X_{i1}, \ldots, X_{in_i}\}$, $i = 1, \ldots, k$, be independent random samples with population d.f.'s $Q_1(H(x)), \ldots, Q_k(H(x))$, respectively, where $H(x)$ is some continuous d.f. and Q_i, $i = 1, \ldots, k$, are absolutely continuous d.f.'s on $(0, 1)$. Then for each $r \in R$,

$$P\{R = r\} = \frac{n_1! \ldots n_k!}{N!} E\left\{\prod_{i=1}^{k} \prod_{j=1}^{n_i} q_i(U_{r_{ij}:N})\right\} \tag{A.4}$$

where q_i is the p.d.f. of Q_i.

Proof: Since the joint distribution of $X_i = (X_{i1}, \ldots, X_{in_i})$ for each $i = 1, \ldots, k$, is

$$W_i(H(x_{i1}), \ldots, H(x_{in_i})) = \prod_{j=1}^{n_i} Q_i(H(x_{ij}))$$

the Corollary follows from Corollary A.1.

Considering the case with $k = 2$ in Corollary A.2 and $Q_1(t) = t$ and $Q_2(t) = Q(t)$ we get the result due to Lehmann (1953, equation 4.1).

6 References

Callfas D, Mohanty SG (1986) On non-null distributions of ranks. Comm. Statist. Comp. and Simula 15:591–608

David HA (1981) Order Statistics. 2nd ed. John Wiley & Sons, New York

Deely JJ, Amos DE, Steck GP (1969) The Exceedance Test for Truncation of a Supplier's Data. Journal of the American Statistical Association 64:823–829

Haga T (1960) A Two-sample Rank Test on Location. Annals of the Institute Statistical Mathematics 11:211–219

Hajek J, Sidak Z (1967) Theory of Rank Tests. Academic Press. New York

Hoeffding W (1951) Optimum Nonparametric Tests. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability 83–92

Katzenbeisser W (1989) The exact power of two-sample location tests based on exceedance statistics against shift alternatives. Statistics 20:47–54

Klotz J (1962) Nonparametric Tests for Scale. Annals of Mathematical Statistics 33:498–512

Klotz J (1964) On the Normal Scores Two-sample Rank Test. Journal of the American Statistical Association 59:652–664

Lehmann EL (1953) The Power of Rank Tests. Annals of Mathematical Statistics 24:23–43

Lin Chiou-Hua, Sukhatme S (1990) Hoeffding type theorem and power comparison of some two-sample rank tests. Preprint No. 90-10, Statistical Laboratory, ISU, Ames, IA 50011

Lin Chiou-Hua, Sukhatme (1991) Two sample rank tests with truncated populations, Preprint No. 91-26, Statistical Laboratory, ISU, Ames, IA 50011

Mann HB, Whitney DR (1947) On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other. Annals of Mathematical Statistics 4:51–67

Mathisen HC (1943) A method of testing the hypothesis that the two samples are from the same population, Annals of Mathematical Statistics 14:188–194

Nelson Wayne (1990) Hazard Plotting of Left Truncated Life Data. Journal of Quality Technology 22:230–238

Rosenbaum S (1953) Tables for a nonparametric test of dispersion, Annals of Mathematical Statistics 24:663–668

Sidak Z, Hojek S (1979) Monte Carlo comparison of some rank tests optimal for uniform distributions. Contributions to Statistics, Jaroslav Hajek Memorial Volume, (Ed. J. Jureckova), D. Reidel, Boston

Sidak Z (1977) Tables for the Two-sample Location *E*-test Based on Exceeding Observations. Aplikace Matematiky 22:166–175

Sidak Z, Vondracek J (1957) A Simple Non-parametric Test of Difference in Location of Two Populations. (In Czech.) Aplikace Matematiky 2:215–221

Sukhatme S (1986) Power of some nonparametric tests, Proceedings of the Social Statistics Section, American Statistical Association, Washington, D.C., 272–277

Wilcoxon F (1945) Individual Comparisons by Ranking Methods. Biometrics 1:80–83

Received 17.09.91
Revised version 04.06.92

Metrika (1993) 40:173–183

Metrika

Further Developments in Estimation of the Largest Mean of K Normal Populations

N. MUKHOPADHYAY, S. CHATTOPADHYAY, S. K. SAHU

Department of Statistics, U Box 120, University of Connecticut, 196 Auditorium Road, Storrs, CT 06269 USA

Abstract: We revisit the bounded maximal risk point estimation problem as well as the fixed-width confidence interval estimation problem for the largest mean among $k(\geq 2)$ independent normal populations having unknown means and unknown but equal variance. In the point estimation setup, we devise appropriate two-stage and modified two-stage methodologies so that the associated *maximal risk* can be *bounded* from above *exactly* by a preassigned positive number. Kuo and Mukhopadhyay (1990), however, emphasized only the asymptotics in this context. We have also introduced, in both point and interval estimation problems, *accelerated sequential* methodologies thereby saving sampling operations tremendously over the purely sequential schemes considered in Kuo and Mukhopadhyay (1990), but enjoying at the same time asymptotic second-order characteristics, fairly similar to those of the purely sequential ones.

Key Words and Phrases: Two-stage procedure; modified two-stage procedure; exact bounded maximal risk; fixed-width confidence interval; accelerated sequential procedure; second-order properties.

1 Introduction

Let $\pi_1, \ldots, \pi_k$ be independent populations with the distribution of the ith population being $N(\mu_i, \sigma^2)$, $i = 1, \ldots, k$. We assume that the means $\mu_1, \ldots, \mu_k$ and the common variance σ^2 are *all unknown parameters.* We write $\mu^* = \max\{\mu_1, \ldots, \mu_k\}$ and $\underset{\sim}{\mu} = (\mu_1, \ldots, \mu_k)$. Recently, Kuo and Mukhopadhyay (1990) developed purely sequential and three-stage sampling methodologies for constructing fixed-width confidence intervals as well as bounded risk point estimators for μ^*. In this context, one should also refer to Saxena and Tong (1969) and Tong (1970).

The present note has direct bearing on the findings of Kuo and Mukhopadhyay's (1990) estimation procedures. First, we look at the point estimation problem in Section 2 and develop both two-stage and accelerated sequential

0026–1335/93/3–4/173–183 $2.50

sampling techniques. Our two-stage and modified two-stage procedures will place an *exact upper bound* on the *maximal risk* associated with the final estimator of μ^*. Kuo and Mukhopadhyay (1990) could only achieve this bound asymptotically. See, Theorem 2.3 (iv) and Theorem 2.4 (iv) in Kuo and Mukhopadhyay (1990). Next, we propose a variant of their purely sequential procedure, called *accelerated sequential procedure*, which is operationally more convenient. In other words, we are making the purely sequential procedure (2.4) of Kuo and Mukhopadhyay (1990) significantly more attractive by suitably modifying it. In Section 3, we introduce a similar accelerated sequential procedure for the corresponding fixed-width confidence interval problem and present the associated second-order properties.

2 Point Estimation

Let $X_{i1}, X_{i2}, \ldots$ be independent random variables from π_i, and having recorded $\{X_{i1}, \ldots, X_{in}\}$ from π_i, define

$$\bar{X}_{in} = n^{-1} \sum_{j=1}^{n} X_{ij} , \qquad \bar{X}_n^* = \max\{\bar{X}_{1n}, \ldots, \bar{X}_{kn}\} ,$$

and

$$S_n^2 = (kn - k)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - \bar{X}_{in})^2$$

for $n \geq 2$, $i = 1, \ldots, k$. For the point estimation problem, suppose that the loss incurred in estimating μ^* by $\bar{X}_n^*$, which is a maximum likelihood estimator, is

$$L_n = A|\bar{X}_n^* - \mu^*|^s \tag{2.1}$$

where A and s are known positive numbers. The risk of L_n depends on $\underset{\sim}{\mu}$ and σ. For fixed σ, the risk associated with (2.1) is maximized when $\mu_1 = \cdots = \mu_k$ and this *maximal risk* is given by

$$\sup_{\underset{\sim}{\mu}} E(L_n) = B\sigma^s / n^{s/2} \tag{2.2}$$

where $B = kA \int_{-\infty}^{\infty} |y|^s \Phi^{k-1}(y)\phi(y)\, dy$, $\phi(u) = (2\pi)^{-1/2} \exp(-\frac{1}{2}u^2)$, and $\Phi(u) =$

$\int_{-\infty}^{u} \phi(t)\,dt$, $-\infty < u < \infty$. Refer to Cohen and Sackrowitz (1982) and Theorem 2.1 in Kuo and Mukhopadhyay (1990).

Our *goal* is to make the maximal risk given in (2.2) not to exceed a preassigned level $W(>0)$. Thus, we first propose a two-stage as well as a modified two-stage procedure in order to attain the *goal* of *bounded maximal risk exactly*. Now, $\text{Sup}_{\mu} E(L_n) \leq W$ if n is the smallest integer $\geq (B/W)^{2/s}\sigma^2 = n^*$, say.

2.1 *A Two-Stage Procedure*

We start with $m(\geq 2$ and $> 1 + sk^{-1})$ observations from each population to begin the experiment. Let

$$h = h(m) = \left\{\frac{1}{2}k(m-1)\right\}\left\{\Gamma\left(\frac{k(m-1)-s}{2}\right)\Big/\Gamma\left(\frac{k(m-1)}{2}\right)\right\}^{2/s},$$

and define

$$N = \max\{m, [(B/W)^{2/s}hS_m^2]^* + 1\} \tag{2.3}$$

where $[x]^*$ stands for the largest integer smaller than x. If $N = m$, then we do not take any more samples from π_i's in the second stage. On the other hand, if $N > m$, then we sample the difference from each π_i. Finally, based on $\{X_{i1}, \ldots, X_{iN}\}$ from π_i, we estimate μ^* by $\bar{X}_N^*$. This Stein (1945)-type procedure has the following properties.

Theorem 1: For the two-stage procedure (2.3) *we have*:

(i) $\underset{\mu}{\text{Sup}}\, E(L_N) \leq W$ *for every fixed* $W(>0)$;
(ii) $E(N)/n^* \to h(>1)$ *as* $W \to 0$;
(iii) $W^{-1} \underset{\mu}{\text{Sup}}\, E(L_N) \to 1$ *as* $W \to 0$;

where $n^* = (B/W)^{2/s}\sigma^2$.

Proof: In view of Kuo and Mukhopadhyay's (1990) Theorem 2.2, we get

$$\underset{\mu}{\text{Sup}}\, E(L_N) = B\sigma^s E(N^{(-1/2)s}) \,. \tag{2.4}$$

From (2.3), it is immediate to note that $N \geq (B/W)^{2/s} h S_m^2$ a.s. and hence (2.4) leads to (since $m \geq 2$ and $m > 1 + sk^{-1}$)

$$\begin{aligned}\sup_{\underset{\sim}{\mu}} E(L_N) &\leq W\{h^{-1}k(m-1)\}^{(1/2)s} E[\{\chi^2_{k(m-1)}\}^{(-1/2)s}] \\ &= W\left\{\frac{1}{2}h^{-1}k(m-1)\right\}^{(1/2)s}\left\{\Gamma\left(\frac{k(m-1)-s}{2}\right)\Big/\Gamma\left(\frac{k(m-1)}{2}\right)\right\} \\ &= W\,,\end{aligned}$$

by the choice of "h". Hence, we have part (i).

For part (ii), note that

$$h(B/W)^{2/s} S_m^2 \leq N \leq h(B/W)^{2/s} S_m^2 + m \tag{2.5}$$

and hence

$$h \leq E(N)/n^* \leq h + mn^{*-1}\,,$$

which implies that $\lim_{W \to 0} E(N)/n^* = h$. Obviously, h is strictly larger than unity which follows from Lemma 2(b) of Mukhopadhyay and Hilton (1986).

In order to prove part (iii), first note from (2.4) that

$$\sup_{\underset{\sim}{\mu}} E(L_N) = WE\{(n^*/N)^{(1/2)s}\}\,. \tag{2.6}$$

From (2.5), we obtain $\lim_{W \to 0} N/n^* = (h^{-1}\sigma^2/S_m^2)^{(1/2)s}$ a.s. Hence,

$$\liminf_{W \to 0} E\{(n^*/N)^{(1/2)s}\} \geq \left\{\frac{k(m-1)}{h}\right\}^{(1/2)s} E[\{\chi^2_{k(m-1)}\}^{(-1/2)s}] = 1$$

by the obvious choice of "h" and Fatou's Lemma. On the other hand, $E\{(n^*/N)^{(1/2)s}\} \leq E\{(h^{-1}\sigma^2/S_m^2)^{(1/2)s}\} = 1$, by our choice of "$h$" as before. Thus part (iii) follows from (2.6).

Remark 1: The goal of *bounding the maximal risk* has been met *exactly* (Theorem 1(i)), however, the characteristic depicted by Theorem 1(ii) is disturbing since it implies that our two-stage procedure (2.3) *oversamples* (in comparison with n^*),

even asymptotically. In order to remedy this drawback, we now move on to implement a modified version of the two-stage procedure (2.3).

2.2 *A Modified Two-Stage Procedure*

We let m, the starting sample size, depend on W and thereby let it "grow" as W gets smaller, however at a rate slower than that of n^*. Along the lines of Mukhopadhyay (1980), we define

$$m = \max\{m_0, [1 + sk^{-1}]^* + 1, [W^{-2/(s+\gamma)}]^* + 1\} \tag{2.7}$$

for an arbitrary, but fixed and chosen, $\gamma(>0)$ and $m_0(\geq 2)$. With this $m = m(W)$, we now define the following Stein (1945)-type two-stage procedure: Let

$$N = \max\{m, [(B/W)^{2/s} h S_m^2]^* + 1\} \tag{2.8}$$

and we estimate μ^* by $\bar{X}_N^*$. The following properties hold.

Theorem 2: For the modified two-stage procedure (2.7)–(2.8) *we have*:

(i) $\sup_{\mu} E(L_N) \leq W$ *for every fixed* $W(>0)$;
(ii) $N/n^* \to 1$ *a.s. and* $E(N)/n^* \to 1$ *as* $W \to 0$;
(iii) $W^{-1} \sup_{\mu} E(L_N) \to 1$ *as* $W \to 0$;
(iv) $\liminf\{E(N) - n^*\} = \infty$ *as* $W \to 0$;

where $n^* = (B/W)^{2/s}\sigma^2$.

Proof: Part (i) follows immediately from Theorem 1(i). To prove part (ii), note that $m = m(W) \to \infty$ as $W \to 0$, but $m(W)/n^* \to 0$ as $W \to 0$. Hence, $h = h(m) = h(m(W)) \to 1$ as $W \to 0$. See Lemma 1 in Mukhopadhyay and Hilton (1986). Part (iii) follows along the lines of Theorem 1(iii) with obvious changes in a couple of places. We now turn to part (iv) and use Lemma 1 of Mukhopadhyay and Hilton (1986).

$$\begin{aligned} E(N) - n^* &\geq (B/W)^{2/s}\sigma^2\{h(W) - 1\} \\ &= O(W^{-2/s})O(m^{-1}) = O(W^{2/(s+\gamma)-2/s}) \\ &= O(W^{-2\gamma/s(s+\gamma)}) \to \infty \quad \text{as} \quad W \to 0\ , \quad \text{which is part (iv)}\ . \end{aligned}$$

The modified two-stage procedure (2.7)–(2.8) is thus *asymptotically efficient* (i.e. $E(N)/n^* \to 1$) in the Chow-Robbins (1965) sense or *asymptotically first-order efficient* in the Ghosh-Mukhopadhyay (1981) sense. However, this modified version is *asymptotically not second-order efficient* (Theorem 2(iv)) in the Ghosh-Mukhopadhyay (1981) sense. In Kuo and Mukhopadhyay (1990), a purely sequential procedure and a three-stage procedure were proposed giving rise to stopping variables, say, N_{Seq} and N_{TS} respectively. There, it was proved that

$$E(N_{Seq}) = n^* + D + o(1) \ , \tag{2.9}$$

$$E(N_{TS}) = n^* + D' + o(1) \ , \tag{2.10}$$

as $W \to 0$, with appropriate and explicit fixed quantities D and D'. In general, one expects that $|D|$ is smaller than $|D'|$, but at the same time, both the purely sequential as well as the three-stage procedures are asymptotically shown to be second-order efficient, even though the purely sequential one is slightly more so than the three-stage one. Hence, it makes sense to make the purely sequential procedure of Kuo and Mukhopadhyay (1990) operationally more convenient and at the same time, essentially maintain the second-order property given in (2.9). In a broad range of problems, it has been noticed through large scale simulations that the *accelerated sequential* procedures' overall moderate sample characteristics lie somewhere in between those of the purely sequential and three-stage procedures, but much more like those of the purely sequential ones. In this context, refer to Mukhopadhyay and Solanky (1991) [written as MS (1991) in the sequel]. See also Hall (1983).

2.3 *An Accelerated Sequential Procedure*

Let us write $S_n^{*2} = (1 - n^{-1})^{-1} S_n^2$ for $n \geq 2$. Kuo and Mukhopadhyay's (1990) purely sequential procedure essentially looks like

$$N = \inf\{n \geq m(\geq 2): n \geq (B/W)^{2/s} S_n^{*2}\} \ . \tag{2.11}$$

Instead, we chose and fix $0 < \rho < 1$ and start the experiment with $m(\geq 2)$ observations from each π. Then, we take one observation at a time from each population according to the following stopping rule: Let

$$R = \inf\{n \geq m: n \geq \rho(B/W)^{2/s} S_n^{*2}\} \ . \tag{2.12}$$

When sampling stops, we have $\{X_{i1}, \ldots, X_{iR}\}$ from π_i, $i = 1, \ldots, k$. Then, define

$$N = \max\{R, [(B/W)^{2/s} S_R^{*2}]^* + 1\} \ . \tag{2.13}$$

If $N = R$, we do not take any more samples from any population. On the other hand, if $N > R$, then we sample the difference from each population. Finally, we estimate μ^* by $\bar{X}_N^*$ obtained from $\{X_{i1}, \ldots, X_{iN}\}$, $i = 1, \ldots, k$. In view of Theorem 2.2 in Kuo and Mukhopadhyay (1990), the *maximal risk* is given by (2.4) where N comes from (2.13). Also note that the *accelerated sequential procedure* (2.12)–(2.13) cuts down the sampling operations by approximately $100(1 - \rho)\%$ when compared with the purely sequential counterpart (2.11).

Theorem 3: For the accelerated sequential procedure (2.12)–(2.13), *we have as* $W \to 0$:

(i) $-2(k\rho)^{-1} \leq \liminf E(N - n^*) \leq \limsup E(N - n^*) \leq 1 - 2(k\rho)^{-1}$ *if* $m \geq 4$;
(ii) $W + s(s + 6)W(4k\rho n^*)^{-1} + o(n^{*-1}) \geq \underset{\mu}{\text{Sup}}\ E(L_N)$
$\geq W + s(s + 6 - 2k\rho)W(4k\rho n^*)^{-1} + o(n^{*-1})$ *if*
$m > 1 + \max\{2.02, (6 + s)/k\}$;

where $n^* = (B/W)^{2/s}\sigma^2$.

Proof: The accelerated sequential procedure (2.12)–(2.13) falls under the general framework of MS (1991) with m *replaced by* $m - 1$, and their $\delta = 1$, $\theta = k$, $\tau^2 = 2k$, $a = \frac{1}{2}k$, $p = 2/k$, $m_0(\delta) = 2.02$ if $k = 2$, and 2 if $k \geq 3$, $h = k/n^*$, and $q = 0$. Hence, part (i) follows immediately from Corollary 2.1 of MS (1991).

The maximal risk is again of the form given in (2.4) where N is given by (2.13). Here, we will utilize Corollary 2.2 of MS (1991) with $l = -\frac{1}{2}s$. Note that $\eta'' = 2/k\rho$. When $s = 2$, we need $m - 1 > \max\left\{m_0(\delta), \dfrac{2}{\delta a}\right\} = \max(m_0(\delta), 4/k)$, that is $m \geq 4$. When $s \in R^+ - \{2\}$, the sufficient condition on m turns out to be $m > 1 + \max\{2.02, (6 + s)/k\}$.

3 Interval Estimation

Let us reconsider the fixed-width confidence interval estimation problems for μ^* as discussed in Kuo and Mukhopadhyay (1990). Given $d(>0)$, we take $I_n = [\bar{X}_n^* \pm d]$ as the confidence interval for μ^* when we have recorded $n(\geq 2)$ samples

from each π_i. Saxena and Tong (1969) proved that

$$\inf_{\underset{\sim}{\mu}} P(\mu^* \in I_n) = \Phi^k(dn^{1/2}/\sigma) - \Phi^k(-dn^{1/2}/\sigma) \tag{3.1}$$

for every fixed d, σ and n, and the infimum in (3.1) is attained when $\mu_1 = \cdots = \mu_k$. Therefore, to achieve the desired coverage probability, that is $P(\mu^* \in I_n) \geq 1 - \alpha$ for a preassigned $\alpha \in (0, 1)$, we first determine "a" such that $\Phi^k(a) - \Phi^k(-a) = 1 - \alpha$, and then "$n$" needs to be the smallest integer $\geq a^2\sigma^2/d^2 = C$, say. Since C is unknown, Tong (1970) proposed a purely sequential procedure for this problem, while Kuo and Mukhopadhyay (1990) studied the associated second-order properties. In view of our comments given at the end of Section 2.2, we now wish to consider an appropriate accelerated version of the sequential procedure discussed in Tong (1970) and Kuo and Mukhopadhyay (1990).

Recall that $S_n^{*2} = (1 - n^{-1})^{-1} S_n^2$ for $n \geq 2$. Kuo and Mukhopadhyay's (1990) purely sequential procedure (3.3) can essentially be written as

$$N = \inf\{n \geq m: n \geq a^2 S_n^{*2}/d^2\} , \tag{3.2}$$

where $m(\geq 2)$ is the starting sample size from each population. Instead, we choose and fix $0 < \rho < 1$ and $q \geq 0$, and start the experiment with $m(\geq 2)$ observations from each π. Then, we take one observation at a time from each population according to the following stopping rule: Let

$$R = \inf\{n \geq m: n \geq \rho a^2 S_n^{*2}/d^2\} . \tag{3.3}$$

When sampling stops, we have $\{X_{i1}, \ldots, X_{iR}\}$ from π_i, $i = 1, \ldots, k$. Then, define

$$N = \max\{R, [a^2 S_R^{*2}/d^2 + q]^* + 1\} . \tag{3.4}$$

If $N = R$, we do not take any more samples from any population. On the other hand, if $N > R$, then we sample the difference from each population. Finally, we estimate μ^* by the confidence interval I_N. Along the lines of Kuo and Mukhopadhyay (1990), we then have

$$\inf_{\underset{\sim}{\mu}} P(\mu^* \in I_N) = E\{\Phi^k(dN^{1/2}/\sigma) - \Phi^k(-dN^{1/2}/\sigma)\} , \tag{3.5}$$

where the infimum is attained when $\mu_1 = \cdots = \mu_k$.

Theorem 4: For the accelerated sequential procedure (3.3)–(3.4), *we have as* $d \to 0$:

$$\inf_{\underset{\sim}{\mu}} P(\mu^* \in I_N) \geq (1 - \alpha) + o(d^2)$$

if $q = \max\left\{0, \frac{1}{k\rho}\left[2 - \frac{1}{2}(a^2 + 1) + \frac{1}{2}(k-1)a\phi(a)\frac{\Phi^{k-2}(a) - \Phi^{k-2}(-a)}{\Phi^{k-1}(a) + \Phi^{k-1}(-a)}\right]\right\}$

when $m \geq 4$.

Proof: Let $F(x) = \Phi^k(x^{1/2}) - \Phi^k(-x^{1/2})$ for $x > 0$. Then, with $l = d^2/\sigma^2$, we have

$$\begin{aligned}\inf_{\underset{\sim}{\mu}} P(\mu^* \in I_N) &= E\{F(lN)\} \\ &= F(lE(N)) + \frac{1}{2}l^2 E\{(N - E(N))^2\}F''(lE(N)) + r_1(l) \ , \quad \text{say} \ ,\end{aligned} \tag{3.6}$$

where $|r_1(l)| = O(l^3 E(|N - E(N)|^3))$. Now, note that the accelerated sequential procedure (3.3)–(3.4) is covered by the unified theory developed in MS (1991) where m is *replaced* by $m - 1$ and their $\delta = 1$, $h = k/C$, $\theta = k$, $\tau^2 = 2k$, $a = \frac{1}{2}k$, $p = 2/k$, $n_0^* = C$, and $m_0(\delta) = 2.02$ or 2 according as $k = 2$ or $k \geq 3$ respectively. Corollary 2.1 of MS (1991) leads to

$$q - 2(k\rho)^{-1} \leq \liminf E(N - C) \leq \limsup E(N - C) \leq q + 1 - 2(k\rho)^{-1} \ , \tag{3.7}$$

if $m > 1 + m_0(\delta)$. Also, Theorem 2.2 (iii) immediately gives $|r_1(l)| = O(l^3 l^{-3/2}) = O(l^{3/2}) = o(l)$ if $m > 1 + 3k^{-1}$. Again, since $F(a^2) = 1 - \alpha$, we write for a suitable $z(>0)$,

$$F(lE(N)) = (1 - \alpha) + \{lE(N) - a^2\}F'(a^2) + r_2(l) \tag{3.8}$$

where $|r_2(l)| = \frac{1}{2}|(lE(N) - a^2)^2 F''(z)| = O(l^2) = o(l)$, if $m > 1 + \max\{m_0(\delta), 3k^{-1}\}$, in view of Theorem 2.2 (ii) and (3.6). Now, combining (3.6)–(3.8), we get

$$\begin{aligned}\inf_{\underset{\sim}{\mu}} P(\mu^* \in I_N) &\geq (1 - \alpha) + lF'(a^2)(q - 2(k\rho)^{-1} + o(1)) \\ &\quad + l\{a^2(k\rho)^{-1}F''(a^2) + o(1)\}\end{aligned} \tag{3.9}$$

if $m > 1 + \max\{m_0(\delta), 3k^{-1}\}$. Now, from (3.9), we obtain,

$$\inf_{\underset{\sim}{\mu}} P(\mu^* \in I_N) \geq (1 - \alpha) + o(d^2) \tag{3.10}$$

if $q \geq (k\rho)^{-1}\left\{2 - a^2\frac{F''(a^2)}{F'(a^2)}\right\}$. Hence, the Theorem, once one simplifies $F''(a^2)/F'(a^2)$ in terms of $\Phi(a)$ and $\phi(a)$.

Remark 2: One can certainly consider any fixed value for the fudge factor "q" utilized in the definition of N given by (3.4). Then, of course, the second-order expansion (3.9) for $\inf_{\underset{\sim}{\mu}} P(\mu^* \in I_N)$ will compare in its form quite favorably with that given in Theorem 3.1 of Kuo and Mukhopadhyay (1990) for the corresponding purely sequential methodology.

Remark 3: The accelerated sequential procedure (3.3)–(3.4) cuts down the sampling operations by approximately $100(1 - \rho)\%$ when compared with the fully sequential methodology discussed in Kuo and Mukhopadhyay (1990).

Remark 4: The accelerated sequential procedures such as (2.12)–(2.13) and (3.3)–(3.4) seem to perform most favorably for moderate values of n^* and C respectively, when ρ is chosen between 0.4 and 0.7. This feature has been noted via large scale simulation exercises for various related problems across the board.

Acknowledgement: This project has been partially supported by the Grant Number 89-0225 from the Air Force Office of Scientific Research.

References

Chow YS, Robbins H (1965) On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* 36:457–462

Cohen A, Sackrowitz H (1982) Estimating the mean of the selected population. *Statistical Decision Theory and Related Topics III* (S.S. Gupta and J.O. Berger, eds.), Academic Press, New York, 243–270

Ghosh M, Mukhopadhyay N (1981) Consistency and asymptotic efficiency of two-stage and sequential estimation procedures. *Sankhya A*, 43:220–227

Hall P (1983) Sequential estimation saving sampling operations. *J. Roy. Statist. Soc. B* 45:219–223

Kuo L, Mukhopadhyay N (1990) Multi-stage point and interval estimation of the largest mean of k normal populations and the associated second-order properties. *Metrika* 37:291–300

Mukhopadhyay N (1980) A consistent and asymptotically efficient two-stage procedure to construct fixed-width confidence intervals for the mean. *Metrika* 27:281–284

Mukhopadhyay N, Hilton GF (1986) Two-stage and sequential procedures for estimating the location parameter of a negative exponential distribution. *South Afr. Statist. J.* 20:117–136

Mukhopadhyay N, Solanky TKS (1991) Second order properties of accelerated stopping times with applications in sequential estimation. *Sequential Analysis* 10:99–123

Stein C (1945) A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* 16:243–258

Saxena KML, Tong YL (1969) Interval estimation of the largest mean of k normal populations with known variances. *J. Amer. Statist. Assoc.* 64:296–299

Tong YL (1970) Multi-stage interval estimation of the largest of k normal means. *J. Roy. Statist. Soc. B* 32:272–277

Received 15.06.1992

Book Review

C. Constantinescu: Maß und Integral, VDF Zürich, 1990, 324 pp., Fr. 58,00

This second, revised and augmented edition (1990) of C. Constantinescu's lecture notes on measure and integral consists of four chapters and an appendix. The approach is set-theoretical with some new and interesting modifications.

Chapter I starts with the treatment of (finite positive) measures defined on semirings. Null sets are always of a local nature and hereditary w.r.t. subsets. The definition of outer measure uses covers up to null sets. The most significant modification concerns the completion of measures. Here, in addition a Beppo Levi property is required. The completion is therefore always defined on a δ-ring.

Chapter II starts with the concept of measurability w.r.t. δ-rings. The integral is then defined w.r.t. the completion and is hence an abstract essential integral. This requires some appropriate modifications of classical proofs, e.g. of the convergence theorems.

Chapter III is devoted to the study of L^p spaces, of real measures and their Hahn-Jordan decomposition, of measures with densities and product measures on δ-rings.

In the comprehensive Chap. IV, Radon measures on arbitrary Hausdorff spaces are treated The approach, meanwhile accepted in topological measure theory, follows the one of Kisynski and Topsoe using inner regularity w.r.t. compact sets. Local boundedness is not required. Finally, representation of set functions and linear functionals by Radon measures is discussed, not only in the locally compact case.

In the appendix, many valuable remarks concerning the history of measure and integral can be found.

Though there are no excercises, these lecture notes can be recommended to students. But even the expert will find an interesting approach to the classical theory.

Erlangen B. Anger

Metrika (1993) 40:185–189

Metrika

Bounds for the Median of the Negative Binomial Distribution

R. VAN DE VEN, N. C. WEBER

School of Mathematics and Statistics (F07), University of Sydney, NSW 2006, Australia

Summary: Bounds are obtained for the median of the negative binomial distribution which are valid for all possible parameter values of the distribution when the median is defined as $\inf\{x: P(X \leq x) \geq \frac{1}{2}\}$.

Key Words: beta distribution, negative binomial distribution, mean, median, mode.

1 Introduction

For the beta distribution Groeneveld and Meeden (1977) outlined a procedure by which it can be shown simply that the median lies between the mean and mode, provided that the two parameters of the distribution are strictly greater than 1. Using this result Payton, Young and Young (1989) obtained bounds for the "median" of the negative binomial distribution (NBD) when the shape parameter is strictly greater than one.

Unfortunately, the value(s) treated by Payton et al. (1989) as median values for the NBD do not, except in the case where there exists a value of x such $P(X \leq x) = \frac{1}{2}$, satisfy the generally accepted definition of the median of a random variable X, that being any value (m say) satisfying $P(X < m) \leq \frac{1}{2} \leq P(X \leq m)$. The bounds thus obtained are not necessarily correct for this definition of the median (e.g. see Payton et al. (1989), *Remark* 2).

This paper hence has two objectives. Firstly, we modify the results of Payton et al. (1989) so that they hold for a rigorously defined median. The definition we employ is

$$\text{Med}(X) = \inf\{x: P(X \leq x) \geq \tfrac{1}{2}\} \ . \tag{1}$$

All reference hereafter to the median will refer to this value. Secondly, we

0026–1335/93/3–4/185–189 $2.50

extend the results obtained by Payton et al. (1989) to cover the complete family of NBDs. In the process we extend the results available for the median of the beta distribution. To complete the paper some remarks are offered on the results obtained.

2 Main Results

Let $X = G_{p,q}$ be a beta(p, q) random variable with distribution function $G_{p,q}(x)$ and density function $g_{p,q}(x)$ given by

$$g_{p,q}(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1} \; ; \qquad 0 \le x \le 1; p > 0, q > 0 \; .$$

Theorem 1: If X has a beta(p, q) distribution then

$$0 < \text{Med}(X) < E(X) < \tfrac{1}{2} \Leftrightarrow p < q \; ,$$

$$\text{Med}(X) = E(X) = \tfrac{1}{2} \Leftrightarrow p = q \; ,$$

$$\tfrac{1}{2} < E(X) < \text{Med}(X) < 1 \Leftrightarrow p > q \; .$$

Proof: To obtain this result separate the parameter space into the disjoint regions $p = q$, $p < q$ and $p > q$. Now by symmetry, if $p = q$ then $\text{Med}(X) = E(X) = \frac{1}{2}$. For the region $p < q$ it can be shown, by considering each of the following sub-regions $1 < p < q$, $1 = p < q$, $p < 1 \le q$ and $p < q < 1$ separately, that $0 < \text{Med}(X) < E(X) < \frac{1}{2}$. Here the result of Groeneveld and Meeden (1977) that

$$0 < \frac{p-1}{p+q-2} = \text{Mode}(X) < \text{Med}(X) < E(X) < \tfrac{1}{2} \; ; \qquad 1 < p < q \qquad (2)$$

gives the result for the first sub-region. For $1 = p < q$ the result can be shown analytically whilst for the remaining two sub-regions the result follows using the approach outlined in Groeneveld and Meeden (1977) and the fact that if $p < 1$, $g_{p,q}(x)$ has a minimum on $[0, 1]$ in the interval $(\frac{1}{2}, 1]$. Similarly, we can obtain the results for the case $p > q$ and then the theorem follows.

Now let Y have a negative binomial distribution with mean $\mu\,(> 0)$ and shape parameter $\alpha\,(> 0)$, written NBD(μ, α). Then we have

$$P(Y = y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha)y!}\left(\frac{\alpha}{\alpha + \mu}\right)^{\alpha}\left(\frac{\mu}{\alpha + \mu}\right)^{y} ; \qquad y = 0, 1, 2, \ldots .$$

Using the result of Patil (1960) that

$$F(y) = P(Y \leq y) = G_{\alpha,[y]+1}\left(\frac{\alpha}{\alpha + \mu}\right) ; \qquad y \geq 0 , \tag{3}$$

where $[x]$ denotes the largest integer not greater than x, together with a tightening of the procedure employed by Payton et al. (1989) we obtain the following result for the median of a NBD(μ, α).

Theorem 2: If $\eta = \text{Med}(Y)$ then for all $\alpha > 0$ and $\mu > 0$

$$\left[\left(\frac{\alpha - 1}{\alpha}\right)\mu\right] + 1 \leq \eta < \mu \qquad \text{if } \mu > \alpha$$

$$]\mu[- 1 \leq \eta < \left(\frac{\alpha - 1}{\alpha}\right)\mu + 1 \qquad \text{if } \mu \leq \alpha$$

where $]x[\ (= -[-x])$ denotes the smallest integer not less than x.

Proof: In the proof below use is made of the fact that, by definition, η is an integer. Consider the case $\mu > \alpha$.

From (1) and (3) we have $\frac{1}{2} \leq F(\eta) = G_{\alpha,\eta+1}\left(\frac{\alpha}{\alpha + \mu}\right)$ giving

$$\text{Med}(G_{\alpha,\eta+1}) \leq \left(\frac{\alpha}{\alpha + \mu}\right) < \frac{1}{2} . \tag{4}$$

So by Theorem 1 we obtain $E(G_{\alpha,\eta+1}) < \frac{1}{2}$ and hence

$$\eta > \alpha - 1 \qquad \text{for all } \mu > \alpha . \tag{5}$$

For $\eta \geq \alpha$ we have, again by Theorem 1,

$$\text{Med}(G_{\alpha,\eta}) \leq E(G_{\alpha,\eta}) = \left(\frac{\alpha}{\alpha + \eta}\right) ,$$

where the last term is strictly greater than $\left(\frac{\alpha}{\alpha+\mu}\right)$ because

$$\frac{1}{2} > F(\eta - 1) = G_{\alpha,\eta}\left(\frac{\alpha}{\alpha+\mu}\right) \Rightarrow \text{Med}(G_{\alpha,\eta}) > \left(\frac{\alpha}{\alpha+\mu}\right) .$$

From here we obtain the result $\eta < \mu$ whenever $\eta \geq \alpha$ and hence

$$\eta < \mu \qquad \text{for all } \mu > \alpha . \tag{6}$$

For the case $\alpha > 1$, we have, from (5), (2) and (4)

$$\left(\frac{\alpha-1}{\alpha+\eta-1}\right) = \text{Mode}(G_{\alpha,\eta+1}) < \text{Med}(G_{\alpha,\eta+1}) \leq \left(\frac{\alpha}{\alpha+\mu}\right)$$

from which it follows that

$$\eta > \left(\frac{\alpha-1}{\alpha}\right)\mu \qquad \text{for all } 1 < \alpha < \mu . \tag{7}$$

Equations (5), (6) and (7) can be combined to give

$$\left[\left(\frac{\alpha-1}{\alpha}\right)\mu\right] + 1 \leq \eta < \mu \qquad \text{if } \mu > \alpha .$$

The second part of Theorem 2 (corresponding to $\mu \leq \alpha$) can be obtained using a similar argument to the above and the complementary result to (2) given by Groeneveld and Meeden (1977), that is,

$$\frac{1}{2} < E(X) < \text{Med}(X) < \text{Mode}(X) = \frac{p-1}{p+q-2} ; \qquad 1 < q < p .$$

3 Remarks

Let $Y \sim NBD(\mu, \alpha)$ and $\eta = \text{Med}(Y)$.

1. If $\mu = \alpha$ then $\eta =]\mu[- 1$.
2. Let $M = \left(\frac{\alpha - 1}{\alpha}\right)\mu$. The mode of Y is $([M] \vee 0)$ if M is not a positive integer whilst the modal values are M and $M - 1$ if M is a positive integer. From the lower bounds in Theorem 2 we have η always bounded below by the mode. Moreover, if M is a positive integer and $\alpha \geq \mu$ then $\eta = M$.
3. It is not the case that the median of Y is always bounded above by the mean of Y. For example, if $Y \sim NBD(1.9, 3)$ then $\eta = 2$.
4. If α and η both take values in $(0, 1]$ then $\eta = 0$.
5. For the case $\mu > \alpha$ it can be seen from Theorem 2 that the range of possible values for the median of $Y \sim NBD(\mu, \alpha)$ can be quite large. However, as

$$\frac{Y}{\mu} \xrightarrow{d} \Gamma\left(\alpha, \frac{1}{\alpha}\right) \qquad \text{as } \mu \to \infty \ (\alpha \text{ fixed}) \ ,$$

there would appear little scope for relative tightening if the bounds obtained by Chen and Rubin (1986) for a gamma random variable

$$(\alpha - 1/3) < \text{Med}(\Gamma(\alpha, 1)) < \alpha$$

are tight as these give the following bounds

$$\left(\frac{\alpha - \frac{1}{3}}{\alpha}\right) < \lim_{\mu \to \infty} \text{median}\left(\frac{Y}{\mu}\right) < 1 \ .$$

References

Chen J, Rubin H (1986) Bounds for the difference between mean and median of gamma and Poisson distributions. Statistics and Probability Letters 4:281–283

Groeneveld RA, Meeden G (1977) The mode, median, and mean inequality. The American Statistician 31:120–121

Patil GP (1960) On the evaluation of the negative binomial distribution with examples. Technometrics 2:501–505

Payton ME, Young LJ, Young JH (1989) Bounds for the difference between median and mean of beta and negative binomial distributions. Metrika 36:347–354

Received 23.06.1992

Book Review

A. D. Barbour, L. Holst, and S. Janson: Poisson Approximation. Oxford Studies in Probability 2, Clarendon Press, Oxford, 1992, X/277 pp., £30.00

This book is dedicated to a thorough study of various aspects concerning Poisson approximation with rates and its applications in one and several dimensions (including stochastic processes), with a particular emphasis on couplings and the so-called Stein-Chen method. Though originally developed in the context of the central limit theorem, this method together with some quite sophisticated refinements by the authors has turned out to be one of the most powerful tools to tackle a large number of diverse approximation problems, not only for Poisson limits. In particular, it allows for asymptotic expansions and very sharp (two-sided) estimations for the approximation error of the distribution of e.g. sums of dependent and independent random variables by suitable Poisson distributions, in various metrics such as total variation, Wasserstein type metrics or the Hellinger distance. A separate section (Chap. 10) is devoted to a generalization of the Stein-Chen method to Poisson process approximation, a field in which research is still going to continue. A novelty might also be the discussion of the Poisson approximation of mixed Poisson distributions which is scattered over several sections (e.g. Chap. 3.3).

Besides the more technical background, the book covers quite a large number of examples in which Poisson approximation is useful and, sometimes, the only possible way to obtain numerical results. This unique collection ranges from classical problems such as matchings, "ménages" and urn problems to random graph theory, order statistics and extreme value theory; some of the results are presented in book form for the first time here.

Although the authors point out in the preface that "the bulk of the material can be understood with no more than a knowledge of elementary discrete probability theory", the book as a whole is by no means elementary. The style of presentation is excellent; each chapter is accompanied by extensive introductory remarks which frequently sketch the basic ideas of the theory in advance and give references to related work. Where possible, direct comparison is made with competing methods (such as characteristic functions, or operator theory). Some basic facts on probability metrics are given in the appendix.

Summarizing, this book is, in my opinion, not only an outstanding up-to-date research monograph (almost 20% of the references date from 1990 and later!), but also a rich source of examples and recent problems for those readers who are more interested in applications. It will surely be one of my favorites for a long time.

Oldenburg D. Pfeifer

Metrika (1993) 40:191–201

Metrika

Bounds on the Efficiency of the Residual Design of Extended BIB Designs

SANPEI KAGEYAMA AND HIROSHI SETOYA

Department of Mathematics, Hiroshima University, Hiroshima 734, Japan

Abstract: This gives a complete proof for bounds of the efficiency, conjectured by Das and Kageyama (1992), on robustness of extended balanced incomplete block designs against the unavailability of any number of observations in a block.

1 Introduction

The robustness problem of block designs against the unavailability of data has been considered in various ways. Das and Kageyama (1992) showed the robustness of extended balanced incomplete block (BIB) designs when all the observations or any one observation in a block are lost. Furthermore, by a *computer search* they also observed the robustness of extended BIB designs against the unavailability of any number of observations in a block, *within some range of design parameters.*

In this note, we shall prove bounds of the efficiency to show generally the robustness of extended BIB designs against the unavailability of any number of observations in a block. This gives a complete answer to the problem presented in Section 3.3 of Das and Kageyama (1992).

2 The Bounds

An extended BIB design d with v treatments and $b + 1$ blocks of size k each is a design obtained by juxtaposing one binary block of size k to a BIB design with parameters v, b, r, k and λ denoted by a BIBD(v, b, r, k, λ).

In the extended BIB design d, suppose that s ($1 \leq s \leq k$) observations in any one block of d are lost. Let p be the number of treatments common to the missing treatments and k treatments in the added block, and let q be the number of treatments common to the remaining treatments (in a block containing the

0026–1335/93/3–4/191–201 $2.50

missing treatments) and k treatments in the added block. The parameters should satisfy

$$0 \leq p \leq s \ , \quad \max\{0, 2k - v - p\} \leq q \leq k - s \ . \tag{2.1}$$

Let d^* be the design obtained by deleting s observations for $1 \leq s \leq k$ in any one block in d. Assume d^* to be connected. As shown in Das and Kageyama (1992), the efficiency of residual design d^* is given by

$$e_s(p, q) = \frac{\phi_2}{\phi_1(s)} \quad (= B, \text{ say})$$

where

$$\phi_2 = k\{\lambda v(v-1) + k(v-k)\}/\{\lambda v(\lambda v + k)\} \ ,$$

$$\phi_1(s) = k\{v - k - s + 2(p-1)\}/(\lambda v) + k(k - p - 2)/(\lambda v + k)$$
$$+ k(s - p - 1)/(\lambda v - k) + 4\lambda v k(k-s)(1/D + 1/E)$$
$$\text{for } 1 \leq s \leq k - 1 \ ,$$

$$D = 2\lambda^2 v^2 (k-s) - \alpha + \sqrt{\alpha^2 - \beta} \ ,$$

$$E = 2\lambda^2 v^2 (k-s) - \alpha - \sqrt{\alpha^2 - \beta} \ ,$$

$$\alpha = (k-s)\{k(k+2p) - p(p+2q)\} - sq(k-q) \ ,$$

$$\beta = 4pk^2(k-s)(k-s-q)(2k-p-q) \ ; \quad \text{or}$$

$$\phi_1(s) = 2\lambda v k(k-p-1)/(\lambda^2 v^2 - k^2) + k\{v - 2(k-p) - 1\}/(\lambda v)$$
$$+ 2\lambda v k/(\lambda^2 v^2 - 2kp + p^2) \qquad \text{for } s = k \ .$$

Remark 2.1: The above expressions of $\phi_1(s)$ for $1 \leq s \leq k-1$ and $s = k$ are derived through Lemmas 3.1 and 3.3 of Das and Kageyama (1992), respectively. In fact, Lemma 3.1 is given under $p \geq 1$, $k - p - q \geq 1$, $q \geq 1$, $k - s - q \geq 1$, $s - p \geq 1$ and $v - 2k + p + q \geq 1$, while Lemma 3.3 is given under $p \geq 1$, $k - p \geq 1$ and $v - 2k + p \geq 1$. These restrictions occur from some patterns of the C-matrix of the residual design d^*. These parametric restrictions form a subset of ranges (2.1). To cover all the ranges as in (2.1), we have to consider

special cases of such restrictions. Simplified expressions of $\phi_1(s)$ for these special cases can be given by Lemmas 3.2.1, 3.2.2, 3.2.3 and 3.2.4 of Das and Kageyama (1992). {Incidentally, expressions derived in such a way may coincide with those given by formally omitting eigenvalues with negative values of multiplicity, as in the results of Lemma 3.1 of Das and Kageyama (1992), used to derive $\phi_1(s)$.} For such special cases Das and Kageyama (1992) implicitly have shown that our theorem described later holds. Thus our theorem is expressed under (2.1).

The following bound is now conjectured by Das and Kageyama (1992; Section 3.3).

Theorem: In an extended BIB design obtained by adding one binary block of size k to a BIB design with parameters v, b, r, k, λ, when s ($1 \leq s \leq k$) observations in any one block are lost, the efficiency B of residual designs satisfies

$$A \geq B \geq C$$

where

$$A = \frac{\lambda v(v-1) + k(v-k)}{\lambda v(v-1) + k(v-k+1)},$$

$$C = \frac{(\lambda v - k)\{\lambda v(v-1) + k(v-k)\}}{\lambda^2 v^2(v-1) - k^2(v-2k+1)} \quad (\text{when } v \geq 2k)$$

or

$$(\lambda v - k)\{\lambda v(v-1) + k(v-k)\}(\lambda^2 v^2 - 2vk + v^2)/[2\lambda^2 v^2\{(v-k-1)$$
$$\times (\lambda^2 v^2 - 2vk + v^2) + \lambda^2 v^2 - k^2\}$$
$$+ (2k - v - 1)(\lambda^2 v^2 - k^2)(\lambda^2 v^2 - 2vk + v^2)]$$
$$(\text{when } k+1 \leq v \leq 2k-1)$$

with

$$0 \leq p \leq s\,, \quad \max\{0, 2k - v - p\} \leq q \leq k - s\,.$$

In fact, $A = e_1(1, k-1)$ and $C = e_k(p^*, 0)$ with $p^* = \max\{0, 2k - v\}$.

The theorem shall be proved by separating the range of s into two cases as $1 \leq s \leq k-2$ and $s = k, k-1$.

3 Proof of the Theorem for $s = k, k-1$

Proposition 3.1: In a BIBD(v, b, r, k, λ), an inequality $A \geq B \geq C$ holds for $s = k$ and $k-1$.

Case: $s = k$. Note that $q = 0$ in this case.

(I) *A proof of* $B \geq C$. Since $\max\{0, 2k-v\} \leq p$, $0 \leq p$ and $2k - v \leq p$. Hence Theorem 3.4 of Das and Kageyama (1992) shows $B \geq C$.

(II) *A proof of* $A \geq B$.

(i) *Case* $p = k$: Here we get $\phi_1(k) = k(v-1)/(\lambda v)$. Hence $B = \{\lambda v(v-1) + k(v-k)\}/\{(\lambda v + k)(v-1)\}$. Thus, $A - B = k(k-2)\{\lambda v(v-1) + k(v-k)\}/[(\lambda v + k)(v-1)\{\lambda v(v-1) + k(v-k+1)\}] \geq 0$.

(ii) *Case* $\max\{0,2k-v\} \leq p \leq k-1$: Now

$$A - B = [\{\lambda v(v-1) + k(v-k)\}(2(\lambda^2 v^2 - k^2)p(2k-p) + (\lambda^2 v^2 - 2kp + p^2)\{\lambda vk(k-2) - k^2(2p-k)\}\}]/[\{\lambda v(v-1) + k(v-k+1)\}N]$$

where

$$N = (\lambda^2 v^2 - 2kp + p^2)(\lambda^2 v^3 - 3\lambda^2 v^2 + 2k^3 - vk^2 - 2k^2 p + k^2) + 2\lambda^2 v^2(\lambda^2 v^2 - k^2).$$

It follows that the denominator of $A - B$ is positive. Because (1) $\lambda v(v-1) + k(v-k+1) > 0$, (2) $\lambda^2 v^2 - 2kp + p^2 > 0$, and (3) $\lambda^2 v^3 - 3\lambda^2 v^2 + 2k^3 - vk^2 - 2k^2 p + k^2 > 0$. Furthermore, the fact that the numerator of $A - B$ is positive can be shown by the following relations:

(4) $2(\lambda^2 v^2 - k^2)p(2k-p) \geq 0$, $\lambda v(v-1) + k(v-k) > 0$;

(5) $\lambda vk(k-2) - k^2(2p-k) > k^2(k-2) - k^2(2p-k) = 2k^2(k-p-1) \geq 0$.

Thus, we get $A \geq B$ when $\max\{0, 2k-v\} \leq p \leq k-1$. Therefore, by cases (i) and (ii), $A \geq B$ holds for $\max\{0, 2k-v\} \leq p \leq k$.

Case: $s = k-1$. As noted in Remark 3.2 of Das and Kageyama (1992), the case of $s = k-1$ is actually equivalent to the case of $s = k$. In fact, we only have $q = 0$ (for $s = k$) and $q = 0$ or 1 (for $s = k-1$). Then, in B, $e_{k-1}(p, 0) = e_k(p, 0)$ and

$e_{k-1}(p, 1) = e_k(p + 1, 0)$ for $p = 0, 1, \ldots, k - 1$. This implies that $A \geq B \geq C$ holds for $s = k - 1$. The proof is thus completed. □

4 Proof of the Theorem for $1 \leq s \leq k - 2$

Lemma 4.1: In the following BIB designs, $A \geq B \geq C$ holds for all s such that $1 \leq s \leq k - 2$: BIBD$(v, b, r, k, \lambda) = (4, 4, 3, 3, 2)$, $(5, 5, 4, 4, 3)$, $(6, 6, 5, 5, 4)$, $(7, 7, 3, 3, 1)$, $(7, 7, 4, 4, 2)$, $(11, 11, 6, 6, 3)$.

This can be checked by calculation of factors A, maximum and minimum of B, and C. Note that the existence of these BIB designs can be seen in Raghavarao (1971).

Lemma 4.2: In a BIBD(v, b, r, k, λ), $F > 0$ and $G > 0$ hold for all s such that $1 \leq s \leq k - 2$, where

$$F = \lambda^4 v^4 (k - s) - \lambda^2 v^2 k(k - s)(k + 2p) + \lambda^2 v^2 p(k - s)(p + 2q)$$
$$+ \lambda^2 v^2 sq(k - q) + pk^2(k - s - q)(2k - p - q) ,$$
$$G = 2(\lambda^2 v^2 - k^2)\{\lambda^4 v^4 (k - s) - pk^2(k - s - q)(2k - p - q)\}$$

with $0 \leq p \leq s$ and $\max\{0, 2k - v - p\} \leq q \leq k - s$.

Proof: It is clear that

$$\lambda^2 v^2 p(k - s)(p + 2q) + \lambda^2 v^2 sq(k - q) + pk^2(k - s - q)(2k - p - q) \geq 0 . \quad (4.1)$$

Some calculation shows that $\lambda^4 v^4 (k - s) - \lambda^2 v^2 k(k - s)(k + 2p) > \lambda^2 v^2 (k - s)[k^2\{k(k - 2) - 2\} + 2k] > 0$, since $k \geq 3$. This with (4.1) implies $F > 0$. Next,

$$\lambda^4 v^4 (k - s) - pk^2(k - s - q)(2k - p - q)$$
$$\geq (k - s)\left\{\frac{v^4 r^4 (k - 1)^4}{(v - 1)^4} - 2pk^3\right\} + pk^2(p + q) + pqk^2(2k - p - q)$$
$$> (k - s)k^3(k - 1)\{(k - 1)^3 - 2\} > 0 ,$$

since $r \geq k$, $p \leq k-1$, $2k-p-q \geq 0$ and $v \geq 3$. This with $\lambda^2 v^2 - k^2 > 0$ implies $G > 0$. □

Lemma 4.3: In a BIBD(v, b, r, k, λ) with $v \leq 2k-1$, an inequality $k \leq r \leq 2\lambda$ holds.

Proof: By $\lambda(v-1) = r(k-1)$, $\lambda v = rk + \lambda - r \leq \lambda(2k-1)$, i.e. $(r-2\lambda)(k-1) \leq 0$. Hence $r \leq 2\lambda$. By Fisher's inequality $r \geq k$. □

Lemma 4.4: In a BIBD(v, b, r, k, λ), $H > 0$ holds for all s such that $1 \leq s \leq k-2$, where

$$H = 4(k-s)\{(\lambda^2 v^3 - 3\lambda^2 v^2 - vk^2 + k^3 + k^2 s - 2k^2 p - \lambda vk^2 + \lambda vks + \lambda vk)F + G\} .$$

Proof: This is given by separating into two cases.

Case (I): $v \geq 2k$. Note that in this case $v \geq 6$. At first, by $k - s > 0$ and Lemma 4.2, $F > 0$ and $G > 0$. Next, it holds that

$$\begin{aligned}
&\lambda^2 v^3 - 3\lambda^2 v^2 - vk^2 + k^3 + k^2 s - 2k^2 p - \lambda vk^2 + \lambda vks + \lambda vk \\
&\quad \geq k^2(k-p) + k^2(s-p) + v[(\lambda v - k)\{\lambda(v-3) + k\} \\
&\qquad - \lambda k(k+1)] \quad (\text{by } s \geq 1) \\
&\quad \geq k^2(k-p) + k^2(s-p) + (\lambda v - k)v\{\lambda(v-3) - 1\} \quad (\text{by } v \geq 2k \text{ and } \lambda \geq 1)
\end{aligned}$$

which is positive. These relations show that $H > 0$.

Case (II): $k+1 \leq v \leq 2k-1$. Note that $v \geq 4$. Since $k \geq 3$, at first we have $k - s > 0$ and, by Lemma 4.2, $F > 0$ and $G > 0$. Next,

$$\begin{aligned}
&\lambda^2 v^3 - 3\lambda^2 v^2 - vk^2 + k^3 + k^2 s - 2k^2 p - \lambda vk^2 + \lambda vks + \lambda vk \\
&\quad = k^2(k-p) + k^2(s-p) + v\{\lambda^2 v(v-3) - \lambda k^2 + (\lambda s + \lambda - k)k\} \\
&\quad > \lambda^2 v(v-3) - \lambda k^2 + (2\lambda - k)k \quad (\text{by } k > p, s \geq p, s \geq 1) \qquad (4.2)
\end{aligned}$$

$$\geq 2\lambda^2[\{k(k-3)-2\}/2] \ . \quad \text{(by Lemma 4.3 with } k \leq 2\lambda \text{ and } k+1 \leq v) \tag{4.3}$$

When $k \geq 4$, (4.3) is positive. When $k = 3$ and $v = 4$, (4.2) $= \lambda(4\lambda - 3) - 9 > 0$. When $k = 3$ and $v = 5$, (4.2) $= \lambda(10\lambda - 3) - 9 > 0$. Thus, $H > 0$ when $k + 1 \leq v \leq 2k - 1$. Hence the proof is completed. □

Lemma 4.5: In a BIBD(v, b, r, k, λ) with $k + 1 \leq v \leq 2k - 1$, $K > 0$ holds, where $K = 2\lambda^2v^2\{(v-k-1)(\lambda^2v^2 - 2vk + v^2) + \lambda^2v^2 - k^2\} + (2k - v - 1) \times (\lambda^2v^2 - k^2)(\lambda^2v^2 - 2vk + v^2)$.

Proof: Since $\lambda \geq 1$ and $v > k$, $\lambda^2v^2 - 2vk + v^2 = (\lambda^2 + 1)v^2 - 2vk > 0$ and $\lambda^2v^2 - k^2 > 0$. Hence $K > 0$. □

Thus, since in the six BIB designs as in Lemma 4.1 our bounds hold, such six BIB designs are, in particular, excluded to prove the following Propositions 4.1, 4.2 and 4.3.

Proposition 4.1: In a BIBD(v, b, r, k, λ), an inequality $A \geq B$ holds for all s such that $1 \leq s \leq k - 2$.

Proof: Since with H as in Lemma 4.4

$$(A - B)/\{\lambda v(v-1) + k(v-k)\}$$

$$= \frac{H - (\lambda v - k)\{\lambda v(v-1) + k(v-k+1)\}DE}{\{\lambda v(v-1) + k(v-k+1)\}H} \quad (=\alpha, \text{ say}) \ ,$$

if $\alpha \geq 0$, then $A \geq B$. It is clear from Lemma 4.4 that the denominator of α is positive. Now denote the numerator of α by α'. So

$$\alpha'/\{4(k-s)\} = (-2\lambda^2v^2 + k^2s - 2k^2p + k^2 + \lambda vks - \lambda vk)F + G \ .$$

(a) *Case* $s = 1$ (and hence $p = 0$ or 1 and $q \leq k - 1$):

(1) When $p = 0$,

$$\alpha'/\{4(k-s)\} = 2\lambda^2v^2(\lambda^2v^2 - k^2)(k^3 - k^2 - kq + q^2)$$

$$\geq 2\lambda^2v^2(\lambda^2v^2 - k^2)\{k^2(k-2) + k + q^2\} > 0.$$

(2) When $p = 1$,

$$\alpha'/\{4(k-s)\}$$

$$= (2k-q-1)(k-q-1)[2\lambda^2 v^2\{v^2 r^2(k-1)^2/(v-1)^2 - 2k^2\} + 2k^4]$$

$$\geq (2k-q-1)(k-q-1)[2\lambda^2 v^2 k^2\{k(k-2)-1\} + 2k^4]$$

which is non-negative, since $k \geq 3$ and $v \geq 4$. Thus $\alpha' \geq 0$.

(b) *Case* $2 \leq s \leq k-2$ (and hence $k \geq 4$): Let

$$\alpha'/\{4(k-s)\} = 2(\lambda^2 v^2 - k^2)L + MF$$

where

$$L = \lambda^2 v^2[\{(k-s)(k+p) - sq\}(k-q) + (k-s)(k-p)(p+q)]$$

$$- 2pk^2(k-s-q)(2k-p-q) ,$$

$$M = k^2 s - 2k^2 p - k^2 + \lambda vks - \lambda vk .$$

(3) Since $\lambda^2 v^2 - k^2 > 0$, it follows from Lemma 4.2 that $F > 0$.

(4) $L \geq \lambda^2 v^2[\{(k-s)(k+p) - sq\}(k-q) + (k-s)(k-p)(p+q)$

$$- (k-s-q)(2k-p-q)]$$

$$\geq \lambda^2 v^2\{(k-s)(k+p) - sq + (k-p)(p+q) - (2k-p-q)\}(k-s-q)$$

$$= \lambda^2 v^2\{(k+p)(k-s-1) - k + (k+1)(p+q) - sq$$

$$+ p(1-p-q)\}(k-s-q)$$

$$\geq 0 .$$

(5) $M \geq 0$ can be shown by considering four cases, $s \geq p+1$, $s = p$, $s \geq 3$ and $s = 2$, separately, after some algebra.

Thus, $\alpha' \geq 0$. Hence the proof is completed. □

Proposition 4.2: In a BIBD(v, b, r, k, λ) with $v \geq 2k$, an inequality $B \geq C$ holds for all s such that $1 \leq s \leq k-2$.

Proof: Since

$$(B - C)/[(\lambda v - k)\{\lambda v(v-1) + k(v-k)\}]$$

$$= \frac{\{\lambda^2 v^2(v-1) - k^2(v-2k+1)\}DE - H}{\{\lambda^2 v^2(v-1) - k^2(v-2k+1)\}H} \quad (=\beta, \text{ say})$$

if $\beta \geq 0$, then $B \geq C$. It follows from Lemma 4.4 that the denominator of β is positive. For, $\lambda^2 v^2(v-1) - k^2(v-2k+1) > 2\lambda^2 v^2(k-1) > 0$. Now denote the numerator of β by β'. Then, since $k \geq 3$ and hence $v \geq 6$,

$$\begin{aligned}
\beta'/\{4(k-s)\} &= \lambda^4 v^4(k^3 + k^2 - k^2 s + 2k^2 p + \lambda v k^2 - \lambda v k s - \lambda v k)(k-s) \\
&\quad + \{(\lambda^2 v^2 - k^2) + k^2(k-s) + 2k^2 p + \lambda v k(k-s) \\
&\quad + \lambda v(\lambda v - k)\}\{\lambda^2 v^2 p(k-s)(p+2q) - \lambda^2 v^2 k(k-s)(k+2p) \\
&\quad + \lambda^2 v^2 s q(k-q)\} + pk^2\{3(\lambda^2 v^2 - k^2) + k^2(k-s) + 2k^2 p \\
&\quad + \lambda v k(k-s) + \lambda v(\lambda v - k)\}(k-s-q)(2k-p-q) \\
&\geq \lambda^4 v^4(k^3 + k^2 - k^2 s + 2k^2 p + \lambda v k^2 - \lambda v k s - \lambda v k)(k-s) \\
&\quad - \lambda^2 v^2 k\{(\lambda^2 v^2 - k^2) + k^2(k-s) + 2k^2 p + \lambda v k(k-s) \\
&\quad + \lambda v(\lambda v - k)\}(k-s)(k+2p) \\
&\geq \lambda^2 v^2 k\{\lambda^2 v^2 - k(k+2p)\}(\lambda v + k)(k-s-1) \\
&\quad + 2\lambda^2 v^2 k p\{\lambda^2 v^2(k-2) - k^2(k+2p)\}
\end{aligned}$$

which is non-negative, since $\lambda^2 v^2 - k(k+2p) = v^2 r^2(k-1)^2/(v-1)^2 - k(k+2p) > k^2\{k(k-2) - 2\} + 4k > 0$, and $\lambda^2 v^2(k-2) - k^2(k+2p) = v^2 r^2(k-1)^2(k-2)/(v-1)^2 - k^2(k+2p) > k^2\{k(k-2)^2 - 2(k-1)\}$ which is positive for $k \geq 4$. For $k = 3$ the original $\lambda^2 v^2(k-2) - k^2(k+2p)$ can be shown to be positive, since $\lambda \geq 2$ when $v = 6$. Thus, $\beta' \geq 0$. Hence the proof is completed. □

Proposition 4.3: In a BIBD(v, b, r, k, λ) with $k+1 \leq v \leq 2k-1$, an inequality $B \geq C$ holds for all s such that $1 \leq s \leq k-2$.

Proof: Since with K as in Lemma 4.5

$$(B-C)/[(\lambda v-k)\{\lambda v(v-1)+k(v-k)\}]$$

$$=\frac{KDE-(\lambda^2v^2-2vk+v^2)H}{HK} \quad (=\beta, \text{ say}),$$

if $\beta \geq 0$, then $B \geq C$. It follows from Lemmas 4.4 and 4.5 that the denominator of β is positive. Now denote the numerator of β by β'. Then, since $k \geq 3$,

$$\begin{aligned}
&\beta'/\{4(k-s)\}\\
&=\lambda^4v^4[\{k(\lambda v-k)(k-s-1)+2k^2(v-k+p-s)\}(\lambda^2v^2-2vk+v^2)\\
&\quad+2v(\lambda^2v^2-k^2)(2k-v)](k-s)+[\{k(\lambda v-k)(k-s-1)\\
&\quad+2k^2(v-k+p-s)\}(\lambda^2v^2-2vk+v^2)+2\lambda^2v^2(\lambda^2v^2-k^2)]\{\lambda^2v^2p\\
&\quad\times(k-s)(p+2q)-\lambda^2v^2k(k-s)(k+2p)+\lambda^2v^2sq(k-q)\}\\
&\quad+pk^2[\{k(\lambda v-k)(k-s-1)+2k^2(v-k+p-s)\}(\lambda^2v^2-2vk+v^2)\\
&\quad+2(\lambda^2v^2-k^2)(2\lambda^2v^2-2vk+v^2)](k-s-q)(2k-p-q)\\
&\geq 2\lambda^2v^2(\lambda^2v^2-k^2)\{v(2k-v)-k(k+2p)+p(p+2q)\}\\
&\quad+k(\lambda v-k)(\lambda^2v^2-2vk+v^2)\{\lambda^2v^2-k(k+2p)\}\\
&=\lambda^2v^2(\lambda v-k)\{\lambda v(\lambda vk-2k^2-4kp)-2k^2(k+2p)\}\\
&\quad+\lambda^2v^2(\lambda v-k)\{2p(2\lambda vq-k^2)+2\lambda vp^2+2kp^2+4kpq+vk(2k-v)\}\\
&\quad+\lambda^2v^2(\lambda v-k)\{2\lambda v^2(2k-v)-k^3\}\\
&\quad+vk^2(\lambda v-k)(k+2p)(2k-v)
\end{aligned}$$

which can be shown to be positive, since the six BIB designs in Lemma 4.1 are excluded from our consideration, by noting the following relations through Lemma 4.3: (i) $\lambda^2v^2>0$, $\lambda v-k>0$, $2k-v\geq 1$, $2\lambda v^2(2k-v)-k^3>0$, (ii) $\lambda v(\lambda vk-2k^2-4kp)-2k^2(k+2p)>k^2\{\lambda vk-2(k+2p)(k+1)\}/2>0$, (iii) $2p(2\lambda vq-k^2)+2\lambda vp^2+2kp^2+4kpq+vk(2k-v)>0$. Thus, $\beta'\geq 0$. Hence the proof is completed. □

Therefore, Propositions 3.1, 4.1, 4.2 and 4.3 can show the validity of the theorem in Section 2. □□

5 Conclusion

In Sections 3 and 4, the conjecture of Das and Kageyama (1992) has been proved mathematically. This means, through Section 3.3 of Das and Kageyama (1992), that extended BIB designs are fairly robust against the unavailability of any number of observations in any one block.

Acknowledgements: The authors are grateful to Dr. Ashish Das for his discussion. Thanks are also due to a referee for his constructive comments and suggestions.

References

Das A and Kageyama S (1992) Robustness of BIB and extended BIB designs against the unavailability of any number of observations in a block. Computational Statistics & Data Analysis 14: 343–358

Raghavarao D (1971) Constructions and Combinatorial Problems in Design of Experiments. John Wiley, New York

Received 27.03.92
Revised version 14.09.92

Book Review

W. J. Anderson (Ed.): Continuous Time Markov Chains: An Applications-Oriented Approach, Springer-Verlag, 1991, 367 pp., DM 128,—

One could quarrel with the choice of sub-title. 'Applications-Oriented' evokes martingales, weak convergence, diffusion approximation, coupling, maybe strong approximation, and their use in describing a host of processes arising in stochastic modelling. Instead, we are told on page 5 that the process X will never figure in any rôle of importance, and that it is enough for applications to know the transition probabilities $P_{ij}(t)$. Well, never mind.

'The P-Q Approach' would be a better sub-title. The main aim of the book is to present the fruits of the Kendall-Reuter-Karlin approach to Markov chains in continuous time. It contains a detailed discussion of the construction of transition semi-groups $P(t)$ from knowledge only of $Q = P'(0)$, starting with the Feller recursion, and continuing to the construction of other solutions, when such exist. Conditions for the existence and uniqueness of solutions, including Hou's theorem in the non-conservative case, receive corresponding emphasis. The classification of states is then introduced. For transient classes, the concepts of λ-recurrence and λ-transience, and their connection with quasi-stationary distributions, are examined: for recurrent classes, there is a chapter on strong and exponential ergodicity. Then come symmetry and reversibility, stochastic monotonicity, dual processes and coupling. In all cases, much effort is given to providing conditions expressed in terms of the Q-matrix for the various phenomena to occur. This is important for the practitioner, who is much more likely to know Q than he is to know $P(t)$. The proofs throughout are analytic rather than probabilistic, and much use is made of resolvents.

There are, nonetheless, applications: to birth and death processes, in the style of Karlin and McGregor, to birth-death-catastrophe processes, to open migration processes and to some ecological and epidemiological processes. Here, however, the philosophy chosen shows its weaknesses, since those models for which the (Laplace transforms of the) transition matrices $P(t)$ can be explicitly calculated are few, and the solutions thereby obtained often uninformative. A prime example is the formula given for the final size distribution in the general stochastic epidemic, involving an ugly and intractable combinatorial sum of products; whereas, if the initial number of susceptibles is at all large, there is a good approximation to the whole evolution of the epidemic, using a mixture of a branching process approximation for those paths that die out early and a diffusion approximation for those that do not.

The strength of the book is clearly in its exposition of P-Q theory rather than in its usefulness in applications; in this respect, it makes a welcome addition to the literature on Markov chains.

Cambridge A. D. Barbour

Metrika (1993) 40:203–210

Metrika

Maximum Likelihood Methods for Fitting the Burr Type XII Distribution to Multiply (Progressively) Censored Life Test Data

DALLAS R. WINGO

Math Stat Associates, Inc., P.O. Box 8385, Red Bank, New Jersey 07701, USA

Summary: This paper develops mathematical and computational methodology for fitting, by the method of maximum likelihood (ML), the Burr Type XII distribution to multiply (or progressively) censored life test data. Mathematical expressions are given for approximating the asymptotic variances and covariances of the ML estimates (MLEs) of the parameters of the Burr Type XII distribution. A rigorous mathematical analysis is undertaken to investigate the existence and uniqueness of the MLEs for arbitrary sample data. The methodology of this paper is applied to progressively censored sample data arising in a life test experiment.

Key Words: Burr distribution, maximum likelihood estimation, multiple censoring, progressively censored samples.

1 Introduction

The Burr Type XII distribution, with cumulative distribution function (CDF)

$$F(x) = 1 - (x^c + 1)^{-k}\ , \qquad x > 0\ , c > 0\ , k > 0 \tag{1.1}$$

and probability density function (PDF)

$$f(x) = ckx^{c-1}/(x^c + 1)^{k+1}, \tag{1.2}$$

is a versatile probability distribution. As shown by Rodriguez (1977) and Tadikamalla (1980), this distribution covers the curve shape characteristics for the Normal, lognormal, Gamma, logistic and exponential (Pearson Type X) distributions, as well as a significant portion of the curve shape characteristics for the Pearson Types I (Beta), II, V, VII, IX and XII families. Because the family of curve shape characteristics exhibited by the Burr Type XII distribution is very rich, this distribution is useful in applications which require a model of functionally simple form but which can also exhibit a wide variety of shapes.

0026–1335/93/3–4/203–210 $2.50

Wingo (1983) has described methods for fitting the Burr Type XII distribution to life test or other (complete sample) data by maximum likelihood (ML) and has also provided an extensive list of references to earlier published work on this distribution. The present paper extends the work of Wingo (1983) by developing mathematical and computational methodology for fitting the Burr Type XII distribution to progressively censored life test or survival data.

2 Notation and Statement of the Problem

Censoring occurs naturally in many life test experiments. Broadly speaking, censoring occurs when exact lifetimes (or survival times) are known only for a portion of the subjects under study. For example, when data from a life test experiment are analyzed, some subjects are still surviving. The survival times of these surviving subjects are known only to be beyond their current lifetimes. These data are called *right-censored* data. When right-censored data have different running times (called *censoring times*) intermixed with exact lifetimes, such data are called *multiply* or *progressively censored.* Multiply censored data arise when subjects go on test at different times. Thus, when the data are recorded, the subjects have different running times. For example, in clinical trials, progressive censoring is often the result of participant dropout.

Let N be the total number of subjects placed on life test, and let n be the number of known, exact survival (or failure) times. Also, suppose censoring occurs progressively in l stages at times T_j, where $T_j < T_{j-1}$, $j = 1, 2, \ldots, l$. If, at stage j of censoring, r_j randomly selected surviving subjects are removed (censored) from further observation, then it follows that

$$N = n + \sum_{j=1}^{l} r_j \ . \tag{2.1}$$

Two types of censoring are generally recognized: Type I and Type II censoring. In Type I progressive censoring, subjects are censored at fixed, pre-determined times T_j. In Type II progressive censoring, the T_j coincide with times of failure and are random variables. For a Type I progressively censored sample, the likelihood function (LF) is

$$L = C \prod_{i=1}^{n} f(x_i) \prod_{j=1}^{l} [1 - F(T_j)]^{r_j} \ , \tag{2.2}$$

where C is a constant, $f(\ )$ is the probability density function and $F(\ )$ is the

cumulative distribution function. For a Type II progressively censored sample, the LF is

$$L = C \prod_{i=1}^{n} f(T_i)[1 - F(T_i)]^{r_i} , \tag{2.3}$$

where, as noted earlier, the T_i are observed values of random variables and coincide with times of failure. Only Type I censoring will be considered in this paper.

As has been pointed out by Nelson (1982), when analyzing progressively censored data by ML it is necessary to assume that subjects censored at any specific time come from the same life distribution as the subjects that survive beyond that time. This assumption does not hold, for example, if subjects are censored when they appear as though they are about to fail.

3 Maximum Likelihood Estimation

3.1 Log-Likelihood Function and Likelihood Equations

For an l-stage progressively censored sample with n independent, completely determined lifetimes $\{x_i > 0\}$ $(i = 1, \ldots, n)$ with probability distribution (1.1)–(1.2) and r_j subjects censored at times $\{T_j > 0\}$ $(j = 1, \ldots, l)$, the log-likelihood function (LLF) is

$$L(c, k) = n \log(c) + n \log(k) + (c - 1)\sum \log(x_i) - (k + 1)\sum \log(x_i^c + 1)$$

$$- k \sum_{j=1}^{l} r_j \log(T_j^c + 1) , \tag{3.1}$$

where $\sum$ denotes the summation over the n observation $\{x_i\}$ and where $\sum_{j=1}^{l}$ denotes the summation over the l censoring times $\{T_j\}$. All logarithms are of the Napierian type.

Differentiation of (3.1) with respect to c and k yields

$$L_c \equiv \frac{n}{c} + \sum \log(x_i) - (k + 1)\sum \frac{x_i^c \log(x_i)}{(x_i^c + 1)} - k \sum_{j=1}^{l} \frac{r_j T_j^c \log(T_j)}{T_j^c + 1} \tag{3.2}$$

and

$$L_k \equiv \frac{n}{k} - \sum \log(x_i^c + 1) - \sum_{j=1}^{l} r_j \log(T_j^c + 1) \ , \tag{3.3}$$

where $L_\alpha \equiv \partial L/\partial\alpha$ for the generic parameter α. The LEs, $L_c = 0$ and $L_k = 0$, are obtained by setting to zero the right-hand sides of (3.2)–(3.3). The MLEs of the parameters of (1.1)–(1.2) are the set(s) of values $\hat{c} > 0$ and $\hat{k} > 0$ which simultaneously satisfy $L_c = 0$ and $L_k = 0$ and for which (3.1) is *globally maximized* over all $c > 0$ and $k > 0$.

The solution of $L_k = 0$ for k yields

$$\hat{k} = \frac{n}{\sum \log(x_i^c + 1) + \sum_{j=1}^{l} r_j \log(T_j^c + 1)} \ . \tag{3.4}$$

The substitution of (3.4) into (3.1) yields the conditional LLF

$$\begin{aligned} L^*(c) &\equiv L(c|\hat{k}(c)) \\ &= \Phi(c) - n \log\left[\sum \log(x_i^c + 1) + \sum_{j=1}^{l} r_j \log(T_j^c + 1)\right] , \end{aligned} \tag{3.5}$$

where

$$\Phi(c) \equiv n[\log(n) - 1] + n \log(c) + (c - 1)\sum \log(x_i) - \sum \log(x_i^c + 1) \tag{3.6}$$

is strictly concave (Mangasarian, 1969) in c for all $c > 0$.

3.2 *Asymptotic Variances and Covariances of MLEs*

An idea of the asymptotic behavior of the MLEs for large samples is obtained from the expected information matrix $I_{ij} \equiv -E\{\partial^2 L/\partial\theta_i\partial\theta_j\}$, where $i, j = 1, 2$ and $\Theta = (c, k)^T$. I is symmetric. Below, we give the observed information matrix $\hat{I}_{ij}$, which is obtained by dropping the expectation operator E. The approximate (i.e., observed) asymptotic variance-covariance matrix $\tilde{V}$ for the MLEs is obtained by inverting the observed information matrix, i.e.,

$$[\tilde{V}_{ij}] = \frac{1}{n}[\hat{I}_{ij}]^{-1} \tag{3.7}$$

where, as noted earlier, $\hat{I}_{ij} = -\{\partial^2 L/\partial\theta_i \partial\theta_j\}$ and $\Theta = (c, k)^T$.

Upon differentiating (3.2)–(3.3) with respect to c and k, we obtain

$$L_{cc} \equiv -\frac{n}{c^2} - (k+1)\sum \frac{x_i^c (\log x_i)^2}{(x_i^c + 1)^2} - k \sum_{j=1}^{l} \frac{r_j T_j^c (\log T_j)^2}{(T_j^c + 1)^2} , \tag{3.8}$$

$$L_{kk} \equiv -\frac{n}{k^2} , \tag{3.9}$$

$$L_{ck} \equiv L_{kc} \equiv -\sum \frac{x_i^c \log(x_i)}{(x_i^c + 1)} - \sum_{j=1}^{l} \frac{r_j T_j^c \log T_j}{T_j^c + 1} , \tag{3.10}$$

where $L_{\alpha\beta} \equiv \partial^2 L/\partial\alpha\partial\beta$ for the generic parameters α and β.

3.3 *Existence and Uniqueness of* MLEs

When c is known and positive, $\hat{k}$ exists, is unique and is given by (3.4). ■

When k is known, fixed and positive, the LLF (3.1) is strictly concave in c, inasmuch as (3.8) implies that $L_{cc} < 0$. It follows from the strict concavity of (3.1) (when $k > 0$ and fixed) that $\hat{c}$ is unique. To show that $\hat{c}$ always exists, we show that the equation $L_c = 0$ (where L_c is given by (3.2)) always has a single, finite zero on the positive real line, as long as $x_i \neq 1$ and/or $T_j \neq 1$ for some $i = 1, \ldots, n$ and $j = 1, \ldots, l$.

Suppose we define

$$\phi(c) \equiv \frac{n}{c} + \sum \log(x_i) - (k+1)\sum \frac{x_i^c \log(x_i)}{x_i^c + 1} - k \sum_{j=1}^{l} \frac{r_j T_j^c (\log T_j)}{T_j^c + 1} , \tag{3.11}$$

where $k > 0$ and fixed. Also, define the index sets

$$M \equiv \{i\,(i = 1, \ldots, n) \mid x_i < 1\} ,$$

$$P \equiv \{i\,(i = 1, \ldots, n) \mid x_i > 1\} ,$$

$$Q \equiv \{j\,(j = 1, \ldots, l) \mid T_j > 1\} .$$

Now, direct calculation shows that, as $c \to \infty$,

$$\sum\left(\frac{x_i^c}{x_i^c + 1}\right)\log(x_i) \to \sum_{i \in P} \log(x_i) ,$$

$$\sum_{j=1}^{l} r_j\left(\frac{T_j^c}{T_j^c + 1}\right)\log(T_j) \to \sum_{j \in Q} r_j \log(T_j) .$$

Direct calculation also shows that

$$\begin{aligned} \phi(0) &= \infty , \\ \phi(\infty) &= \sum_{i \in M} \log(x_i) - k\left[\sum_{i \in P} \log(x_i) + \sum_{j \in Q} r_j \log(T_j)\right] . \end{aligned} \tag{3.12}$$

A close examination of (3.12) reveals that its right-hand side will always be negative, as long as either of the index sets M, P or Q is nonempty. Under these conditions, it therefore follows that $\phi(\infty) < 0$ for all $c > 0$. Since $\phi(0) = \infty$ and $\phi(\infty) < 0$, there exists at least one real root of $\phi(c) = 0$ on the positive real line. To show that this root $\hat{c}$ is unique, it suffices to show that $d\phi(c)/dc < 0$, i.e., that $\phi(c)$ is monotone decreasing in c for $c > 0$. Now, straightforward differentiation of $\phi(c)$ shows that $d\phi(c)/dc < 0$, since $\sum x_i^c \log(x_i)/(x_i^c + 1)$ and $\sum_{j=1}^{l} r_j T_j^c \log(T_j)/(T_j^c + 1)$ are both monotonically increasing in c for $c > 0$. Hence, when k is known, fixed and positive, then, under the above conditions on the sample data, the MLE $\hat{c}$ always exists and is unique. ■

When c and k are both unknown, $\hat{c}$ can, in principle, be obtained by maximizing the conditional LLF $L^*(c)$ in (3.5) over all $c > 0$. It turns out that, when both c and k are unknown, $L^*(c)$ is unimodal, provided that certain conditions on the sample data are satisfied. We shall prove this assertion by examining the behaviour of the equation $\phi(c) = 0$, where $\phi(c) \equiv dL^*(c)/dc$. More specifically, we prove below that the equation $\phi(c) = 0$ always possesses a single, finite zero on the positive real line if and only if the set M is nonempty. Let

$$\begin{aligned} \phi(c) = \frac{n}{c} + \sum \log(x_i) - \sum \frac{x_i^c \log(x_i)}{x_i^c + 1} \\ - \frac{n\left\{\sum \dfrac{x_i^c \log(x_i)}{x_i^c + 1} + \displaystyle\sum_{j=1}^{l} \frac{r_j T_j^c \log(T_j)}{T_j^c + 1}\right\}}{\sum \log(x_i^c + 1) + \displaystyle\sum_{j=1}^{l} r_j \log(T_j^c + 1)} . \end{aligned} \tag{3.13}$$

Now, direct calculation shows that

$$\phi(0) = \infty \ ,$$
$$\phi(\infty) = \sum_{i \in M} \log(x_i) \ . \tag{3.14}$$

As before, the right-hand side of (3.14) will be negative if and only if the set M is nonempty. Therefore, provided that these conditions on the sample data hold, $\phi(\infty) < 0$ for all $c > 0$. Since $\phi(0) = \infty$ and $\phi(\infty) < 0$, there exists at least one positive real root of the equation $\phi(c) = 0$, where $\phi(c)$ is defined by (3.13). Differentiation and straightforward (though tedious) algebraic manipulations show that $d\phi(c)/dc < 0$ for all $c > 0$, implying that $\phi(c) = 0$ has a single, finite zero on the positive real line. This zero is the MLE $\hat{c}$. ■

For the case in which both c and k are unknown and must be estimated jointly, it is instructive to compare the conclusions drawn in the present paper with those of Theorem 1 of Wingo (1983). It should be noted that Theorem 1 of Wingo (1983) contains a serious misprint that renders the theorem false. However, all is not lost. If every occurrence of the expression "$x_i \geq 1$" is replaced by "$x_i \leq 1$", then Theorem 1 remains valid.

4 Example

Table 1 gives data from a clinical trial that was conducted to assess the effectiveness of an anesthetic antibiotic ointment in relieving pain caused by superficial skin wounds. Because of its flexibility to model a variety of clinical data, the Burr Type XII distribution was used as a tentative model for the distribution of patient relief times.

Table 1. Pain Relief Times (in Hours) for 20 Patients

0.828	0.881	1.138	0.879	0.554	0.653	0.698	0.566	0.665	0.917
0.529	0.786	1.110	0.866	1.037	0.788	1.050	0.899	0.683	0.829

A total of 30 patients was involved in the clinical trial. Table 1 gives the time to pain reflief (in hours) of 20 patients. The remaining 10 patients were subjected to the following Type I progressive censoring scheme: $l = 3$, $T_1 = 0.25$, $T_2 = 0.50$, $T_3 = 0.75$, $r_1 = 5$, $r_2 = 1$ and $r_3 = 4$. The computed MLEs for these data are

$\hat{c} = 6.560$ and $\hat{k} = 2.597$. Approximate variances and covariances of the MLEs are $\text{var}\{\hat{c}\} = 0.014$, $\text{var}\{\hat{k}\} = 0.017$ and $\text{cov}\{\hat{c}, \hat{k}\} = 0.002$.

If we define $\mu = E\{x\}$, then the MLE of the distribution mean is (Wingo, 1983)

$$\hat{\mu} = \frac{\Gamma\left(\hat{k} - \frac{1}{\hat{c}}\right)\Gamma\left(1 + \frac{1}{\hat{c}}\right)}{\Gamma(\hat{k})} . \tag{3.15}$$

The estimated average pain relief time is computed from (3.15) as $\hat{\mu} = 0.817$ hours.

5 Conclusions

This paper has described mathematical methodology for estimating the parameters of the Burr Type XII distribution from progressively censored sample data. Mathematical expressions are given for the approximate (large-sample) variances and covariances of the MLEs. Necessary and sufficient conditions on the sample data were derived which ensure that the MLEs exist, are unique and finite. The results of the present paper are therefore expected to facilitate the computational calculations faced by clinicians or other researchers who wish to use the Burr Type XII distribution as a tentative model to describe life test or other experimental data.

References

Mangasarian, OL (1969) Nonlinear Programming. Mc-Graw Hill, New York

Nelson, W (1982) Applied Life Data Analysis. John Wiley & Sons, New York

Rodriguez, RN (1977) A guide to the Burr Type XII distributions. Biometrika 64:129–134

Tadikamalla, PR (1980) A look at the Burr and related distributions. Internat. Statist. Review 48:337–344

Wingo, DR (1983) Maximum likelihood methods for fitting the Burr Type XII distribution to life test data. Biom J. 25:77–84

Received 19.03.1992
Revised version 25.09.1992

Metrika (1993) 40:211–221

Metrika

Some Statistical Problems in the Presence of an Outlier When Sampling from Truncation Parameter Densities

VIJAY ROHATGI

Departmént of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403-0221, USA

KANDASAMY SELVAVEL

Department of Mathematics, Claflin College, Orangeburg, SC 29115, USA

Summary: In this paper we derive some recurrence relations for moments of order statistics of a random sample from a truncation parameter density when one of the observations is an outlier. We also derive uniform minimum variance unbiased estimator of a parametric function.

1 Introduction

A variety of outlier models have been the subject of discussion in the literature. We restrict attention to the case when the outlier has a distribution that is different from the population distribution. An outlier is not necessarily a bad or an erroneous observation; it may be an indication of some unexpected useful industrial treatment or successful agricultural variety. One approach is to remove such outliers by means of some test of significance and to base analysis only on the remaining observations. In this study, however, we shall be interested in the estimation of parameters and the moments of order statistics based on the complete sample.

We consider one – truncation parameter probability density functions (pdfs) of the forms

$$f(x;\theta) = \begin{cases} q(\theta)h(x), & a < \theta < x < b \\ 0, & \text{otherwise} \end{cases},$$

and

$$f(x;\theta) = \begin{cases} q(\theta)h(x), & a < x < \theta < b \\ 0, & \text{otherwise} \end{cases},$$

where $-\infty < a < b < \infty$ are known constants, $h(x)$ is a positive, absolutely continuous function and $q(\theta)$ is everywhere differentiable. In Section 2 we obtain some recurrence relations between moments of order statistics in the presence of

0026–1335/93/3–4/211–221 $2.50

an outlier. In Section 3 we derive UMVU estimator of a U-estimable function, say $g(\theta)$. Some applications of the theorems are given in Section 4.

2 Recurrence Relations for Moments of Order Statistics

Recurrence relations for moments of order statistics have been the subject of intensive investigation. Balakrishnan (1987) derived two simple identities involving single moments of order statistics from a sample of size n in the presence of an outlier. His results generalize those of Joshi (1973).

In this section we obtain some new recurrence relations between single moments for the following outlier model. Let

$$f_1(x;\theta) = \begin{cases} q_1(\theta)h_1(x), & a < \theta < x < b \\ 0, & \text{otherwise} \end{cases}, \tag{2.1}$$

and

$$f_2(x;\theta) = \begin{cases} q_2(\theta)h_2(x), & a < \theta < x < b \\ 0, & \text{otherwise} \end{cases}. \tag{2.2}$$

Let us represent the sample by n (≥ 2) independent random variables X_r ($r = 1, 2, \ldots, n-1$) and Y, such that X_r has pdf $f_1(x;\theta)$ and Y has pdf $f_2(x;\theta)$. Furthermore, let $Z_{1:n} \leq Z_{2:n} \leq \cdots \leq Z_{n:n}$ be the order statistics obtained by arranging the n independent observations in increasing order of magnitude. Then the pdf $h_{i:n}(x)$ of $Z_{i:n}$ ($1 \leq i \leq n$) is given by (David and Shu (1978))

$$\begin{aligned} h_{i:n}(x) = q_1^{n-1}(\theta)q_2(\theta)\Bigg[& \frac{(n-1)!}{(i-2)!\,(n-i)!}\left\{\int_\theta^x h_1(u)du\right\}^{i-2} \\ & \cdot\left\{\int_x^b h_1(u)du\right\}^{n-i}\left\{\int_\theta^x h_2(u)du\right\}h_1(x) \\ & + \frac{(n-1)!}{(i-1)!\,(n-i)!}\left\{\int_\theta^x h_1(u)du\right\}^{i-1}\left\{\int_x^b h_1(u)du\right\}^{n-i}h_2(x) \\ & + \frac{(n-1)!}{(i-1)!\,(n-i-1)!}\left\{\int_\theta^x h_1(u)du\right\}^{i-1}\left\{\int_x^b h_1(u)du\right\}^{n-i-1} \\ & \cdot\left\{\int_x^b h_2(u)du\right\}h_1(x)\Bigg], \text{ if } x > \theta\ , \end{aligned}$$

and

$$= 0 \quad \text{otherwise} ,$$

where the first term drops out if $i = 1$, and the last if $i = n$. We denote the k^{th} order moment $E(Z_{i:n}^k)$ of $Z_{i:n}$ by $\mu_{i:n}^{(k)}(\theta)$ $(1 \leq i \leq n)$. Further, let $f_{i:n}(x)$ be the pdf of the i^{th} order statistic in a sample of size n from a continuous population with pdf $f_1(x; \theta)$, and denote by

$$v_{i:n}^{(k)}(\theta) = \int_{\theta}^{b} x^k f_{i:n}(x) dx,$$

the moment of order k of $X_{i:n}$. For convenience we denote the antiderivative of $f_i(\theta; \theta)$, $i = 1, 2$ by

$$d_i(\theta) = \int f_i(\theta; \theta) d\theta , \qquad i = 1, 2 \tag{2.3}$$

and for $r \geq 0$, let

$$J_r = r d_1(\theta) + d_2(\theta) . \tag{2.4}$$

Let f denote the integral operator defined by

$$f\omega(\theta) = \int e^{-d_1(\theta)} f_1(\theta; \theta) \omega(\theta) d\theta \tag{2.5}$$

and for $n \geq 1$ define f^n iteratively as follows.

$$f^1\omega(\theta) = f\omega(\theta) \quad \text{and} \quad f^m\omega(\theta) = f(f^{m-1}\omega(\theta)), \qquad \text{if } m > 1 . \tag{2.6}$$

We find a relation between $\mu_{i:n}^{(k)}(\theta)$ and f operating on $\mu_{2:n-i+2}^{(k)}(\theta)$ and $v_{i-r:n-r}^{(k)}(\theta)$.

Theorem 2.1: Let d_i, $i = 1, 2$, J_r and f be as defined in (2.3) to (2.6). Then for nonnegative integers k and i $(2 \leq i < n)$

$$\mu_{i:n}^{(k)}(\theta) = (-1)^{i-1} \frac{(n-1)!}{(n-i)!} e^{J_{n-1}} f^{i-1} \left\{ e^{-J_{n-i}} \mu_{1:n-i+1}^{(k)}(\theta) \right\}$$

$$+ e^{J_{n-1}} \sum_{r=1}^{i-1} (-1)^r \frac{(n-1)!}{(n-r)!} f^r \left\{ e^{-J_{n-r-1}} \frac{f_2(\theta; \theta)}{f_1(\theta; \theta)} v_{i-r:n-r}^{(k)}(\theta) \right\}$$

and

$$\mu_{n:n}^{(k)}(\theta) = (-1)^{n-2}(n-1)!\, e^{J_{n-1}} f^{n-2}\left\{e^{-d_1-d_2}\mu_{2:2}^{(k)}(\theta)\right\}$$

$$+ e^{J_{n-1}} \sum_{r=1}^{n-2} (-1)^r \frac{(n-1)!}{(n-r)!} f^r \left\{e^{-(n-r-1)d_1-d_2} \frac{f_2(\theta;\theta)}{f_1(\theta;\theta)} \nu_{n-r:n-r}^{(k)}(\theta)\right\} . \tag{2.7}$$

Proof: For any nonnegative integer k

$$\mu_{i:n}^{(k)}(\theta) = \int_{\theta}^{b} x^k h_{i:n}(x)dx \ . \tag{2.8}$$

Note that

$$q_i(\theta) = \left\{\int_{\theta}^{b} h_i(u)du\right\}^{-1} , \qquad i = 1, 2 \tag{2.9}$$

and

$$q_i'(\theta) = \frac{dq_i(\theta)}{d\theta} = q_i^2(\theta)h_i(\theta) , \qquad i = 1, 2 \tag{2.10}$$

Differentiating (2.8) with respect to θ, and using (2.9) and (2.10) we get the following differential equation

$$\frac{d\mu_{i:n}^{(k)}(\theta)}{d\theta} - [(n-1)f_1(\theta;\theta) + f_2(\theta;\theta)]\mu_{i:n}^{(k)}(\theta)$$

$$= -(n-1)f_1(\theta;\theta)\mu_{i-1:n-1}^{(k)}(\theta) - f_2(\theta;\theta)\nu_{i-1:n-1}^{(k)}(\theta) \ . \tag{2.11}$$

Solving (2.11) we get

$$\mu_{i:n}^{(k)}(\theta) = -e^{J_{n-1}(\theta)} \int e^{-J_{n-1}(\theta)}(n-1)f_1(\theta;\theta)\mu_{i-1:n-1}^{(k)}(\theta)d\theta$$

$$- e^{J_{n-1}(\theta)} \int e^{-J_{n-1}(\theta)} f_2(\theta;\theta)\nu_{i-1:n-1}^{(k)}(\theta)d\theta \ . \tag{2.12}$$

Using the definition of f and iterating (2.12) we get (2.7).

Theorem 2.2: Let d_i, $i = 1, 2$, J_r and f be as defined in (2.3) to (2.6). Then for any nonnegative integer k

$$\mu_{1:n}^{(k)}(\theta) = -e^{J_{n-1}(\theta)} \int \theta^k e^{-J_{n-1}(\theta)} [(n-1) f_1(\theta; \theta) + f_2(\theta; \theta)] d\theta \ .$$

Proof: For any nonnegative integer k

$$\begin{aligned} \mu_{1:n}^{(k)}(\theta) &= EZ_{1:n}^k \\ &= q_1^{n-1}(\theta) q_2(\theta) \int_\theta^b x^k \left[\left\{ \int_x^b h_1(u) du \right\}^{n-1} h_2(x) \right. \\ &\quad \left. + (n-1) \left\{ \int_x^b h_1(u) du \right\}^{n-2} \left\{ \int_x^b h_2(u) du \right\} h_1(x) \right] dx \ . \end{aligned} \tag{2.13}$$

Differentiating (2.13) with respect to θ, and using (2.9) and (2.10) we get

$$\begin{aligned} \frac{d\mu_{1:n}^{(k)}(\theta)}{d\theta} &= [q_2(\theta) h_2(\theta) + (n-1) q_1(\theta) h_1(\theta)] \theta^k \\ &\quad + [(n-1) q_1(\theta) h_1(\theta) + q_2(\theta) h_2(\theta)] \mu_{1:n}^{(k)}(\theta) \ . \end{aligned}$$

That is

$$\begin{aligned} &\frac{d\mu_{1:n}^{(k)}(\theta)}{d\theta} - [(n-1) q_1(\theta) h_1(\theta) + q_2(\theta) h_2(\theta)] \mu_{1:n}^{(k)}(\theta) \\ &= [(n-1) q_1(\theta) h_1(\theta) + q_2(\theta) h_2(\theta)] \theta^k \ . \end{aligned}$$

Solving the above differential equation we get

$$\begin{aligned} \mu_{1:n}^{(k)}(\theta) &= -e^{J_{n-1}(\theta)} \int \theta^k [(n-1) q_1(\theta) h_1(\theta) + q_2(\theta) h_2(\theta)] e^{-J_{n-1}(\theta)} d\theta \\ &= -e^{J_{n-1}(\theta)} \int \theta^k [(n-1) f_1(\theta; \theta) + f_2(\theta; \theta)] e^{-J_{n-1}(\theta)} d\theta \ . \end{aligned}$$

Remark 2.1: When $f_1(x; \theta) = f_2(x; \theta)$ we get

$$\mu_{1:n}^{(k)}(\theta) = -n e^{n d_1(\theta)} \int f_1(\theta; \theta) \theta^k e^{-n d_1(\theta)} d\theta \ .$$

Let

$$f_3(x;\theta) = \begin{cases} q_3(\theta)h_3(x), & a < x < \theta < b \\ 0 & \text{otherwise} \end{cases},$$

and

$$f_4(x;\theta) = \begin{cases} q_4(\theta)h_4(x), & a < x < \theta < b \\ 0, & \text{otherwise} \end{cases}.$$

For convenience we denote the antiderivative of $f_i(\theta;\theta)$, $i = 3, 4$ by

$$d_i(\theta) = \int f_i(\theta;\theta)d\theta \ , \qquad i = 3, 4 \tag{2.14}$$

and let

$$K_r(\theta) = rd_3(\theta) + d_4(\theta), \qquad \text{for } r \geq 1 \ . \tag{2.15}$$

Let I denote the integral operator defined by

$$I\omega(\theta) = \int e^{d_3(\theta)} f_3(\theta;\theta)\omega(\theta)d\theta \tag{2.16}$$

and for $1 \leq n$ define I^n iteratively as follows.

$$I^1\omega(\theta) = I\omega(\theta) \quad \text{and} \quad I^n\omega(\theta) = I(I^{n-1}\omega(\theta)) \ , \quad \text{if } n > 1 \ . \tag{2.17}$$

Then the following analogous result holds for the pdfs $f_3(x;\theta)$ and $f_4(x;\theta)$.

Theorem 2.3: Let d_i, $i = 3, 4$, $K_r(\theta)$ and I be as defined in (2.14) to (2.17). Then for any nonnegative integer k and for nonnegative integer $1 \leq i \leq n$

$$\mu_{i:n}^{(k)}(\theta) = \frac{(n-1)!}{(i-1)!} e^{-K_{n-1}(\theta)} I^{n-1}\left(e^{K_{i-1}(\theta)}\mu_{i:i}^{(k)}(\theta)\right)$$

$$+ e^{-K_{n-1}(\theta)} \sum_{r=1}^{n-i} \frac{(n-1)!}{(n-r)!} I^r\left(\frac{d_4}{d_3} e^{K_{n-r-1}(\theta)} \nu_{i:n-r}^{(k)}(\theta)\right) .$$

Applications of Theorem 2.1 will be given in Section 4.

3 Unbiased Estimation of Parameters

The problem of estimation of parameters in the presence of outliers has been considered by many authors. Dixit (1989) derived the maximum likelihood estimators and moment estimators of parameters for samples from a gamma distribution in the presence of outliers. Joshi (1988) discussed several estimators of the mean of an exponential distribution, when an unidentified single outlier is present. In this section we derive UMVU estimators of parameters from truncation parameter families in the presence of an outlier.

We see from Section 2 that the pdfs $h_{1:n}(x)$ of $Z_{1:n}$ $(n \geq 2)$ and $h_{n:n}(x)$ of $Z_{n:n}$ $(n \geq 2)$ are given by

$$h_{1:n}(x) = (n-1)q_1^{n-1}(\theta)q_2(\theta)\left\{\int_x^b h_1(u)du\right\}^{n-2}\left\{\int_x^b h_2(u)du\right\}h_1(x)$$

$$+ q_1^{n-1}(\theta)q_2(\theta)\left\{\int_x^b h_1(u)du\right\}^{n-1}h_2(x), \qquad \text{if } \theta < x\ ,$$

and

$$h_{n:n}(x) = (n-1)q_1^{n-1}(\theta)q_2(\theta)\left\{\int_\theta^x h_1(u)du\right\}^{n-2}\left\{\int_\theta^x h_2(u)du\right\}h_1(x)$$

$$+ q_1^{n-1}(\theta)q_2(\theta)\left\{\int_\theta^x h_1(u)du\right\}^{n-1}h_2(x), \qquad \text{if } \theta < x\ ,$$

respectively.

Theorem 3.1: Let $Z_{1:n} \leq Z_{2:n} \leq \cdots \leq Z_{n:n}$ be as defined in Section 2. Then the UMVU estimator of a U-estimable function $g(\theta)$ is given by

$$\phi(Z_{1:n}) = -\frac{q_1^{n-1}(Z_{1:n})q_2(Z_{1:n})}{[(n-1)f_1(Z_{1:n}, Z_{1:n}) + f_2(Z_{1:n}, Z_{1:n})]}$$

$$\cdot\frac{d}{d\theta}\left(\frac{g(\theta)}{q_1^{n-1}(\theta)q_2(\theta)}\right)\Bigg|_{\theta=Z_{1:n}} \tag{3.1}$$

Proof: We can easily prove that $Z_{1:n}$ is complete for θ using equation (3.3) below. In view of Lehmann-Scheffe theorem it follows that the UMVU estimator of a U-estimable function $g(\theta)$ is given by the solution of the following equation.

$$E\phi(Z_{1:n}) = g(\theta) \ .$$

That is

$$\int_{\theta}^{\infty} \phi(x)\left(\int_{x}^{\infty} h_1(u)du\right)^{n-2}\left[\left(\int_{x}^{\infty} h_1(u)du\right)h_2(x) + (n-1)\left(\int_{x}^{\infty} h_2(u)du\right)h_1(x)\right]dx = \frac{g(\theta)}{q_1^{n-1}(\theta)q_2(\theta)} \ . \tag{3.2}$$

Differentiating (3.2) with respect to θ we get

$$\phi(\theta)\frac{[(n-1)f_1(\theta;\theta) + f_2(\theta;\theta)]}{q_1^{n-1}(\theta)q_2(\theta)} = -\frac{d}{d\theta}\left(\frac{g(\theta)}{q_1^{n-1}(\theta)q_2(\theta)}\right) \ .$$

That is

$$\phi(\theta) = -\frac{q_1^{n-1}(\theta)q_2(\theta)}{[(n-1)f_1(\theta;\theta) + f_2(\theta;\theta)]}\frac{d}{d\theta}\left(\frac{g(\theta)}{q_1^{n-1}(\theta)q_2(\theta)}\right) \ . \tag{3.3}$$

Substituting $Z_{1:n}$ for θ in (3.3) we get the UMVU estimator of $g(\theta)$.

Let f_3 and f_4 be as defined in Section 2. Then the following analogous result holds for the pdfs f_3 and f_4.

Theorem 3.2: Let $Z_{1:n} \le Z_{2:n} \le \cdots \le Z_{n:n}$ be as defined in Section 2. Then the UMVU estimator of a U-estimable function $g(\theta)$ is given by

$$\phi(Z_{n:n}) = \frac{q_1^{n-1}(Z_{n:n})q_2(Z_{n:n})}{[(n-1)f_1(Z_{n:n}, Z_{n:n}) + f_2(Z_{n:n}, Z_{n:n})]} \cdot \frac{d}{d\theta}\left(\frac{g(\theta)}{q_1^{n-1}(\theta)q_2(\theta)}\right)\Bigg|_{\theta=Z_{nn}} \ .$$

4 Some Applications

Example 4.1: As an example of Theorem 2.1, consider the Pareto pdfs

$$f_1(x;\theta,\alpha_1) = \begin{cases} \alpha_1\theta^{\alpha_1}x^{-\alpha_1-1}, & \text{if } x > \theta \\ 0, & \text{otherwise} \ , \end{cases}$$

and

$$f_2(x;\theta,\alpha_2)=\begin{cases}\alpha_2\theta^{\alpha_2}x^{-\alpha_2-1}, & \text{if } x>\theta\\ 0, & \text{otherwise}\end{cases}.$$

We see easily that

$$\mu^{(k)}_{2:2}(\theta)=\frac{\alpha_2\theta^{k+\alpha_2-\alpha_1}}{\alpha_1(2-k/\alpha_1)(1-k/\alpha_1)}+\frac{\alpha_1\theta^{k+\alpha_1-\alpha_2}}{\alpha_2(2-k/\alpha_2)(1-k/\alpha_2)}, \qquad \text{if } \alpha_1,\alpha_2>k$$

and

$$v^{(k)}_{r:r}(\theta)=\frac{\Gamma(r+1)\Gamma(1-k/\alpha_2)}{\Gamma(r+1-k/\alpha_2)}\theta^k, \qquad \text{if } \alpha_2>k.$$

Using the values of $\mu^{(k)}_{2:2}(\theta)$ and $v^{(k)}_{r:r}(\theta)$ given above, and applying Theorem 2.1 we get

$$\begin{aligned}\mu^{(k)}_{n:n}(\theta)=&\frac{(n-1)\alpha_2\Gamma(1-k/\alpha_1)\theta^k}{\alpha_1\Gamma(3-k/\alpha_1)\binom{n+\alpha_2/\alpha_2-k/\alpha_1-1}{n-2}}\\ &+\frac{(n-1)\alpha_1\Gamma(1-k/\alpha_2)\theta^k}{\alpha_2\Gamma(3-k/\alpha_2)\binom{n+2\alpha_2/\alpha_1-k/\alpha_1-2}{n-2}}\\ &+\frac{\alpha_2}{\alpha_1}\Gamma(n)\Gamma(1-k/\alpha_1)\theta^k\\ &\cdot\sum_{r=1}^{n-2}\frac{1}{\Gamma(r+1)\Gamma(n-r-k/\alpha_2+1)\binom{n+\alpha_2/\alpha_1/k/\alpha_1-1}{r}}\end{aligned}$$

if $\alpha_1,\alpha_2>k$.

Note that when $f_1(x;\theta,\alpha_1)=f_2(x;\theta,\alpha_2)$, that is when $\alpha_1=\alpha_2$ we get

$$\mu^{(k)}_{n:n}(\theta)=\frac{\Gamma(n+1)\Gamma(1-k/\alpha_1)}{\Gamma(n-k/\alpha_1+1)}\theta^k, \qquad \text{if } \alpha_1>k.$$

Example 4.2: We now compute $\mu_{1:n}^{(k)}(\theta)$ for the Pareto pdfs considered in Example 4.1. Using Theorem 2.2 we get

$$\mu_{1:n}^{(k)}(\theta) = -e^{(n-1)\alpha_1 \ln\theta + \alpha_2} \int \theta^k e^{-\alpha_1 \ln\theta - \alpha_2 \ln\theta} \left[(n-1)\frac{\alpha_1}{\theta} + \frac{\alpha_2}{\theta} \right] d\theta$$

$$= -[(n-1)\alpha_1 + \alpha_2]\theta^{(n-1)\alpha_1 + \alpha_2} \int \theta^{k-(n-1)\alpha_1 - \alpha_2 - 1} d\theta$$

$$= \frac{[(n-1)\alpha_1 + \alpha_2]}{[(n-1)\alpha_1 + \alpha_2 - k]} \theta^k , \qquad \text{if } (n-1)\alpha_1 + \alpha_2 > k .$$

Note that when $f_1(x; \theta, \alpha_1) = f_2(x; \theta, \alpha_2)$, that is when $\alpha_1 = \alpha_2$ we get

$$\mu_{1:n}^{(k)}(\theta) = \frac{n\alpha_1}{n\alpha_1 - k} \theta^k , \qquad \text{if } n\alpha_1 > k .$$

Example 4.3: As an example of Theorem 3.1, consider the following pdfs.

$$f_1(x; \theta, \alpha) = \begin{cases} \alpha\theta^\alpha x^{-\alpha-1}, & \text{if } x > 0 \\ 0, & \text{otherwise} , \end{cases}$$

where α is a known constant, and

$$f_2(x; \theta) = e^{-(x-\theta)}, \quad \text{if } x > \theta \quad \text{and} \quad = 0, \quad \text{otherwise} .$$

Let $g(\theta) = \theta$. Then using Theorem 3.1 we have

$$\phi(\theta) = -\frac{(\alpha\theta^\alpha)^{n-1} e^\theta}{(n-1)\alpha/\theta + 1} \frac{d}{d\theta} \left(\frac{\theta}{(\alpha\theta^\alpha)^{n-1} e^\theta} \right) = \theta - \frac{\theta}{\theta + (n-1)\alpha} .$$

Hence, the UMVU estimator of θ is given by $Z_{1:n} - \dfrac{Z_{1:n}}{(n-1)\alpha + Z_{1:n}}$.

References

Balakrishnan N (1987) Two identities involving order statistics in the presence of an outlier. *Commun. Statist.-Theor. aand Metho.* 16(8): 2385–2389

David HA, Shu VS (1978) Robustness of location estimators in the presence of an outlier. In: David, HA (Ed.) *Contributions to survey sampling and applied statistics: Papers in honor of HO Hartley*, 235–250, Academic Press, New York

Dixit UJ (1989) Estimation of parameters of the gamma distribution in the presence of outliers. *Commun. Statist.-Theor. and Metho.* 18(8): 3071–3085

Joshi PC (1973) Two identities involving order statistics. *Biometrika* 60: 428–429

Joshi PC (1988) Estimation and testing under exchangeable exponential model with a single outlier. *Commun. Statist.-Theor. and Metho.* 17(7): 2315–2326

Received 28.09.1992

Book Review

K. V. Mardia (Ed.): The Art of Statistical Science. A Tribute to G. S. Watson, John Wiley & Sons, 1992, 317 pp., £49, 95

Geoffrey Stuart Watson turned 70 on December 3, 1991. Twenty-five colleagues and friends took the opportunity to honor and dedicate this book to this most remarkable statistician, who is soon to retire as professor of statistics at Princeton University. A total of seven subdivisions of this book reflect the wide range of interests, inspirations, and contributions by G. S. Watson to the field of statistics.

In part I, time series are treated by T. W. Anderson, J. Durbin, and E. J. Hannan. Directional data are discussed in part II by N. I. Fisher and P. Hall, E. Ronchetti, and M. A. Stephens. In part III, J . Aitchison, D. G. Kendall and J. T. Kent review compositional and shape data analysis. Five chapters in part IV contain discussions of technical problems in inference by O. E. Barndorff-Nielsen and P. Bloesild, R. J. Beran and P. W. Millar, H. E. Daniels, and J. W. Tukey. Part V is devoted to spatial statistics, with contributions from N. A. C. Cressie and M. O. Grondona, C. Jennison and B. W. Silverman. W. J. Ewens and S. Karlin discuss statistics and genetics in part VI. Part VII, the final section, is concerned with case studies on issues of public policy with chapters by N. T. J. Bailey and P. Bloomfield.

Geoffrey Watson is not only praised for his great and lasting impact in a diversity of areas of statistics ranging from econometrics, physics, and biology to earth sciences and space; he is also acclaimed as a successful painter. This volume is an eloquent and comprehensive tribute to a great scientist and a most amiable friend, to whom many applicants of his various forms of art, including the referee, are greatly indebted.

Four photos document the ontogeny of the boy from the Australian bush to Prof. Dr. G. S. Watson.

Tübingen K. Schmidt-Koenig

Metrika (1993) 40:223–235

Metrika

The φ-Divergence Statistic in Bivariate Multinomial Populations Including Stratification[1]

L. Pardo, D. Morales

Departamento de Estadística e I.O., Facultad de Matemáticas, Universidad Complutense de Madrid, 28040 Madrid, Spain

M. Salicrú

Departamento de Estadística, Universidad de Barcelona, 08028 Barcelona, Spain

M. L. Menéndez

Departamento de Matemática Aplicada, E.T.S. de Arquitectura, Universidad Politécnica de Madrid, 28040 Madrid, Spain

Abstract: φ-divergence statistics quantify the divergence between a joint probability measure and the product of its marginal probabilities on the basis of contingency tables. Asymptotic properties of these statistics are investigated either considering random sampling or stratified random sampling with proportional allocation and independence among strata. To finish same tests of hypotheses of independence are presented.

1 Introduction

Let X and Y be two random variables taking on the values x_i $(i = 1, \ldots, M)$ and y_j $(j = 1, \ldots, M)$ with probabilities $P = (p_1, \ldots, p_M)$ and $Q = (q_1, \ldots, q_M)$, respectively. Csiszár (1969) defined a measure of information in Q about P, or φ-divergence between P and Q, as follows

$$I^C(P \| Q) = \sum_{i=1}^{M} q_i \varphi\left(\frac{p_i}{q_i}\right)$$

where φ is a real valued convex function on $(0, \infty)$ with

$$\lim_{u \to 0} \varphi(u) = \varphi(0) , \qquad 0\, \varphi\left(\frac{0}{0}\right) = 0 , \qquad 0\, \varphi\left(\frac{a}{0}\right) = a \lim_{u \to \infty} \frac{\varphi(u)}{u} .$$

[1] The research in this paper was supported in part by DGICYT Grants No. PS91-0387 and No. PB91-0155. Their financial support is gratefully acknowledged.

0026–1335/93/3–4/223–235 $2.50

φ-divergence statistics are obtained by either replacing both distributions involved in the argument of the φ-divergence measure by their sample estimates or by replacing one distribution and considering the other as given. In Zografos, Ferentinos and Papaioannou (1990) the sampling properties of estimated divergence type measures are studied. Approximate means and variances are derived and asymptotic distributions are obtained. Furthermore, tests of goodness of fit of observed frequencies to expected ones and tests of equality of divergences based on two or more multinomial samples are given by them. With a different family of divergence measures, Morales et al. (1992) have studied the same problem in a stratified random sampling with proportional allocation and independence among strata. Other interesting works in this line can be seen in Menéndez et al. (1991) and Salicrú et al. (1992).

In this paper we will consider a random variable (X, Y) of the discrete type taking on pairs of values $(x_i, y_j), i = 1, \ldots, M$, and $j = 1, 2, \ldots, K$. We denote by

$$P_{XY} = (p_{ij})_{\substack{i=1,\ldots,M \\ j=1,\ldots,K}} = (P(X = x_i, Y = y_j))_{\substack{i=1,\ldots,M \\ j=1,\ldots,K}}$$

the joint probability mass function of (X, Y) and by $P_X = (p_{i.})_{i=1,\ldots,M}$ and $P_Y = (p_{.j})_{j=1,\ldots,K}$ the corresponding marginal probability distributions, respectively, i.e. $p_{i.} = \sum_{j=1}^{K} p_{ij}$ and $p_{.j} = \sum_{i=1}^{M} p_{ij}$.

Csiszar's measure of information in $P_X * P_Y$ about P_{XY} is given by

$$I^{\varphi}(P_{XY} \| P_X * P_Y) = \sum_{i=1}^{M} \sum_{j=1}^{K} p_{i.} p_{.j} \varphi\left(\frac{p_{ij}}{p_{i.} p_{.j}}\right)$$

In this paper we analyze the properties of the analogue estimate of $I^{\varphi}(P_{XY} \| P_X * P_Y)$ in a simple and stratified random sampling set up as well as its applications to test statistical hypotheses.

As we were informed by the referees similar results as in Theorems 1 and 2 of Section 2 have been obtained by Zografos (1991a). Also related with the stratified sampling setup is a recent paper by the same author (Zografos (1991b)).

2 Asymptotic Distribution of $I^{\varphi}(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y)$

Consider a sample of n members drawn at random with replacement from the population. We denote by $\hat{p}_{ij} = n_{ij}/n$, $\hat{p}_{i.} = n_{i.}/n$, $\hat{p}_{.j} = n_{.j}/n$ the sample estimators of p_{ij}, $p_{i.}$ and $p_{.j}$, respectively, where n_{ij} is the number of observations of

the value (x_i, x_j) $(i = 1,\ldots,M,\ j = 1,\ldots,K)$ in the sample, $n_{i.} = \sum_{j=1}^{K} n_{ij}$ and $n_{.j} = \sum_{i=1}^{M} n_{ij}$. The φ-divergence in the sample may be quantified as follows:

$$I^c(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y) = \sum_{i=1}^{M} \sum_{j=1}^{K} \hat{p}_{i.}\hat{p}_{.j}\varphi\left(\frac{\hat{p}_{ij}}{\hat{p}_{i.}\hat{p}_{.j}}\right) .$$

When the sample is drawn at random and with replacement from the population, the random vector $(n\hat{p}_{11},\ldots,n\hat{p}_{MK})$ has a multinomial distribution with parameters $(n; p_{11},\ldots,p_{MK})$.

The asymptotic distribution of $I^c(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y)$ in the previous random sampling set up is given in the following theorem:

Theorem 1: If we consider the analogue estimate $I^c(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y)$ obtained by replacing p_{ij}, $p_{i.}$ and $p_{.j}$ by the observed frequencies $\hat{p}_{ij}$, $\hat{p}_{i.}$ and $\hat{p}_{.j}$, then

$$n^{1/2}(I^c(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y) - I^c(P_{XY} \| P_X * P_Y)) \xrightarrow[n\uparrow\infty]{L} \mathcal{N}(0, v^2) ,$$

where

$$v^2 = \sum_{i=1}^{M} \sum_{j=1}^{K} k_{ij}^2 p_{ij} - \left(\sum_{i=1}^{M} \sum_{j=1}^{K} k_{ij} p_{ij}\right)^2$$

and

$$k_{ij} = \sum_{r=1}^{M} \left\{ p_{r.}\varphi\left(\frac{p_{rj}}{p_{r.}p_{.j}}\right) - \frac{p_{rj}}{p_{.j}}\varphi'\left(\frac{p_{rj}}{p_{r.}p_{.j}}\right)\right\}$$

$$+ \sum_{s=1}^{K} \left\{ p_{.s}\varphi\left(\frac{p_{is}}{p_{i.}p_{.s}}\right) - \frac{p_{is}}{p_{i.}}\varphi'\left(\frac{p_{is}}{p_{i.}p_{.s}}\right)\right\} + \varphi'\left(\frac{p_{ij}}{p_{i.}p_{.j}}\right)$$

provided $v^2 > 0$.

Proof: Bickel and Doksum (1977, pp. 135) have shown that if $\frac{\partial}{\partial x_i} h(x_1,\ldots,x_M)$ exist and is continuous for all $i = 1,\ldots,M$, then the asymptotic distribution of $T_n = h(\hat{p}_1,\ldots,\hat{p}_M)$ in a random sample is given by

$$n^{1/2}(T_n - h(p_1, \ldots, p_M)) \xrightarrow[n\uparrow\infty]{L} \mathcal{N}(0, \sigma^2) \ ,$$

where

$$\sigma^2 = \sum_{i=1}^{M} p_i \left(\frac{\partial h}{\partial p_i}(p_1, \ldots, p_M) \right)^2 - \left(\sum_{i=1}^{M} p_i \frac{\partial h}{\partial p_i}(p_1, \ldots, p_M) \right)^2 .$$

The derivatives are calculated treating $p_1, \ldots, p_M$ as "independent variables" not linked by $p_1 + p_2 + \cdots + p_M = 1$. If we consider the function

$$h(p_{11}, \ldots, p_{MK}) = I^c(P_{XY} \| P_X * P_Y) \ ,$$

then we obtain the stated result.

Remark 1:

(i) If we consider $\varphi(x) = x \log x$, $\varphi(x) = (1 - x)^2$, $\varphi(x) = (1 - x^{1/2})^2$, and $\varphi(x) = (\alpha - 1)^{-1}(x^\alpha - x)$, $\alpha > 0$, $\alpha \neq 1$, we obtain the following measures of divergence, between P_{XY} and $P_X * P_Y$: Kullback-Leibler, $I^{KL}(P_{XY} \| P_X * P_Y)$; Kagan, $I^{Ka}(P_{XY} \| P_X * P_Y)$; Matusita's square, $I^{Ma}(P_{XY} \| P_X * P_Y)$ and Havrda and Charvat, $I^{Ha}(P_{XY} \| P_X * P_Y)$; respectively. The corresponding expression of v^2 for the Kullback-Leibler's divergence is given by:

$$\sigma^2_{KL} = \sum_{i=1}^{M} \sum_{j=1}^{K} p_{ij} \log^2 \frac{p_{ij}}{p_{i.} p_{.j}} - \left(\sum_{i=1}^{M} \sum_{j=1}^{K} p_{ij} \log \frac{p_{ij}}{p_{i.} p_{.j}} \right)^2 .$$

In the remaining cases we will only write the expression of k_{ij}:

$$k_{ij}^{Ka} = 2 \frac{p_{ij}}{p_{i.} p_{.j}} - \sum_{r=1}^{M} \frac{p_{rj}^2}{p_{r.} p_{.j}} - \sum_{s=1}^{K} \frac{p_{is}^2}{p_{i.} p_{.s}}$$

$$k_{ij}^{Ma} = 3 - \sum_{r=1}^{M} \frac{p_{rj}^{1/2} p_{r.}^{1/2}}{p_{.j}^{1/2}} - \sum_{s=1}^{K} \frac{p_{.s}^{1/2} p_{is}^{1/2}}{p_{i.}^{1/2}} - \frac{p_{i.}^{1/2} p_{.j}^{1/2}}{p_{ij}^{1/2}}$$

$$k_{ij}^{Ha} = - \sum_{i=1}^{M} \frac{p_{rj}^{\alpha}}{p_{r.}^{\alpha-1} p_{.j}^{\alpha}} - \sum_{s=1}^{K} \frac{p_{is}^{\alpha}}{p_{.s}^{\alpha-1} p_{i.}^{\alpha}} + \frac{1}{1 - \alpha} \left(1 - \alpha \frac{p_{ij}^{\alpha}}{p_{i.}^{\alpha-1} p_{.j}^{\alpha-1}} \right)$$

(ii) In this context Zvarova (1973) obtained the asymptotic distribution of sample estimates of Renyi's divergence measure.

Theorem 2: If $P_{XY} = P_X{}^*P_Y$, then

$$\frac{2n(I^c(\hat{P}_{XY} \| \hat{P}_X{}^*\hat{P}_Y) - \varphi(1))}{\varphi''(1)} \xrightarrow[n\uparrow\infty]{L} \chi^2_{(M-1)(K-1)}$$

provided that $\varphi''(1) \neq 0$.

Proof: A Taylor series expansion of $\varphi(x)$ around the point 1 for

$$x = (\hat{p}_{ij})/(\hat{p}_{i.}\hat{p}_{.j})$$

yields

$$\varphi\left(\frac{\hat{p}_{ij}}{\hat{p}_{i.}\hat{p}_{.j}}\right) = \varphi(1) + \left(\frac{\hat{p}_{ij}}{\hat{p}_{i.}\hat{p}_{.j}} - 1\right)\varphi'(1) + \frac{1}{2}\left(\frac{\hat{p}_{ij}}{\hat{p}_{i.}\hat{p}_{.j}} - 1\right)^2(\varphi''(1) + \varepsilon_{ijn})$$

$i = 1, \ldots, M$, $j = 1, \ldots, K$, where $\varepsilon_{i,j,n} \xrightarrow[n\uparrow\infty]{P} 0$ as $n \to \infty$.

Multiplying both sides of the last expression by $n\hat{p}_{i.}\hat{p}_{.j}$ and suming over $i = 1, \ldots, M$, $j = 1, \ldots, K$, we get

$$nI^c(\hat{P}_{XY} \| \hat{P}_X{}^*\hat{P}_Y) = n\varphi(1) + \frac{n}{2}\varphi''(1)\sum_{i=1}^{M}\sum_{j=1}^{K}\frac{1}{\hat{p}_{i.}\hat{p}_{.j}}(\hat{p}_{ij} - \hat{p}_{i.}\hat{p}_{.j})^2$$

$$+ \frac{n}{2}\sum_{i=1}^{M}\sum_{j=1}^{K}\frac{(\hat{p}_{ij} - \hat{p}_{i.}\hat{p}_{.j})^2}{p_{i.}p_{.j}}\varepsilon_{ijn} .$$

Note that

$$n\sum_{i=1}^{M}\sum_{j=1}^{K}\frac{1}{\hat{p}_{i.}\hat{p}_{.j}}(\hat{p}_{ij} - \hat{p}_{i.}\hat{p}_{.j})^2 = \sum_{i=1}^{M}\sum_{j=1}^{K}\frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}} \xrightarrow[n\uparrow\infty]{L} \chi^2_{(M-1)(K-1)}$$

Finally, as $\varepsilon_{ijn} \xrightarrow[n\uparrow\infty]{P} 0$, we obtain the stated result.

The expressions of $\dfrac{2n(I^c(\hat{P}_{XY} \| \hat{P}_X{}^*\hat{P}_Y) - \varphi(1))}{\varphi''(1)}$, for the φ-divergences given in remark 1, can be seen in the following table.

Kullback-Leibler: $2n \sum_{i=1}^{M} \sum_{j=1}^{K} \hat{p}_{ij.} \log \frac{\hat{p}_{ij.}}{\hat{p}_{i..}\hat{p}_{.j.}}$

Kagan: $n \sum_{i=1}^{M} \sum_{j=1}^{K} \frac{(\hat{p}_{i..}\hat{p}_{.j.} - \hat{p}_{ij.})^2}{\hat{p}_{i..}\hat{p}_{.j.}}$

Matusita: $4n \sum_{i=1}^{M} \sum_{j=1}^{K} (\hat{p}_{i..}^{1/2}\hat{p}_{.j.}^{1/2} - \hat{p}_{ij.})^2$

Havrda-Charvat: $\alpha^{-1}(\alpha - 1)^{-1} 2n \left(\sum_{i=1}^{M} \sum_{j=1}^{K} \frac{\hat{p}_{ij.}^{\alpha}}{\hat{p}_{i..}^{\alpha-1}\hat{p}_{.j.}^{\alpha-1}} - 1 \right)$

Now we suppose that the population associated with the bivariate random variable (X, Y) is finite, N elements, and it can divided into L non-overlapping subpopulations, called strata as homogeneous as possible. Let N_l be the number of individuals into the lth stratum $\left(\text{so that } \sum_{l=1}^{L} N_l = N \right)$ and let p_{ijl} be the probability that a randomly selected member belongs to the lth stratum and takes on the value (x_i, y_j), $i = 1, \ldots, M$, $j = 1, \ldots, K$, $l = 1, \ldots, L$. Thus

$$\sum_{i=1}^{M} \sum_{j=1}^{K} p_{ijl} = \frac{N_l}{N} , \qquad \sum_{i=1}^{M} \sum_{j=1}^{K} \sum_{l=1}^{L} p_{ijl} = 1 .$$

Let $p_{ij.}$ be the probability that a randomly selected member in the whole population takes on the value (x_i, x_j), $p_{ij.} = \sum_{l=1}^{L} p_{ijl}$ $i = 1, \ldots, M$ and $j = 1, \ldots, K$, $p_{i.l} = \sum_{j=1}^{K} p_{ijl}$ be the marginal probability of the value x_i in the lth stratum and $p_{.jl} = \sum_{i=1}^{M} p_{ijl}$ be the marginal probability of the value y_j in the lth stratum. We will also define $p_{i..} = \sum_{l=1}^{L} p_{i.l}$ and $p_{.j.} = \sum_{l=1}^{L} p_{.jl}$. The φ-divergence in this context is given by

$${}_{st}I^{\varphi}(P_{XY} \| P_X * P_Y) = \sum_{i=1}^{M} \sum_{j=1}^{K} p_{.j.} p_{i..} \varphi \left(\frac{p_{ij.}}{p_{i..} p_{.j.}} \right) ,$$

where $P_{XY} = (p_{ij.};\ i = 1, \ldots, M,\ j = 1, \ldots, K)$, $P_X = (p_{i..};\ i = 1, \ldots, M)$ and $P_Y = (p_{.j.};\ j = 1, \ldots, K)$.

Consider a stratified sample of size n drawn at random from the population independently in different strata. We hereafter suppose that the sample is choosen by proportional allocation in each stratum. Assume also that a sample

of size n_l is drawn independently at random with replacement from the lth stratum where $n_l/n = N_l/N$. If $\hat{p}_{ijl}$ denotes the relative frequency of the values (x_i, y_j) into the lth stratum $\left(\text{and hence } \sum_{i=1}^{M}\sum_{j=1}^{K} \hat{p}_{ijl} = n_l/n,\ \hat{p}_{i.l} = \sum_{j=1}^{K} \hat{p}_{ijl},\ \hat{p}_{.jl} = \sum_{i=1}^{M} \hat{p}_{ijl},\ \hat{p}_{i..} = \sum_{l=1}^{L} \hat{p}_{i.l} \text{ and } \hat{p}_{.j.} = \sum_{l=1}^{L} \hat{p}_{.jl}\right)$, then the φ-divergence in the sample may be quantified by ${}_{st}I^{c}(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y)$, where $\hat{P}_{XY} = (\hat{p}_{ij.};\ i = 1, \ldots, M,\ j = 1, \ldots, K)$, $\hat{P}_X = (\hat{p}_{i..};\ i = 1, \ldots, M)$ and $\hat{P}_Y = (p_{.j.};\ j = 1, \ldots, K)$.

In this context we establish the following theorem.

Theorem 3:

$$n^{1/2}({}_{st}I^{c}(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y) - {}_{st}I^{c}(P_{XY} \| P_X * P_Y)) \xrightarrow[n \uparrow \infty]{L} \mathcal{N}(0, {}_{st}v^2) \ ,$$

where

$${}_{st}v^2 = \sum_{i=1}^{M}\sum_{j=1}^{K} b_{ij}^2 p_{ij.} - \sum_{l=1}^{L} \frac{N}{N_l}\left(\sum_{i=1}^{M}\sum_{j=1}^{K} b_{ij} p_{ijl}\right)^2$$

and

$$b_{ij} = \sum_{r=1}^{M}\left(p_{r..}\varphi\left(\frac{p_{rj.}}{p_{r..}p_{.j.}}\right) - \frac{p_{rj.}}{p_{.j.}}\varphi'\left(\frac{p_{rj.}}{p_{r..}p_{.j.}}\right)\right)$$

$$+ \sum_{s=1}^{K}\left(p_{.s.}\varphi\left(\frac{p_{is.}}{p_{i..}p_{.s.}}\right) - \frac{p_{is.}}{p_{i..}}\varphi'\left(\frac{p_{is.}}{p_{i..}p_{.s.}}\right)\right) + \varphi'\left(\frac{p_{ij.}}{p_{i..}p_{.j.}}\right)$$

Proof: If we consider Taylor series expansion of ${}_{st}I^{c}(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y)$ in a neighbourhood of P_{XY}, then we obtain that the random variables

$$n^{1/2}({}_{st}I^{c}(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y) - {}_{st}I^{c}(P_{XY} \| P_X * P_Y))$$

and

$$n^{1/2} \sum_{i=1}^{M}\sum_{j=1}^{K} b_{ij}(\hat{p}_{ij.} - p_{ij.})$$

have asymptotically the same p.d.f., where the partial derivatives

$$b_{ij} = \frac{\partial_{st} I^c(P_{XY} \| P_X * P_Y)}{\partial p_{ij.}} \; ; \qquad i = 1, \ldots, M \; , \quad j = 1, \ldots, K \; ,$$

are calculated treating the $p_{ij.}$'s as "independent" not linked by

$$\sum_{i=1}^{M} \sum_{j=1}^{K} p_{ij.} = 1 \; .$$

Furthermore, the random vectors

$$\left(\frac{n}{n_l} \hat{p}_{11l}, \ldots, \frac{n}{n_l} \hat{p}_{Mkl} \right) , \qquad l = 1, \ldots, L \; .$$

are independent and multinomially distributed with parameters

$$\left(n_l; \frac{N}{N_l} p_{11l}, \ldots, \frac{N}{N_l} p_{Mkl} \right) , \qquad l = 1, \ldots, L$$

Applying the MK-dimensional Central Limit Theorem, we obtain

$$n_l^{1/2} \left(\left(\frac{n}{n_l} \hat{p}_{11l} - \frac{N}{N_l} p_{11l} \right), \ldots, \left(\frac{n}{n_l} \hat{p}_{Mkl} - \frac{N}{N_l} p_{Mkl} \right) \right) \xrightarrow[n \uparrow \infty]{L} \mathcal{N}(0, \Sigma(l))$$

$$l = 1, \ldots, L$$

where

$$\Sigma(l) = \left(\frac{N}{N_l} p_{(i_1, j_1)l} \left(\delta_{(i_1, j_1)(i_2, j_2)} - \frac{N}{N_l} p_{(i_2, j_2)l} \right) \right)_{\substack{i_1, i_2 = 1, \ldots, M \\ j_1, j_2 = 1, \ldots, K}} ,$$

$\delta_{(i_1, j_1)(i_2, j_2)} = 1$ if $(i_1 = i_2, j_1 = j_2)$, $\delta_{(i_1, j_1)(i_2, j_2)} = 0$ otherwise and the index (i, j) is read in the following order: $(1, 1), (1, 2), \ldots, (1, K), (2, 1), (2, 2), \ldots, (2, K), \ldots, (M, 1), (M, 2), \ldots, (M, K)$.

As $n/n_l = N/N_l$ and $n_l^{1/2} = n^{1/2}(N_l/N)^{1/2}$, we have that

$$X_l^t = n^{1/2} \left(\frac{N}{N_l} \right)^{1/2} ((\hat{p}_{11l} - p_{11l}), \ldots, (\hat{p}_{Mkl} - p_{Mkl})) \xrightarrow[n \uparrow \infty]{L} \mathcal{N}(0, \Sigma(l)) \; .$$

$$l = 1, \ldots, L$$

As $X_1, X_2, \ldots, X_L$ are independent vectors, then

$$\sum_{l=1}^{L} \left(\frac{N_l}{N}\right)^{1/2} X_l^t \xrightarrow[n\uparrow\infty]{L} \mathcal{N}\left(0, \sum_{l=1}^{L} \frac{N_l}{N} \Sigma(l)\right) .$$

Therefore

$$\sum_{l=1}^{L} \left(\frac{N_l}{N}\right)^{1/2} b^t X_l^t = n^{1/2} \sum_{i=1}^{M} \sum_{j=1}^{K} b_{ij}(\hat{p}_{ij.} - p_{ij.})$$

has a normal asymptotic probability distribution function with mean zero and variance

$${}_{st}v^2 = \frac{1}{N} \sum_{l=1}^{L} N_l b^t \Sigma(l) b ,$$

where $b^t = (b_{ij};\ i = 1, \ldots, M,\ j = 1, \ldots, K)$.

Now we calculate ${}_{st}v^2$ explicitly.

$$b_{ij} = \frac{\partial_{st} I^c(P_{XY} \| P_X * P_Y)}{\partial p_{ij.}} = \sum_{r=1}^{M} \left(p_{r..} \varphi\left(\frac{p_{rj.}}{p_{r..} p_{.j.}}\right) - \frac{p_{rj.}}{p_{.j.}} \varphi'\left(\frac{p_{rj.}}{p_{r..} p_{.j.}}\right)\right)$$

$$+ \sum_{s=1}^{K} \left(p_{.s.} \varphi\left(\frac{p_{is.}}{p_{i..} p_{.s.}}\right) - \frac{p_{is.}}{p_{i..}} \varphi'\left(\frac{p_{is.}}{p_{i..} p_{.s}}\right)\right) + \varphi'\left(\frac{p_{ij.}}{p_{i..} p_{.j.}}\right) ,$$

Finally, we have

$${}_{st}v^2 = \sum_{i=1}^{M} \sum_{j=1}^{K} b_{ij}^2 p_{ij.} - \sum_{l=1}^{L} \frac{N}{N_l} \left(\sum_{i=1}^{M} \sum_{j=1}^{K} b_{ij} p_{ijl} \right)^2 .$$

Remark 2: Applying Jensen's inequality to the convex function φ, we obtain ${}_{st}v^2 \leq v^2$. Equality holds if and only if $L = 1$ or

$$\sum_{i=1}^{M} \sum_{j=1}^{K} \frac{N}{N_l} p_{ijl} b_{ij}$$

does not depend on l $(l = 1, \ldots, L)$.

Theorem 4: If $P_{XY} = P_X{}^*P_Y$ and $\varphi''(1) \neq 0$, then

$$\frac{2n[{}_{st}I^c(\hat{P}_{XY} \| \hat{P}_X{}^*\hat{P}_Y) - \varphi(1)]}{\varphi''(1)} \xrightarrow[n\uparrow\infty]{L} \sum_{i=1}^{MK} \beta_i \chi_1^2$$

where the χ_1's are independent, the β_i's are the eigenvalues of the matrix $A\Sigma^*$, being

$$A = (a_{(i_1,j_1)(i_2,j_2)})_{\substack{i_1,i_2=1,\ldots,M \\ j_1,j_2=1,\ldots,K}}, \qquad \Sigma^* = \sum_{l=1}^{L} \frac{N_l}{N} \Sigma(l),$$

$$a_{(i_1,j_1)(i_2,j_2)} = \begin{cases} \dfrac{1}{p_{i..}p_{.j.}} - \dfrac{1}{p_{i..}} - \dfrac{1}{p_{.j.}} & \text{if } i_1 = i_2 = i, \quad j_1 = j_2 = j \\ -\dfrac{1}{p_{i..}} & \text{if } i_1 = i_2 = i, \quad j_1 \neq j_2 \\ -\dfrac{1}{p_{.j.}} & \text{if } i_1 \neq i_2, \quad j_1 = j_2 = j \\ 0 & \text{if } i_1 \neq i_2, \quad j_1 \neq j_2 \end{cases}$$

$$\Sigma(l) = \left(\frac{N}{N_l} p_{(i_1,j_1)l} \left(\delta_{(i_1,j_1)(i_2,j_2)} - \frac{N}{N_l} p_{(i_2,j_2)l} \right) \right)_{\substack{i_1,i_2=1,\ldots,M \\ j_1,j_2=1,\ldots,K}}$$

and the index (i, j) is read in the following order: $(1, 1), (1, 2), \ldots, (1, K), (2, 1), (2, 2), \ldots, (2, K), \ldots, (M, 1), (M, 2), \ldots, (M, K)$.

Proof: If $P_{XY} = P_X{}^*P_Y$, then

$$n^{1/2} \sum_{i=1}^{M} \sum_{j=1}^{K} b_{ij}(\hat{p}_{ij.} - p_{ij.}) = 0$$

and

$${}_{st}I^c(P_{XY} \| P_X{}^*P_Y) = \varphi(1) .$$

So, we must use Taylor series expansion of ${}_{st}I^c(\hat{P}_{XY} \| \hat{P}_X{}^*\hat{P}_Y)$ in a neighbourhood of $P_X{}^*P_Y$ including the second order partial derivatives term. After some

straightforward calculus, it can be proved the following equality between matrices:

$$\left(\frac{\partial^2{}_{st} I^c(P_{XY} \| P_X{}^* P_Y)}{\partial p_{i_1 j_1 \cdot} \partial p_{i_2 j_2 \cdot}}\right)_{\substack{i_1, i_2 = 1, \ldots, M \\ j_1, j_2 = 1, \ldots, K}} = \varphi''(1) A + B \ ,$$

where

$$A = (a_{(i_1, j_1)(i_2, j_2)})_{\substack{i_1, i_2 = 1, \ldots, M \\ j_1, j_2 = 1, \ldots, K}} \ , \qquad B = (b_{(i_1, j_1)(i_2, j_2)})_{\substack{i_1, i_2 = 1, \ldots, M \\ j_1, j_2 = 1, \ldots, K}} \ ,$$

$$b_{(i_1, j_1)(i_2, j_2)} = 2[\varphi''(1) + \varphi(1) - \varphi'(1)] \ ,$$

$$a_{(i_1, j_1)(i_2, j_2)} = \begin{cases} \dfrac{1}{p_{i\cdot\cdot} p_{\cdot j \cdot}} - \dfrac{1}{p_{i\cdot\cdot}} - \dfrac{1}{p_{\cdot j \cdot}} & \text{if} \quad i_1 = i_2 = i, \quad j_1 = j_2 = j \\ -\dfrac{1}{p_{i\cdot\cdot}} & \text{if} \quad i_1 = i_2 = i, \quad j_1 \neq j_2 \\ -\dfrac{1}{p_{\cdot j \cdot}} & \text{if} \quad i_1 \neq i_2, \quad j_1 = j_2 = j \\ 0 & \text{if} \quad i_1 \neq i_2, \quad j_1 \neq j_2 \end{cases}$$

and the index (i, j) is read in the following order: (1, 1), (1, 2), ..., (1, K), (2, 1), (2, 2), ..., (2, K), ..., (M, 1), (M, 2), ..., (M, K).

As $(\hat{P}_{XY} - P_X{}^* P_Y)^t B (\hat{P}_{XY} - P_X{}^* P_Y) = 0$, then the random variables

$$\frac{2n[{}_{st} I^c(\hat{P}_{XY} \| \hat{P}_X{}^* \hat{P}_Y) - \varphi(1)]}{\varphi''(1)} \qquad \text{and} \qquad (\hat{P}_{XY} - P_X{}^* P_Y)^t A (\hat{P}_{XY} - P_X{}^* P_Y)$$

converge in law to the same distribution. We know that if $P_{XY} = P_X{}^* P_Y$, then $(\hat{P}_{XY} - P_X{}^* P_Y)$ is normally distributed with parameters $\mu = 0$ and $\Sigma^* = \sum_{l=1}^{L} \frac{N_l}{N} \Sigma(l)$, where

$$\Sigma(l) = \left(\frac{N}{N_l} p_{(i_1, j_1) l} \left(\delta_{(i_1, j_1)(i_2, j_2)} - \frac{N}{N_l} p_{(i_2, j_2) l}\right)\right)_{\substack{i_1, i_2 = 1, \ldots, M \\ j_1, j_2 = 1, \ldots, K}} \ .$$

Therefore, see Mardia et al, 1982, pp. 68, we have the stated result.

3 Tests of Independence

Let (X, Y) be as in Section 1. n observations are drawn at random and with replacement from the population. The hypotheses to be tested is $H_0 : p_{ij} = p_{i.}p_{.j}$ $i = 1, 2, \ldots, M$, $j = 1, 2, \ldots, K$,

We consider the statistics given in theorem 2

$$T_1 = \frac{2n(I^c(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y) - \varphi(1))}{\varphi''(1)}$$

If H_0 is true, then T_1 will be small. Thus a large value of T_1 indicates data less compatible with the null hypothesis. Hence for large n a level α test is given by

$$\Phi(\hat{p}_{11}, \ldots, \hat{p}_{MK}) = \begin{cases} 1 & \text{if } T_1 > \chi^2_{(M-1)(K-1),\alpha} \\ 0 & \text{otherwise} \end{cases}$$

Theorem 1 can be used to evaluate the asymptotic power of the previous test when p_{ij}'s are known but not equal to the $p_{i.}p_{.j}$'s

$$\beta(P_{XY}, P_X * P_Y) = P_{XY}\left(\frac{2n(I^c(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y) - \varphi(1))}{\varphi''(1)} > \chi^2_{(M-1)(K-1),\alpha}\right)$$

$$= 1 - \Psi\left(\frac{\varphi''(1)\chi^2_{(M-1)(K-1),\alpha} - 2n[I^c(P_{XY} \| P_X * P_Y) - \varphi(1)]}{2n^{1/2} v(P_{XY} \| P_X * P_Y)}\right),$$

where $v^2(P_{XY} \| P_X * P_Y)$ is the expression of v^2 given in theorem 1 and Ψ denotes the mean zero and variance one normal c.d.f.

On the other hand, if we have n observations on (X, Y) obtained by a stratified random sampling with proportional allocation, then we consider the statistic given in theorem 4.

$$T_2 = \frac{2n({}_{st}I^c(\hat{P}_{XY} \| \hat{P}_X * \hat{P}_Y) - \varphi(1))}{\varphi''(1)}$$

If H_0 is true, then T_2 will be small. Hence for large n, when $T_2 = t$ one must reject H_0 at a level α if

$$P\left(\sum_{h=1}^{MK} \beta_h \chi_1^2 > t\right) \leq \alpha$$

where the β_h's, $h = 1, \ldots, MK$ are given in theorem 4.

As in simple random sample set up, theorem 3 can be used to evaluate the asymptotic power of the previous test when $(p_{11.}, \ldots, p_{MK.})$ is not equal to $(\hat{p}_{1.}.p_{.1.}, \ldots, \hat{p}_{M.}.p_{.K.})$.

Acknowledgements: The authors thanks the referees for their valuable suggestions.

References

Bickel PJ and Doksum KA (1977) *Mathematical Statistics.* Holden-Day, Inc. Oakland, California

Csiszár I (1967) Information-type measure of difference of probability distributions and indirect observations. Studia Scientiarum Mathematicarum Hungarica 2:299–318

Mardia KV, Kent JT and Bibby JM (1982) Multivariate Analysis. Academic Press

Menéndez ML, Morales D, Pardo L and Salicrú M (1992) Some Applications of (r, s)-Directed divergences. To appear in Utilitas Mathematica

Morales D, Pardo L, Salicrú M and Menéndez ML (1992) The λ-divergence and the λ-mutual Information: Estimation in the Stratified Sampling. To appear in Journal of Computational and Applied Mathematics.

Salicrú M, Menéndez ML, Morales D and Pardo L (1992) Asymptotic properties of (r, s)-Directed Divergence in a Stratified Sampling. To appear in Applied Mathematics and Computation

Zografos K, Ferentinos K and Papaioannou T (1990) φ-Divergence Statistics: Sampling Properties and Multinomial Goodness of fit and Divergence Tests. *Commun. Statist.* (*Theory Meth.*), 19(5): 1785–1802

Zografos K (1991a) Asymptotic properties of φ-divergence statistic and its application in contingency tables. Tech. Report #185, Math. Dept., Univ. of Ioannina, to appear in the Int'l J. of Mathematical and Statistical Sciences

Zografos K (1991b) Asymptotic distributions of estimated f-dissimilarity between populations in a stratified random sampling. Tech. Report #191, Math. Dept. Univ. of Ioannina

Zvarova J (1973) On asymptotic behaviour of a sample estimator of Renyi's information of order α. Trans. 6th Prague Conf. on Inf. Theory Stat. Funct. and Proc., Prague, Czech. Acad. of Sci. 914–924

Received 26.07.1992
Revised version 06.10.1992

Book Review

Ch. E. Minder/G. U. H. Seeber (Eds.): Multivariate Modelle; Medizinische Informatik, Biometrie und Epidemiologie 74, Springer-Verlag, 1991, 165 pp., DM 54,—

In this book written in German and edited by G. U. H. Seeber and Ch. E. Minder we find several papers presented at a seminar of the Swiss-Austria Region of the International Biometric Society held in September 1991 in Biel.

M. Berres. Non-Linear Multivariate Analysis of a Field Experiment.

R. Hatzinger. Quasi-Likelihood Method of Analysis of Independent and Dependent Observations.

G. Tutz. Kernel Functions in Density Estimation Problems, Nonparametric Regression and Discrimination Analysis.

S. Frühwirth-Schnatter. Monitoring of Ecological and Biometric Processes with Statistical Filtering.

W.-J. Stronegger. Kalman Filter for On-Line-Discriminant-Analysis of Growth Curves.

Ch. E. Minder. Global Goodness of Fit Tests for a Wide Class of Statistical Models.

The individual articles demonstrate examples of biometric research that use more complex methods than the classical multivariate techniques.

Bern H. Riedwyl

Metrika (1993) 40:237–242

Metrika

A Characterization of the Binomial and Negative Binomial Distribution by Regression of Products

JERZY PUSZ

Institute of Mathematics, University of Technology, Pl. Politechniki 1, 00-661 Warszawa, Poland

Summary: Products of independent random variables with different shifts are considered. The binomial and negative binomial populations are characterized by proportionality of regressions of these products with respect to residuals.

1 Introduction

Characterizations of probability distributions by properties of regression of some functions in independent random variables with respect to residuals have begun with the celebrated Kagan-Linnik-Rao theorem – see Kagan et al. (1965). The most recent results are gathered in Kagan and Rao (1988), Wesolowski (1989) and Kagan (1991). In the second paper a characterization of the Poisson law by a constant regression of product was obtained.

The main results of the present paper is somewhat similar since we also deal with regressions of products. However we are not interested in the exact form of these conditional moments. Instead we relate regression of products of independent and identically distributed random variables with different shifts. A characterization of the binomial and negative binomial distributions follows.

2 Characterization

It is well known that

$$f(t) = (pe^{ir_1t} + qe^{ir_2t})^v, \qquad t \in \mathbb{R}, \qquad p + q = 1, \qquad r_1 \neq r_2 \ , \tag{1}$$

is the general form of a characteristic function of the binomial and negative binomial distributions. The distribution is binomial iff $pq > 0$ and v is a positive integer; it is negative binomial iff $pq < 0$ and $v < 0$.

0026–1335/93/3–4/237–242 $2.50

Let X be a random variable with a distribution function F and the expectation m. Consider two samples $(Y_1, \ldots, Y_n)$ and $(Z_1, \ldots, Z_n)$, $n \geq 3$ from shifted population $F(x - \theta_1)$ and $F(x - \theta_2)$, $\theta_1 \neq \theta_2$, respectively. Let

$$\overline{Y} = \frac{1}{n} \sum_{k=1}^{n} Y_k, \qquad \overline{Z} = \frac{1}{n} \sum_{k=1}^{n} Z_k \,. \tag{2}$$

Theorem: The relation

$$\begin{aligned} &(m + \theta_2)^n E\left(\prod_{k=1}^{n} Y_k | Y_1 - \overline{Y}, \ldots, Y_n - \overline{Y} \right) \\ &\quad = (m + \theta_1)^n E\left(\prod_{k=1}^{n} Z_k | Z_1 - \overline{Z}, \ldots, Z_n - \overline{Z} \right) \end{aligned} \tag{3}$$

holds almost everywhere iff

1° X has the binomial distribution and $(m + \theta_1)(m + \theta_2) < 0$,
2° X has the negative binomial distribution and $(m + \theta_1)(m + \theta_2) > 0$,
3° $X = m$ a.s. and $(m + \theta_1)(m + \theta_2) = 0$.

Proof: To proof the necessity we assume that (3) is satisfied. We first observe that (3) is equivalent to

$$\begin{aligned} &(m + \theta_2)^n E\left(\prod_{k=1}^{n} Y_k | Y_2 - Y_1, \ldots, Y_n - Y_1 \right) \\ &\quad = (m + \theta_1)^n E\left(\prod_{k=1}^{n} Z_k | Z_2 - Z_1, \ldots, Z_n - Z_1 \right) . \end{aligned} \tag{4}$$

Hence (see, for example, Lemma 1.1.1 in Kagan et al. (1973))

$$\begin{aligned} &(m + \theta_2)^n E\left(\prod_{k=1}^{n} Y_k \exp\left(i \sum_{k=2}^{n} t_k Y_k - i Y_1 \sum_{k=2}^{n} t_k \right) \right) \\ &\quad = (m + \theta_1)^n E\left(\prod_{k=1}^{n} Z_k \exp\left(i \sum_{k=2}^{n} t_k Z_k - i Z_1 \sum_{k=2}^{n} t_k \right) \right) \end{aligned} \tag{5}$$

for $t_k \in \mathbb{R}$, $k = 1, \ldots, n$.

Let $\varphi(t) = E\exp(itX)$ be the characteristic function of X. From (5)

$$(m+\theta_2)^n\left(i\theta_1\varphi\left(-\sum_{k=2}^{n}t_k\right)+\varphi'\left(-\sum_{k=2}^{n}t_k\right)\right)\prod_{k=2}^{n}(i\theta_1\varphi(t_k)+\varphi'(t_k))$$
$$=(m+\theta_1)^n\left(i\theta_2\varphi\left(-\sum_{k=2}^{n}t_k\right)+\varphi'\left(-\sum_{k=2}^{n}t_k\right)\right)\prod_{k=2}^{n}(i\theta_2\varphi(t_k)+\varphi'(t_k))\ . \tag{6}$$

Consequently for $m+\theta_j=0$, $j=1$ or 2, we have $im\varphi(t)=\varphi'(t)$ in a neighborhood of the origin. Hence $\varphi(t)=\exp(imt)$.

Assume now that $(m+\theta_1)(m+\theta_2)\neq 0$. Hence in a neighborhood V of the origin $i\theta_j\varphi+\varphi'$ is non-zero, $j=1,2$ so we can introduce a function

$$H(t)=\ln\frac{(m+\theta_1)(i\theta_2\varphi(t)+\varphi'(t))}{(m+\theta_2)(i\theta_1\varphi(t)+\varphi'(t))}\ , \qquad t\in\mathbb{R} \tag{7}$$

(we take the principal branch of the logarithm).

Then by (6)

$$H\left(-\sum_{k=2}^{n}t_k\right)+\sum_{k=2}^{n}H(t_k)=0\ . \tag{8}$$

As the solution of the equation (8) we obtain (see, for example, Lemma 1.5.1 in Kagan et al. (1973))

$$H(t)=at+d\ , \qquad a,d\in\mathbb{C}\ . \tag{9}$$

Since $H(0)=0$ then $d=0$.

From (7) we have

$$(m+\theta_1)(i\theta_2\varphi(t)+\varphi'(t))=e^{at}(m+\theta_2)(i\theta_1\varphi(t)+\varphi'(t))\ . \tag{10}$$

Hence

$$\left(\frac{m+\theta_2}{m+\theta_1}e^{at}-1\right)\varphi'(t)=i\left(-\frac{m+\theta_2}{m+\theta_1}e^{at}\theta_1+\theta_2\right)\varphi(t)\ . \tag{11}$$

Consequently φ is infinitely often differentiable. Taking $t=0$ in (11) before and after differentiation we get

$$a = -ib\ , \qquad b = \frac{(\theta_2 - \theta_1)\sigma^2}{(m + \theta_1)(m + \theta_2)}\ , \tag{12}$$

where $\sigma^2 = Var(X)$.

We solve the differential equation (11) and obtain (1) in V with

$$p = \frac{m + \theta_2}{\theta_2 - \theta_1}\ , \qquad q = -\frac{m + \theta_1}{\theta_2 - \theta_1}\ , \qquad v = -\frac{(m + \theta_1)(m + \theta_2)}{\sigma^2}\ ,$$
$$r_1 = \frac{\theta_1\sigma^2}{(m + \theta_1)(m + \theta_2)}\ , \qquad r_2 = \frac{\theta_2\sigma^2}{(m + \theta_1)(m + \theta_2)}\ . \tag{13}$$

Consequently, if $(m + \theta_1)(m + \theta_2) < 0$ then $pq > 0$, $v > 0$ and X has a binomial distribution and if $(m + \theta_1)(m + \theta_2) > 0$ then $pq < 0$, $v < 0$ and X has a negative binomial distribution, since both the distributions are uniquely determined by moments.

Now let us assume that X is a binomial random variable with the characteristic function (1). Then

$$H(t) = \ln \frac{[v(pr_1 + qr_2) + \theta_1][p(\theta_2 + r_1 v)e^{ir_1 t} + q(\theta_2 + r_2 v)e^{ir_2 t}]}{[v(pr_1 + qr_2) + \theta_2][p(\theta_1 + r_1 v)e^{ir_1 t} + q(\theta_1 + r_2 v)e^{ir_2 t}]}\ . \tag{14}$$

Hence the equation (8) is fulfilled for any real t_k, $k = 2, \ldots, n$ and

$$\theta_1 = -r_1 v\ , \qquad \theta_2 = -r_2 v\ . \tag{15}$$

As the consequence we obtain (3). ■

Remark 1: For $n = 2$ the formula (8) takes the form $H(t) = -H(-t)$. And the problem in this case remains open.

Remark 2: Putting $t_2 = \cdots = t_n = -\frac{t}{n}$ in (5) we find that it is equivalent to

$$(m + \theta_2)^n E\left(\prod_{k=1}^{n} Y_k | Y_1 - \overline{Y}\right) = (m + \theta_1)^n E\left(\prod_{k=1}^{n} Z_k | Z_1 - \overline{Z}\right)\ . \tag{16}$$

Hence, the assumptions in the Theorem can be weakened replacing there (3) by (16).

Similarly as in Wesolowski (1989) we can omit the equidistribution assumption of the Theorem. Also the shifts may differ.

Let X_k be a random variable with a distribution function F_k and an expectation m_k, $k = 1, \ldots, n$, $n \geq 3$; $X_1, \ldots, X_n$ are independent. Consider random variables $Y_k = X_k + \theta_{1k}$, $Z_k = X_k + \theta_{2k}$, $\theta_{1k}, \theta_{2k} \in \mathbb{R}$, $\theta_{1k} \neq \theta_{2k}$, $k = 1, \ldots, n$. Additionally assume that $\prod_{k=1}^{n} (m_k + \theta_{1k}) \neq 0 \neq \prod_{k=1}^{n} (m_k + \theta_{2k})$.

Proposition: The relation

$$\prod_{k=1}^{n} (m_k + \theta_{2k}) E\left(\prod_{k=1}^{n} Y_k | Y_1 - \bar{Y}, \ldots, Y_n - \bar{Y}\right) = \prod_{k=1}^{n} (m_k + \theta_{1k}) E\left(\prod_{k=1}^{n} Z_k | Z_1 - \bar{Z}, \ldots, Z_n - \bar{Z}\right) \tag{17}$$

holds almost everywhere iff $X_1, \ldots, X_n$ are binomial or negative binomial random variables.

Proof: Let $\varphi_k(t) = E \exp(itX_k)$ be the characteristic function of X_k and

$$H_k(t) = \ln \frac{(m_k + \theta_{2k})(i\theta_{1k}\varphi_k(t) + \varphi_k'(t)}{(m_k + \theta_{1k})(i\theta_{2k}\varphi_k(t) + \varphi_k'(t)}, \qquad k = 1, \ldots, n . \tag{18}$$

Then from (17)

$$H_1\left(-\sum_{k=2}^{n} t_k\right) + \sum_{k=2}^{n} H_k(t_k) = 0 , \qquad t_k \in \mathbb{R} , \qquad k = 2, \ldots, n , \tag{19}$$

and as consequence H_k, $k = 1, \ldots, n$ are linear functions (see, for example, Lemma 1.5.1 in Kagan et al. (1973)). Solving the differential equation (18) we obtain the final conclusion. ■

Acknowledgment: The author is greatly indebted to the referee for the suggestions, which helped to strengthen the main result.

References

Kagan AM, Linnik Yu V, Rao CR (1965) On a characterization of the normal law based on a property of the sample average. *Sankhya*, Ser. A. 27:405–406

Kagan AM, Linnik Yu V, Rao CR (1973) Characterization Problems in Mathematical Statistics. Wiley, New York

Kagan AM, Rao CR (1988) Constancy of regression of a polynomial of sample average on residuals characterizes normal distribution. Technical Report No. 88-30, Center for Multivariate Analysis, Univ. of Pittsburgh (to be published in Proc. Raj Bose Symposium on Probability and Statistics)

Kagan AM (1991) A method for solving a class of functional equations arising in characterization problems. Proceedings, 1990 Taipei Symposium in Statistics, Inst. Statist. Sci., Acad. Sinica, 321–336

Wesolowski J (1989) A regressional characterization of the Poisson distribution. In: Stability Problems for Stochastic Models (Proceedings, Sukhumi, 1987), Springer, *Lecture Notes in Mathematics* 1412:349–351

Received 27.04.1992
Revised version 06.10.1992

Metrika (1993) 40:243–256

Metrika

Bootstrap Based Goodness-Of-Fit-Tests

WINFRIED STUTE

Mathematisches Institut, Universität Gießen, Arndtstraße 2, 6300 Gießen, Germany

WENCESLAO GONZÁLES MANTEIGA

Facultad de Matemáticas, Universidad de Santiago, 15706 Santiago de Compostela, Spain

MANUEL PRESEDO QUINDIMIL

Facultad de Veterinaria, Universidad de Santiago, 27002 Lugo, Spain

Summary: Let $\mathcal{F} = \{F_\theta\}$ be a parametric family of distribution functions, and denote with F_n the empirical d.f. of an i.i.d. sample. Goodness-of-fit tests of a composite hypothesis (contained in $\mathcal{F}$) are usually based on the so-called estimated empirical process. Typically, they are not distribution-free. In such a situation the bootstrap offers a useful alternative. It is the purpose of this paper to show that this approximation holds with probability one. A simulation study is included which demonstrates the validity of the bootstrap for several selected parametric families.

1 Introduction and Main Result

Assume that $\mathcal{F} = \{F(\cdot;\theta): \theta \in \Theta\}$ is a parametric family of distribution functions (d.f.) on m-dimensional Euclidean space. We observe a sample $X_1, \ldots, X_n$ of i.i.d. random vectors in $\mathbb{R}^m$ with common unknown d.f. F. An important issue then is if the underlying F belongs to $\mathcal{F}$ or not. More precisely, we may write $\theta = (\theta_1, \theta_2)$, where θ_1 is a p_1-dimensional parameter of interest and θ_2 is a p_2-dimensional nuisance parameter. In such a situation one would like to test, e.g., the hypothesis H_0: $\theta_1 = \bar{\theta}_1$. When $p_2 = 0$, this reduces to the case of a simple hypothesis, while for $p_1 = 0$, H_0 becomes equivalent to the aforementioned full model check $F \in \mathcal{F}$. A general recipe to test H_0 is to

(a) compute the empirical d.f. F_n of $X_1, \ldots, X_n$
(b) estimate θ_2 by some $\hat{\theta}_{2n}$

and then to compare

(c) F_n and $F(\cdot; \bar{\theta}_1, \hat{\theta}_{2n})$.

Set $\hat{\theta}_n = (\bar{\theta}_1, \hat{\theta}_{2n})$. Goodness-of-fit procedures are often based on proper discrepancies of the so-called estimated empirical process

$$\hat{\alpha}_n(x) = n^{1/2}[F_n(x) - F(x; \hat{\theta}_n)] , \qquad x \in \mathbb{R}^m .$$

0026–1335/93/3–4/243–256 $2.50

E.g., we may consider the Kolmogorov-Smirnov distance

$$D_n = n^{1/2} \sup_x |F_n(x) - F(x; \hat{\theta}_n)| = \|\hat{\alpha}_n\|_\infty$$

or the Cramér-von-Mises distance

$$W_n^2 = n \int [F_n(x) - F(x; \hat{\theta}_n)]^2 F(dx; \hat{\theta}_n) = \int \hat{\alpha}_n^2(x) F(dx; \hat{\theta}_n) \ .$$

Unfortunately, estimation of θ_2 may lead to test statistics which are no longer distribution-free. Consequently, computation of critical values has to be carried through for each particular $\mathscr{F}$. Moreover, the distribution of the test statistic may depend on the unknown parameter. D'Agostino and Stephens (1986) provide a comprehensive account of available results for some selected parametric models. See also Stephens (1974), (1976). Durbin and Knott (1972) and Durbin, Knott and Taylor (1975) provide an orthogonal series decomposition of $\hat{\alpha}_n$ yielding a representation of W_n^2 as a weighted series of independent χ^2-variables. It is the purpose of this paper to show that the bootstrap offers an alternative universal approach for the computation of critical values, under broad assumptions on $\mathscr{F}$ and $\hat{\theta}_n$. The main result is stated as *Theorem 1.1*: It provides a bootstrap version of Durbin's (1973) invariance principle for $\hat{\alpha}_n$. The proof is postponed to the Appendix. Section 2 contains several simulation results which, for the sake of comparison, show that the bootstrap offers an excellent alternative approximation in situations when the true quantiles are tabled. In case no tables exist it constitutes the only possibility. In the following we state the assumptions required for Theorem 1.1 below. (A1) is concerned with regularity of $\hat{\theta}_{2n}$.

(A1): Under H_0, i.e. $F = F(\cdot; \theta)$ with $\theta = (\bar{\theta}_1, \theta_2)$

$$n^{1/2}(\hat{\theta}_{2n} - \theta_2) = n^{-1/2} \sum_{i=1}^{n} l(X_i, \theta) + \varepsilon_{1n} \ , \tag{i}$$

where

$$\mathbb{E}_\theta l(X, \theta) = 0, \qquad \mathbb{E}_\theta [l(X, \theta) l^t(X, \theta)] \text{ exists}$$

and $\varepsilon_{1n} \to 0$ in probability.

Moreover, as $\theta_n \to \theta$, for each $x_0 \in \mathbb{R}^m$

$$\int_{(-\infty, x_0]} l^t(x, \theta_n) F(dx; \theta_n) \to \int_{(-\infty, x_0]} l^t(x, \theta) F(dx; \theta) \tag{ii}$$

and

$$d_2[l(X^n, \theta_n), l(X, \theta)] \to 0 \qquad \text{as } n \to \infty \ . \tag{iii}$$

Here $X^n \sim F(\cdot; \theta_n)$ and d_2 denotes Mallows' metric.

Assumption (A2) guarantees that $F(x; \theta)$ is smooth as a function of x and θ.

(A2): For any open neighborhood U of θ

$$F(\cdot; \underset{\sim}{\theta}) \ , \qquad \underset{\sim}{\theta} \in U \ , \qquad \text{is equicontinuous}$$

and

$$g(x, \bar{\theta}_1, \theta_2) = \partial F(x; \bar{\theta}_1, \theta_2)/\partial\theta_2 \text{ is uniformly continuous on } \mathbb{R}^m \times U \ .$$

In a bootstrap situation we shall need (i) with $\hat{\theta}_n$ in place of θ. "$\varepsilon_{1n} \to 0$ in probability" then means that for each $\delta > 0$

$$P_{\hat{\theta}_n}\{|\varepsilon_{1n}| \geq \delta\} \to 0 \qquad \text{as } n \to \infty \ .$$

Though it might look artificial at first sight, (i) just reminds us of how asymptotic normality is achieved in proofs, namely via (i) and a Cramér-Slutzky argument. Cf. Serfling (1980), pp. 144, for the maximum likelihood estimator. Conditions (ii) and (iii) yield distributional smoothness of the leading term in (i). Introducing Mallows' metric only describes, in a concise way, the fact that $l(X^n, \theta_n) \to l(X, \theta)$ weakly together with convergence of their second moments; cf. Bickel and Freedman (1981). Durbin (1973) formulated his main result for an $\hat{\alpha}_n$-process properly transformed so as to become an element of the Skorokhod space $D[0, 1]$. We prefer to state our result for the original process, which is in $D[-\infty, \infty]^m$; cf. Pollard (1984). In such a situation, under (A1) and (A2),

$$\hat{\alpha}_n \to Z \text{ in distribution} \ ,$$

where $Z = \{Z(x)\}$ is a zero means Gaussian process with continuous sample paths and covariance function

$$\begin{aligned} \operatorname{Cov}[Z(x_1), Z(x_2)] &= F(x_1 \wedge x_2; \theta) - F(x_1; \theta)F(x_2; \theta) \\ &- \int_{(-\underline{\infty}, x_2]} l^t(x, \theta) g(x_1, \theta) F(dx; \theta) - \int_{(-\underline{\infty}, x_1]} l^t(x, \theta) g(x_2, \theta) F(dx; \theta) \\ &+ \int g^t(x_1, \theta) l(x, \theta) l^t(x, \theta) g(x_2, \theta) F(dx; \theta) \ . \end{aligned}$$

To adequately define the bootstrap version of $\hat{\alpha}_n$, take an i.i.d. random sample $X_1^*, \ldots, X_n^*$ from $F(\cdot; \hat{\theta}_n)$. Let $\hat{\theta}_{2n}^*$ denote the bootstrap value of $\hat{\theta}_{2n}$. The bootstrap version of $\hat{\alpha}_n$ is then given by

$$\alpha_n^*(x) = n^{1/2}[F_n^*(x) - F(x; \bar{\theta}_1, \hat{\theta}_{2n}^*)] \ , \qquad x \in \mathbb{R}^m \ .$$

We are now in the position to state

1.1 Theorem: Under H_0, assume that (A1) and (A2) are satisfied. Then, if $\hat{\theta}_{2n} \to \theta_2$ w.p.1,

$$\alpha_n^* \to Z \qquad \text{in distribution}$$

with probability one.

As a consequence, for each Z-continuity set A

$$\mathbb{P}_n^*(\alpha_n^* \in A) - \mathbb{P}(\hat{\alpha}_n \in A) \to 0 \ .$$

Here, $\mathbb{P}_n^*$ is a probability measure carrying $X_1^*, \ldots, X_n^*$. In particular, since Z has continuous sample paths, the sup is continuous Z-almost everywhere. The continuous mapping theorem thus yields

$$\mathbb{P}_n^*(\|\alpha_n^*\|_\infty \leq x) - \mathbb{P}(\|\hat{\alpha}_n\|_\infty \leq x) \to 0 \qquad \text{w.p.1}$$

for each x at which $\mathbb{P}(\|Z\| \leq x)$ is continuous. In other words, under the stated regularity assumptions, the distribution of D_n allows for a bootstrap approximation. Similarly for W_n^2.

2 Simulation Study

In this section we shall report on various simulation results which will demonstrate the validity of the bootstrap approximation in several selected parametric families. As test statistics we considered D_n and W_n^2. The parametric families to be discussed in greater detail include

(a) The normal distributions (univariate and bivariate)
(b) The exponential distributions

(c) The logistic distributions
(d) The Weibull distributions

M random samples, each of sample size n, were generated and analyzed according to the following algorithm:

1. Generate a random sample $X_1, \ldots, X_n$ from some $F(\cdot; \theta)$
2. Calculate the estimator $\hat{\theta}_{2n}$ of θ_2
3. Evaluate $Z_i = F(X_i; \bar{\theta}_1, \hat{\theta}_{2n})$
4. Rearrange the Z's in ascending order: $Z_{1:n} < \cdots < Z_{n:n}$
5. For real-valued data ($m = 1$) calculate D_n and W_n according to

$$D_n = n^{1/2} \max_{1 \le i \le n} \left\{ \frac{i}{n} - Z_{i:n}, Z_{i:n} - \frac{(i-1)}{n} \right\}$$

$$W_n^2 = \sum_{i=1}^{n} \left[Z_{i:n} - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}$$

6. Draw B bootstrap samples as described below:

(a) Generate random samples $X_1^*(j), \ldots, X_n^*(j)$ from $F(\cdot; \bar{\theta}_1, \hat{\theta}_{2n})$, $1 \le j \le B$
(b) Calculate the bootstrap estimator $\hat{\theta}_{2n}^*(j)$ of θ_2, $1 \le j \le B$
(c) Evaluate $Z_i^*(j) = F(X_i^*(j); \bar{\theta}_1, \theta_{2n}^*(j))$, $1 \le j \le B$, $1 \le i \le n$
(d) For each $1 \le j \le B$, put the $Z_i^*(j)$ in ascending order
(e) Compute the $D_n^*(j)$ and $W_n^*(j)$ as above, with the $Z^*(j)$'s in place of the Z's

7. Suppose that for a particular $\mathcal{F}$ of interest, a given $n \ge 1$ and a significance level α, the exact $1 - \alpha$ quantile $x_{1-\alpha}$ of D_n respectively W_n^2 is available. Then reject H_0 if the test statistic exceeds $x_{1-\alpha}$.
8. If 7. does not apply, or just for comparison, the bootstrap based methodology suggests rejecting H_0 if, e.g., D_n exceeds $D^*_{(B(1-\alpha)+1)}$, the $B(1-\alpha) + 1 -$ largest order statistic among $D_n^*(1), \ldots, D_n^*(B)$. Similarly for the W^*'s.
9. Obtain the estimators $\hat{\alpha}_D$, $\hat{\alpha}_W$, $\hat{\alpha}_D^*$ and $\hat{\alpha}_W^*$ of the significance level α, where

$$\hat{\alpha} = \frac{\text{Number of times } H_0 \text{ is rejected}}{M} .$$

In each case we were interested in a full model check: $F \in \mathcal{F}$ versus $F \notin \mathcal{F}$. Hence $p_1 = 0$ and $\hat{\theta}_{2n} = \hat{\theta}_n$. For $\hat{\theta}_{2n}$ we always took the MLE of θ. Finally, $M = 1000$ and $B = 500$.

Test for a Normal Distribution

Our first example deals with the univariate normal distribution $N(\mu, \sigma^2)$. Parameter estimators are

$$\hat{\mu} = \overline{X} = n^{-1} \sum_{i=1}^{n} X_i \qquad \text{and} \qquad \hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \ .$$

Random samples have been generated using the program GGNPM (generation of random deviates from a normal distribution using the polar method) from the IMSL library. Also, we used the program MDNOR from the same library to evaluate the Z's. Critical values for D_n and W_n^2 may be found in table 4.7 of D'Agostino and Stephens (1986). See also Lilliefors (1967). Table 1 compares the actual significance levels.

Table 1. Normal Distribution

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$n = 20$	$\hat{\alpha}_D = 0.102$ $\hat{\alpha}_W = 0.099$ $\hat{\alpha}_D^* = 0.100$ $\hat{\alpha}_W^* = 0.094$	$\hat{\alpha}_D = 0.042$ $\hat{\alpha}_W = 0.051$ $\hat{\alpha}_D^* = 0.043$ $\hat{\alpha}_W^* = 0.050$	$\hat{\alpha}_D = 0.015$ $\hat{\alpha}_W = 0.017$ $\hat{\alpha}_D^* = 0.016$ $\hat{\alpha}_W^* = 0.018$
$n = 50$	$\hat{\alpha}_D = 0.119$ $\hat{\alpha}_W = 0.108$ $\hat{\alpha}_D^* = 0.114$ $\hat{\alpha}_W^* = 0.109$	$\hat{\alpha}_D = 0.054$ $\hat{\alpha}_W = 0.045$ $\hat{\alpha}_D^* = 0.055$ $\hat{\alpha}_W^* = 0.045$	$\hat{\alpha}_D = 0.014$ $\hat{\alpha}_W = 0.010$ $\hat{\alpha}_D^* = 0.012$ $\hat{\alpha}_W^* = 0.009$
$n = 100$	$\hat{\alpha}_D = 0.108$ $\hat{\alpha}_W = 0.094$ $\hat{\alpha}_D^* = 0.099$ $\hat{\alpha}_W^* = 0.098$	$\hat{\alpha}_D = 0.056$ $\hat{\alpha}_W = 0.066$ $\hat{\alpha}_D^* = 0.061$ $\hat{\alpha}_W^* = 0.064$	$\hat{\alpha}_D = 0.016$ $\hat{\alpha}_W = 0.012$ $\hat{\alpha}_D^* = 0.013$ $\hat{\alpha}_W^* = 0.014$

Test for an Exponential Distribution

Recall that X is exponentially distributed with parameters α and β, iff the pertaining d.f. is

$$F(x; \alpha, \beta) = 1 - \exp\left\{-\frac{x - \alpha}{\beta}\right\}, \qquad x > \alpha, \beta > 0$$

Table 2. Exponential Distribution

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$n = 20$	$\hat{\alpha}_D = 0.075$ $\hat{\alpha}_W = 0.085$ $\hat{\alpha}_D^* = 0.080$ $\hat{\alpha}_W^* = 0.084$	$\hat{\alpha}_D = 0.049$ $\hat{\alpha}_W = 0.050$ $\hat{\alpha}_D^* = 0.054$ $\hat{\alpha}_W^* = 0.053$	$\hat{\alpha}_D = 0.008$ $\hat{\alpha}_W = 0.005$ $\hat{\alpha}_D^* = 0.010$ $\hat{\alpha}_W^* = 0.009$
$n = 50$	$\hat{\alpha}_D = 0.091$ $\hat{\alpha}_W = 0.094$ $\hat{\alpha}_D^* = 0.096$ $\hat{\alpha}_W^* = 0.097$	$\hat{\alpha}_D = 0.041$ $\hat{\alpha}_W = 0.045$ $\hat{\alpha}_D^* = 0.043$ $\hat{\alpha}_W^* = 0.043$	$\hat{\alpha}_D = 0.003$ $\hat{\alpha}_W = 0.007$ $\hat{\alpha}_D^* = 0.005$ $\hat{\alpha}_W^* = 0.008$
$n = 100$	$\hat{\alpha}_D = 0.083$ $\hat{\alpha}_W = 0.081$ $\hat{\alpha}_D^* = 0.090$ $\hat{\alpha}_W^* = 0.085$	$\hat{\alpha}_D = 0.043$ $\hat{\alpha}_W = 0.049$ $\hat{\alpha}_D^* = 0.045$ $\hat{\alpha}_W^* = 0.052$	$\hat{\alpha}_D = 0.015$ $\hat{\alpha}_W = 0.009$ $\hat{\alpha}_D^* = 0.011$ $\hat{\alpha}_W^* = 0.009$

and zero otherwise. In our simulation study, exponential variables were generated using the program GGEXN. $\alpha = 0$ was assumed to be known, and β was estimated by its MLE $\hat{\beta} = \bar{X}$. Critical values for D_n and W_n^2 may be found in table 4.11 in D'Agostino and Stephens (1986). See also Lilliefors (1969). Table 2 presents the actual significance levels.

Test for a Logistic Distribution

Now, X has d.f.

$$F(x; \alpha, \beta) = \frac{1}{1 + \exp\{-(x - \alpha)/\beta\}}, \qquad x \in \mathbb{R}, \beta > 0 .$$

The random variables were generated using the rejection method for the Laplace distribution. See Devroye (1986), p. 471. $\beta = 1$ was assumed to be known. The MLE of α solves the equation

$$n^{-1} \sum_{i=1}^{n} \left[\frac{1}{1 + \exp(X_i - \alpha)} \right] = 1/2$$

respectively

$$n^{-1} \sum_{i=1}^{n} [1/2 - F(X_i; \alpha, 1)] = 0 .$$

Table 3. Logistic Distribution

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$n = 20$	$\hat{\alpha}_D = 0.096$ $\hat{\alpha}_W = 0.109$ $\hat{\alpha}_D^* = 0.099$ $\hat{\alpha}_W^* = 0.108$	$\hat{\alpha}_D = 0.054$ $\hat{\alpha}_W = 0.054$ $\hat{\alpha}_D^* = 0.060$ $\hat{\alpha}_W^* = 0.051$	$\hat{\alpha}_D = 0.008$ $\hat{\alpha}_W = 0.010$ $\hat{\alpha}_D^* = 0.012$ $\hat{\alpha}_W^* = 0.009$
$n = 50$	$\hat{\alpha}_D = 0.113$ $\hat{\alpha}_W = 0.107$ $\hat{\alpha}_D^* = 0.116$ $\hat{\alpha}_W^* = 0.101$	$\hat{\alpha}_D = 0.052$ $\hat{\alpha}_W = 0.043$ $\hat{\alpha}_D^* = 0.049$ $\hat{\alpha}_W^* = 0.047$	$\hat{\alpha}_D = 0.004$ $\hat{\alpha}_W = 0.010$ $\hat{\alpha}_D^* = 0.005$ $\hat{\alpha}_W^* = 0.010$
$n = 100$	$\hat{\alpha}_D = 0.102$ $\hat{\alpha}_W = 0.084$ $\hat{\alpha}_D^* = 0.096$ $\hat{\alpha}_W^* = 0.083$	$\hat{\alpha}_D = 0.056$ $\hat{\alpha}_W = 0.047$ $\hat{\alpha}_D^* = 0.053$ $\hat{\alpha}_W^* = 0.048$	$\hat{\alpha}_D = 0.015$ $\hat{\alpha}_W = 0.018$ $\hat{\alpha}_D^* = 0.014$ $\hat{\alpha}_W^* = 0.015$

We have used a Newton-Raphson algorithm to solve the last equation:

$$\alpha^{(m+1)} = \alpha^{(m)} + \frac{\sum_{i=1}^{n} \psi(X_i - \alpha^{(m)})}{\sum_{i=1}^{n} \psi'(X_i - \alpha^{(m)})} ,$$

where $\psi(u) = 1/2 - F(u; 0, 1)$. Alternatively,

$$\alpha^{(m+1)} = \alpha^{(m)} - \frac{\frac{n}{2} - \sum_{i=1}^{n} F(X_i; \alpha^{(m)}, 1)}{\sum_{i=1}^{n} f(X_i; \alpha^{(m)}, 1)} ,$$

with f denoting the density of F and $\alpha^{(0)} = \overline{X}$. Critical values of D_n and W_n^2 are contained in tables 4.22 and 4.23 of D'Agostino and Stephens (1986). See also Stephens (1979). Table 3 presents the estimated significance levels.

Test for a Weibull Distribution

Recall that X has a Weibull distribution with parameters α and β, iff

$$F(x; \alpha, \beta) = 1 - \exp\{-(x/\beta)^{\alpha}\} , \qquad x > 0, \alpha > 0 \text{ and } \beta > 0 .$$

Table 4. Weibull Distribution

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$n = 20$	$\hat{\alpha}_D = 0.106$ $\hat{\alpha}_W = 0.114$ $\hat{\alpha}_D^* = 0.109$ $\hat{\alpha}_W^* = 0.115$	$\hat{\alpha}_D = 0.053$ $\hat{\alpha}_W = 0.054$ $\hat{\alpha}_D^* = 0.049$ $\hat{\alpha}_W^* = 0.060$	$\hat{\alpha}_D = 0.010$ $\hat{\alpha}_W = 0.010$ $\hat{\alpha}_D^* = 0.008$ $\hat{\alpha}_W^* = 0.012$
$n = 50$	$\hat{\alpha}_D = 0.108$ $\hat{\alpha}_W = 0.103$ $\hat{\alpha}_D^* = 0.107$ $\hat{\alpha}_W^* = 0.097$	$\hat{\alpha}_D = 0.041$ $\hat{\alpha}_W = 0.042$ $\hat{\alpha}_D^* = 0.039$ $\hat{\alpha}_W^* = 0.043$	$\hat{\alpha}_D = 0.009$ $\hat{\alpha}_W = 0.012$ $\hat{\alpha}_D^* = 0.009$ $\hat{\alpha}_W^* = 0.011$
$n = 100$	$\hat{\alpha}_D = 0.078$ $\hat{\alpha}_W = 0.094$ $\hat{\alpha}_D^* = 0.089$ $\hat{\alpha}_W^* = 0.086$	$\hat{\alpha}_D = 0.036$ $\hat{\alpha}_W = 0.037$ $\hat{\alpha}_D^* = 0.045$ $\hat{\alpha}_W^* = 0.039$	$\hat{\alpha}_D = 0.011$ $\hat{\alpha}_W = 0.014$ $\hat{\alpha}_D^* = 0.011$ $\hat{\alpha}_W^* = 0.012$

In our simulation study we considered α to be known, $\alpha = 2$. Generation of random samples has been done using the program GGWIB from the IMSL library. As is well known, the log-transformation leads to random variables $Y = -\log X$ having an extreme-value distribution of Gumbel type:

$$F(y) = \exp\left[-\exp\left[-\frac{y-\Phi}{\delta}\right]\right], \qquad y \in \mathbb{R},$$

with $\delta = \alpha^{-1}$ and $\Phi = -\log \beta$. Hence testing about a Weibull distribution is equivalent to testing about an extreme value distribution. The MLE of Φ equals

$$\hat{\Phi} = -\delta \log\left[n^{-1} \sum_{i=1}^{n} \exp(-Y_i/\delta)\right].$$

Critical values of D_n and W_n^2 are contained in tables 4.17 and 4.18 of D'Agostino andd Stephens (1986). See also Stephens (1977) and Chandra, Singpurwalla and Stephens (1981). Table 4 presents the estimated significance levels.

Test for a Bivariate Normal Distribution

For our last example we generated a random sample from a bivariate normal distribution using the procedure as described in Devroye (1986), p. 566–567. Only tests based on the Kolmogorov-Smirnov distance will be considered. Note

that given a sample $(X_1^1, X_1^2), \ldots, (X_n^1, X_n^2)$ from $\mathcal{N}(\mu, \Sigma)$, then D_n can be computed as follows:

$$D_n = \sqrt{n} \max_{1 \le k \le n} \left\{ \max_{1 \le j \le k} \left[\frac{j}{n} - Z_{k,j}; Z_{k,j} - \frac{j-1}{n}; \hat{F}_1(X_{k:n}^1) - \frac{k-1}{n} \right] \right\} ,$$

with $Z_{k,j} = \hat{F}(X_{k:n}^1, X_{[j:n]}^2)$, $1 \le k \le n$, $1 \le j \le k$ and $X_{[j:n]}^2$ denoting the concomitant of $X_{j:n}^1$. $\hat{F} = F(\hat{\mu}, \hat{\Sigma})$ is the estimated normal distribution and $\hat{F}_1$ denotes its first marginal.

We made a comparison between a goodness-of-fit test based on D_n, using bootstrap based percentiles, and Mardia's test for multivariate normality, based on multivariate measures of skewness β_1 and kurtosis β_2. See, e.g., Seber (1984), p. 149. To compute the $Z_{k,j}$ and $F_1(X_{k:n}^1)$, we used the programs MOBNOR and MDNOR, respectively, from the IMSL library. As estimators of μ and Σ we considered the sample mean $\hat{\mu}$ and sample covariance matrix $\hat{\Sigma}$, respectively.

To apply Mardia's test, note that under H_0: F is normal, $\beta_1 = 0$ and $\beta_2 = 8$. Tests are based on

$$\hat{\beta}_1 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}^3 \qquad \text{and} \qquad \hat{\beta}_2 = \frac{1}{n} \sum_{i=1}^{n} g_{ii}^2 ,$$

where

$$g_{ij} = (X_i - \hat{\mu})^t \hat{\Sigma}^{-1} (X_j - \hat{\mu}) .$$

Under the null hypothesis, we have approximately

$$A_n \equiv \frac{1}{6} n \hat{\beta}_1 \sim \chi_4^2 \qquad \text{and} \qquad B_n \equiv n^{1/2} \frac{\hat{\beta}_2 - 8}{8} \sim \mathcal{N}(0, 1) .$$

For a given critical level α, H_0 is rejected if $A_n \ge \chi_{1-\alpha,4}^2$ or if $|B_n| \ge \Phi^{-1}((1-\alpha)/2)$. In each case $M = 1000$ samples were drawn. The number of bootstrap replicates was $B = 100$ in this example. Table 5 presents the estimated significance levels $\hat{\alpha}_S$ and $\hat{\alpha}_K$ for Mardia's tests based on skewness and kurtosis as well as for the bootstrap based Kolmogorov-Smirnov test. It turns out that the bootstrap based test is satisfactory if not excellent while Mardia's tests often have a low power.

Table 5. Bivariate Normal Distribution

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$n = 20$	$\hat{\alpha}_S = 0.061$ $\hat{\alpha}_K = 0.005$ $\hat{\alpha}_D^* = 0.091$	$\hat{\alpha}_S = 0.025$ $\hat{\alpha}_K = 0.003$ $\hat{\alpha}_D^* = 0.052$	$\hat{\alpha}_S = 0.006$ $\hat{\alpha}_K = 0.000$ $\hat{\alpha}_D^* = 0.009$
$n = 50$	$\hat{\alpha}_S = 0.069$ $\hat{\alpha}_K = 0.014$ $\hat{\alpha}_D^* = 0.099$	$\hat{\alpha}_S = 0.039$ $\hat{\alpha}_K = 0.011$ $\hat{\alpha}_D^* = 0.067$	$\hat{\alpha}_S = 0.022$ $\hat{\alpha}_K = 0.009$ $\hat{\alpha}_D^* = 0.010$
$n = 100$	$\hat{\alpha}_S = 0.087$ $\hat{\alpha}_K = 0.025$ $\hat{\alpha}_D^* = 0.085$	$\hat{\alpha}_S = 0.046$ $\hat{\alpha}_K = 0.016$ $\hat{\alpha}_D^* = 0.043$	$\hat{\alpha}_S = 0.015$ $\hat{\alpha}_K = 0.010$ $\hat{\alpha}_D^* = 0.011$

3 Appendix

It is worthwhile remembering the main steps in proving Durbin's result. We only deal with real-valued data here. Write, in terms of a uniform empirical process $\bar{\alpha}_n$,

$$\hat{\alpha}_n(x) = n^{1/2}[F_n(x) - F(x;\theta)] + n^{1/2}[F(x;\theta)] - F(x;\hat{\theta}_n)]$$

$$= \bar{\alpha}_n(F(x;\theta)) - n^{-1/2}\sum_{i=1}^{n} l^t(X_i,\theta)g(x,\theta) + o_{\mathbb{P}}(1)\ ,$$

where $o_{\mathbb{P}}(1)$ is uniform in x. Observe that both terms are asymptotically C-tight: the first by C-tightness of $\bar{\alpha}_n$ and continuity of $F(\cdot,\theta)$, and the second by uniform continuity of $g(\cdot,\theta)$ and asymptotic normality (and hence boundedness in probability) of $n^{-1/2}\sum_{i=1}^{n} l^t(X_i,\theta)$. Convergence of the fidis follows from the Cramér-Wold device and the CLT.

Now, for α_n^* the above decomposition becomes

$$\alpha_n^*(x) = n^{1/2}[F_n^*(x) - F(x;\hat{\theta}_n)] + n^{1/2}[F(x;\hat{\theta}_n) - F(x;\bar{\theta}_1,\theta_{2n}^*)]$$

$$= \bar{\alpha}_n(F(x;\hat{\theta}_n)) - n^{-1/2}\sum_{i=1}^{n} l^t(X_i^*,\hat{\theta}_n)g(x,\tilde{\theta}_n) - \varepsilon_{1n}^{*t}g(x,\tilde{\theta}_n) \tag{3.1}$$

for some $\tilde{\theta}_n = (\bar{\theta}_1, \tilde{\theta}_{2n})$, with $\tilde{\theta}_{2n}$ between $\hat{\theta}_{2n}$ and $\hat{\theta}_{2n}^*$. To analyze the finite dimensional distributions of α_n^* observe that for real numbers $x_1, \ldots, x_p$ and $a_1, \ldots, a_p$

$$\sum_{j=1}^{p} a_j \alpha_n^*(x_j) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{p} a_j [1_{\{X_i^* \le x_j\}} - F(x_j; \hat{\theta}_n) - l^t(X_i^*, \hat{\theta}_n) g(x_j, \tilde{\theta}_n)]$$

$$- \sum_{j=1}^{p} a_j \varepsilon_{1n}^{*t} g(x_j, \tilde{\theta}_n) \ .$$

Lemma: With probability one,

$$\mathbb{P}_n^*(|\hat{\theta}_{2n}^* - \hat{\theta}_{2n}| \ge \varepsilon) \to 0 \ , \qquad \text{for each } \varepsilon > 0 \ .$$

Proof: By assumption, $\varepsilon_{1n}^* \to 0$ in $\mathbb{P}_n^*$-probability. Furthermore

$$\mathrm{Cov}_n^*[l(X_1^*, \hat{\theta}_n) l^t(X_1^*, \hat{\theta}_n)] = \mathrm{Cov}[l(X_1, \hat{\theta}_n) l^t(X_1, \hat{\theta}_n)] \ ,$$

where $X_1 \sim F(\cdot; \hat{\theta}_n)$. Since $\hat{\theta}_n \to \theta$ with probability one by assumption, the covariances converge to $\mathrm{Cov}[l(X_0, \theta) l^t(X_0, \theta)]$, $X_0 \sim F(\cdot; \theta)$, by (iii). Conclude that $\hat{\theta}_{2n}^* - \hat{\theta}_{2n} \to 0$ in quadratic mean as $n \to \infty$, with probability one. This proves the Lemma. □

Lemma: With probability one,

$$\mathbb{P}_n^*(|\hat{\theta}_{2n}^* - \theta_2| \ge \varepsilon) \to 0 \ , \qquad \text{for each } \varepsilon > 0 \ .$$

Proof: We have

$$\mathbb{P}_n^*(|\hat{\theta}_{2n}^* - \theta_2| \ge \varepsilon) \le \mathbb{P}_n^*(|\hat{\theta}_{2n}^* - \hat{\theta}_{2n}| \ge \varepsilon/2) + 1_{\{|\hat{\theta}_{2n} - \theta_2| \ge \varepsilon/2\}} \to 0 \qquad \mathbb{P} - \text{a.s.} \ ,$$

upon using the last Lemma and $\hat{\theta}_{2n} \to \theta_2$ with probability one. □

We are now in the position to give the

Proof of Theorem 1.1: It follows from the last lemma that along with $\hat{\theta}_{2n}^*$ and $\hat{\theta}_{2n}$, $\tilde{\theta}_n \to \theta_2$ in $\mathbb{P}_n^*$-probability, for $\mathbb{P}$-almost all basic sample sequences. By continuity of $g(x_j, \cdot)$ it thus suffices to analyze

$$n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{p} a_j [1_{\{X_i^* \le x_j\}} - F(x_j; \hat{\theta}_n) - l^t(X_i^*, \hat{\theta}_n) g(x_j, \theta)] \ .$$

It is a sum of centered i.i.d. random variables w.r.t. $\mathbb{P}_n^*$. Its variance equals

$$\sum_{j=1}^{p}\sum_{k=1}^{p} a_j a_k \Bigg[F(x_j \wedge x_k; \hat{\theta}_n) - F(x_j; \hat{\theta}_n)F(x_k; \hat{\theta}_n) - \int_{-\infty}^{x_j} l^t(x, \hat{\theta}_n) g(x_k, \hat{\theta}_n) F(dx; \hat{\theta}_n) - \int_{-\infty}^{x_k} l^t(x, \hat{\theta}_n) g(x_j, \hat{\theta}_n) F(dx; \hat{\theta}_n) + \int g^t(x_j, \hat{\theta}_n) l(x, \hat{\theta}_n) l^t(x, \hat{\theta}_n) g(x_k, \hat{\theta}_n) F(dx; \hat{\theta}_n) \Bigg] .$$

Since $\hat{\theta}_n \to \theta$ with probability one we find from (ii), (iii) and the continuity of F, that the last sum converges to

$$\mathrm{Var}\left[\sum_{j=1}^{p} a_j Z(x_j)\right] \qquad \mathbb{P} - \text{a.s.} .$$

The verification of Lindeberg's condition is similar. Finally, asymptotic C-tightness of α_n^* follows from (3.1), upon using the asymptotic C-tightness of $\bar{\alpha}_n$ and the equicontinuity of $F(\cdot; \theta)$, $\theta \in U$, on the one hand, and the asymptotic boundedness of $n^{-1/2} \sum_{i=1}^{n} l^t(X_i^*, \hat{\theta}_n)$ and equicontinuity of $g(\cdot, \tilde{\theta})$, $\tilde{\theta} \in U$, on the other hand. The proof is complete. □

Acknowledgment: The first author's work was supported by a grant from the Xunta Galicia.

References

Bickel PJ, Freedman DA (1981) Some asymptotic theory for the bootstrap. Ann Statist 9:1196–1217

Chandra M, Singpurwalla ND and Stephens MA (1981) Kolmogorov statistics for tests of fit for the extreme value and Weibull distributions. JASA 76:719–731

D'Agostino RB, Stephens MA (1986) Goodness-of-fit techniques. Marcel Dekker, New York

Devroye L (1986) Non-Uniform random variate generation. Springer, New York – Berlin – Heidelberg

Durbin J, Knott M (1972) Components of Cramér-von Mises statistics. I J Roy Statist Soc B 34:290–307

Durbin J (1973) Weak convergence of the sample distribution function when parameters are estimated. Ann Statist 1:279–290
Durbin J, Knott M and Taylor CC (1975) Components of Cramér-von Mises statistics. II J Roy Statist Soc B 37:216–237
Lilliefors HW (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. JASA 62:399–402
Lilliefors HW (1969) On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. JASA 64:387–389
Pollard D (1984) Convergence of stochastic processes. Springer, New York – Berlin – Heidelberg
Seber GAF (1984) Multivariate observation. Wiley, New York
Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley, New York
Stephens MA (1974) EDF statistics for goodness-of-fit and some comparisons. JASA 69:730–737
Stephens MA (1976) Asymptotic results for goodness-of-fit statistics with unknown parameters. Ann Statist 4:357–369
Stephens MA (1977) Goodness-of-fit for the extreme value distribution. Biometrika 64:583–588
Stephens MA (1979) Tests of fit for the logistic distribution based on the empirical distribution function. Biometrika 66:591–595

Received 04.06.1992
Revised version 28.10.1992

Book Reviews

Lee AL: *U-Statistics, Theory and Practice.* Statistics: Textbooks and Monographs. Marcel Dekker, New York Basel 1990, Vol. 110, p. xi+302

This monograph on nonparametric statistic fits well into the series of books which appeared during the last decade on this topic. Its merit lies in a good combination of theory and practice. Although the presentation is kept on an elementary level, the reader will find many topics presented which lead to mathematically very interesting and fruitful areas of research. Likewise, the problems arising in applications are treated with great care. It is clear that a book of this size cannot give all aspects of the theory, however, the choices made by the author agree in large parts with the common taste. This may be seen from the following brief sketch of its content.

Since U-statistics are unbiased minimum variance estimators, they play a central role in statistical inference for parametric and nonparametric models. This motivation is explained at the beginning of this book on an elementary level. The reader will not get a deeper insight from the viewpoint of sufficiency and completeness and should contact other sources. A second motivation for studying U-statistics is purely analytic, since these statistics may be regarded as a generalized averaging method. Therefore a good deal of the monograph is concerned with the development of the analysis for U-statistics. There are two fundamental and elementary analytical properties, the Hoeffding decomposition and the representation of variances. These topics are covered in Chapter 1 together with some standard examples, illustrating the results.

The author takes care of recent developments in the following chapter, extending the basic analysis to variants of U-statistics under the classical assumption of i.i.d. observations to other models. In particular, the reader will find results on generalized U-statistics, non-identical distributed observations, m-dependent and weakly dependent process, sampling from finite populations and generalized L-statistics. The book gives an excellent overview over the mathematical applicability of the theory, but examples for applications in the non i.i.d. case are missing. (This might be too ambitious.) The results are not always stated in full generality or under the weakest assumptions, but this also seems to be too much expected.

Chapter 3 contains the asymptotic distribution theory for U-statistics. Hoeffding's result about the asymptotic normality in the non-degenerate case is covered and also the degenerate case for degree 2 kernels, where the asymptotic distribution is given by a "weighted" χ^2 distribution. This is followed by a brief description of the approach of Rubin, Vitale, Dynkin, Mandelbaum and Taqqu on the representation of limits as multiple Wiener–Ito integrals. The reader will find more details in Chapter 4 in connection with symmetric statistics. The alternative approach (and conceptually simpler) of using empirical distributions is not at all treated in the book. This is certainly a disadvantage. The distributional convergence is completed by a result on convergence to Poisson laws.

More advanced convergence results are covered in the following chapter. A basic result on the Berry Esseen approximation is proved and the problems concerning Edgeworth expansion are outlined. A third analytic tool is introduced next, showing the martingale properties of U-statistics. From this the a.s. convergence (in fact 3 different proofs are given) and maximal inequalities are obtained. As a result one obtains the law of the iterated logarithm and invariance principles. While the non-degenerate case is proved (an easy proof), the mathematically more interesting degenerate case is treated expositorily. The remaining part of this chapter contains some asymptotic results for variants of U-statistics, the case when the kernel contains estimated parameters is of special interest for applications.

Chapter 4 on related statistics builds the bridge to other types of statistics. Symmetric statistics were mentioned before. The close connection of U-statistics to V-statistics (differential statistical functionals in the sense of v. Mises) is discussed only briefly, important examples are missing in this context. A larger part is devoted to incomplete U-statistics, which are especially interesting for applications (they are easier to compute). Two different designs (balanced designs and random designs) are discussed in great detail, including, for example, efficiency considerations.

One of the essential problems in applications of U-statistics is the estimation of the asymptotic variance of U-statistics. The jackknife method is very often used and hence treated in Chapter 5 extensively, and not only for the variance estimate. The reader will find it helpful that this estimate is compared with the bootstrap method in some special cases. The final chapter contains thoroughly chosen examples: estimation of correlations, tests for symmetry, normality, independence, several sample problems, analysis of variance, spatial statistics and sequential estimation.

The monograph presents an excellent overview of both, theory and applications. The subjects chosen are certainly due to the authors own interest and taste, so the reader will find some parts lengthy, some topics missing or mentioned only briefly. The reviewer found it a bit inadequate and inconsistent that large parts are written on an 'elementary' statistical basis while other parts (sections 3 and 4) use 'heavy' machinery, which is introduced and explained insufficiently for the unexperienced reader. Besides this, the mathematical part is overall well written.

This book on U-statistics is one of the best and can be recommended without hesitation for teaching and research.

Göttingen *M. Denker*

Torgersen E: *Comparison of Statistical Experiments.* Cambridge University Press, Cambridge New York Port Chester Melbourne Sydney 1991, p. xviii+675, $ 99.50, £ 60.00

Comparison of statistical experiments emerged during the Fifties from the fundamental research work of D. Blackwell and L. Le Cam. Based on the decision theory of J. Neyman and A. Wald they compared the risk functions the experiments produce for varying losses. In the hands of Le Cam comparison of experiments became an important tool for asymptotical statistics. This is documented in his famous monograph [2].

Inspired by this work the author of the present book has made substantial contributions to the theory of comparison of experiments. Choosing his own original line of approach he developed important fields within this theory, e.g. comparison of finite experiments, of linear normal models, with totally (non-)informative experiments, of (signed) measure families, and geometric criteria for comparison. This profound research work and lectures of the author given at the University of Oslo have formed the basis for the book under review.

The volume is divided into ten chapters, each opening with a substantial introduction and concluding with a comprehensive list of references.

In Chapter 1 (Statistical experiments within the measure theoretical framework) the basic concept of a statistical experiment is introduced as $E = (X, A, P_\theta : \theta \in \Theta)$ where every P_θ is a probability measure on the measurable space (X, A) and where Θ is a parameter set. Special emphasis is laid on coherent experiments as a natural and useful extension of dominated experiments.

In Chapter 2 (Convexity) standard results from convex analysis (together with full proofs) are assembled, for application in the sequel.

In Chapter 3 (Two-person, zero-sum games) those parts of game theory are discussed that are especially useful for decision theory.

In Chapter 4 (Statistical decision problems) a decision rule is defined as a Markov kernel p from the sample space (X, A) of E to some (measurable) decision space (T, S). Given a loss function $L : \Theta \times T \to]-\infty, \infty]$ one defines the corresponding risk (expected loss) of p at θ as $r(\theta : p) = P_\theta p L_\theta$.

In Chapter 5 (Vector lattices) a short introduction to Riess spaces (in particular to L- and M-spaces) is given. The utility of this concept in the theory of statistical experiments has been known since Le Cam's famous 1964 paper.

In Chapter 6 (Deficiencies) the central subject of the book is taken up: Given two experiments E and F with the same parameter set Θ the deficiency $\delta(E, F)$ of E with respect to F measures what one may lose by basing oneself on E rather than on F under the most unfavourable circumstances for E for this comparison. All of the general criteria for comparison are discussed including reduction by invariance.

In Chapter 7 (Equivalence, representations and functionals of experiments) – together with Chapter 6 the core of this work – the well known characterizations of types of experiments are discussed, e.g. by standard experiments, Hellinger transforms, sufficiency, conical measures, likelihood experiments.

In Chapter 8 (Comparison of linear models) the general theory is applied to finite dimensional linear (normal) experiments. This illuminating topic is intensively discussed.

In Chapter 9 (Majorization and approximate majorization) the comparison is extended to (signed) measure families thus shedding new light on the comparison of experiments. The theory may be considered as a vast generalization of well known theories on majorization.

In Chapter 10 (Complements: further examples, problems and comments) interesting additional information on the main themes is provided in a mixed example/problem form. Broad scope in this book is given to motivate and apply the general theory, and there are many examples. Special emphasis is laid on the contributions of the author and his school. Thus many results appear for the first time in book form.

There are only a few books which have a certain overlap with Torgersen's; principially, the books by Heyer [1], Le Cam [2], and Strasser [3]. In [1] and [3] a good portion of classical statistics is also included e.g. on testing and estimation. In [2] and [3] asymptotical methods are intensively discussed. These themes are only touched in Torgersen's book, which deals almost exclusively with comparison for a fixed sample size. But within this context it covers much more material in much more detail than in the other books mentioned above.

This monograph is indispensable for every mathematician who wants to learn what it means and what is going on in 'comparison of experiments'. The mathematical treatment is concise but not boring. The presentation is excellent; and there are only a few misprints. For short an enjoyable book; not only for statisticians but also for mathematicians.

References

[1] Heyer H (1982) Theory of Statistical Experiments. Springer
[2] Le Cam L (1986) Asymptotic Methods in Statistical Decision Theory. Springer
[3] Strasser H (1985) Mathematical Theory of Statistics. De Gruyter

Tübingen *E. Siebert*

Rozovskii BL: *Stochastic Evolution Systems.* Kluwer Acad. Publishers, Mathematics and its Applications (Soviet Series). Dordrecht 1991, p. xviii+315, Dfl. 210.00, US $ 129.00, UK £ 73.00

The book gives a first comprehensive account on the theory of stochastic partial differential equations. This is a fairly recently developed new branch in the field of stochastic differential equations, which is beyond the scope of the nowadays classical theory of ordinary stochastic differential equations and the Ito calculus. The need for this new type of stochastic equations has several origins. Since the mid-seventies an increasing number of models (mainly from physics, biology) occurred, which could be described by such stochastic partial differential equations. On the other hand, there is also an inner need for studying such equations. For example, the problem of non-linear filtering of diffusion processes leads naturally to a second order stochastic partial differential equation.

In principle, the book is the English translation of the Russian edition, published already in 1983. But the last chapter (using the Malliavin calculus) is newly added and also other parts of the book have been extended.

An illustrative survey of examples of stochastic partial differential equations arising in different fields is given at the beginning of the first chapter. Here and then mainly in the second chapter, the basic stochastic calculus is presented which is needed for the general theory. Especially, a thorough introduction of stochastic integration in a Hilbert space is given.

Chapter 3 presents the main tool for solving stochastic partial differential equations. This is the theory of so-called linear stochastic evolution systems in Hilbert spaces. A stochastic evolution system is governed by the equation $du(t) = A(t, u(t))dt + B(t, u(t))dW(t)$. Here, the process $u(t)$ takes values in a function space X, the coefficients A and B (the drift and the diffusion term) are in general unbounded operators, and $W(t)$ is a function space valued "white noise". The unboundedness of the coefficients A and B is handled by embedding the space X of the process into a certain Gelfand triple (X, H, X') of Hilbert spaces. In the later applications such triples are mostly triples of certain Sobolev spaces. Under certain conditions on A and B (coercivity or dissipativity) general theorems on the existence and uniqueness of solutions of linear stochastic evolution equations are proved. These results are then applied to the Dirichlet problem for stochastic partial differential equations (of arbitrary finite order).

In Chapter 4 stochastic second order parabolic equations are treated. Beside a general result on solutions of the Cauchy problem, based on the theory developed in the previous chapter, further analytic properties of the solutions are proved. Chapter 5 continues with the study of the Cauchy problem. It is shown that there are natural mutual connections between the stochastic parabolic equations and certain associated diffusion processes.

These relations to the classical theory of ordinary stochastic differential equations turn out to be essential for the applications to the theory of filtering, extrapolation and interpolation presented in Chapter 6. There it is shown that the so-called filtering measure of a filtering problem given by a diffusion process has

a density with respect to Lebesgue measure, and that this density is the solution of a stochastic parabolic differential equation (the forward filtering equation). A corresponding backward filtering equation is studied for the problems of interpolation and extrapolation.

The last chapter is concerned with the hypoellipticity of second order stochastic differential equations and in particular of the filtering equations.

In the preface it is said that "the book was written to be understood by researchers of different interests and educational background" and that the necessary prerequisites for a reader are "familiarity with functional analysis and the theory of stochastic processes within the framework of standard university courses". This is surely an understatement and the book is only readable and valuable for people already working at least in related fields. More references and motivations between the standard definition-theorem pattern would be helpful. The prerequisites are not sufficient for an understanding. One rather annoying reason is that the book contains too many misprints or wrong notations. Probably, there was never any proofreading. I mention only one typical example: At the beginning of Chapter 4 the defining equation of the second order stochastic parabolic equation is written completely false – and this is the subject of the whole chapter! The symbols for certain variables change almost from line to line, and it is a real puzzle for the reader to finally find his own correct version of the parabolic equation. The usage of a summation convention is – admittedly – a matter of taste, but I believe that a consequent coordinate free notation is surely preferable to a notation that is too often wrong if more than one index variable occurs.

Nevertheless, this is a book on a new and fascinating subject and there are also large parts which are very well presented. So the book is surely of high value for anybody who is getting interested or already working in this field.

Tübingen *E. Dettweiler*

Metrika (1993) 40:263–269

Metrika

MSE-Improvement of the Least Squares Estimator by Dropping Variables

ERKKI P. LISKI[1]

University of Tampere, P.O. Box 607, SF-33101 Tampere, Finland

GÖTZ TRENKLER[2]

University of Dortmund, P.O. Box 50 05 00, 4600 Dortmund 50, Germany

Abstract: It is well known that dropping variables in regression analysis decreases the variance of the least squares (LS) estimator of the remaining parameters. However, after elimination estimates of these parameters are biased, if the full model is correct. In his recent paper, Boscher (1991) showed that the LS-estimator in the special case of a mean shift model (cf. Cook and Weisberg, 1982) which assumes no "outliers" can be considered in the framework of a linear regression model where some variables are deleted. He derived conditions under which this estimator outperforms the LS-estimator of the full model in terms of the mean squared error (MSE)-matrix criterion. We demonstrate that this approach can be extended to the general set-up of dropping variables. Necessary and sufficient conditions for the MSE-matrix superiority of the LS-estimator in the reduced model over that in the full model are derived. We also provide a uniformly most powerful F-statistic for testing the MSE-improvement.

Key Words and Phrases: Dropping variables, Mean squared error superiority, least squares estimator, F-test.

1 Introduction

The implications of dropping variables in regression analysis have been investigated by a number of authors. For example, Rao (1971), Hocking (1974) and Judge et al. (1985, pp. 857–862) have derived sufficient conditions for situations where omission of variables leads to better estimates with respect to the matrix mean squared error (MMSE).

Boscher (1991) was successful to derive necessary and sufficient conditions for dominance of the dropping variables least squares estimator (LSE) over the ordinary LSE in terms of MMSE for the special case of a mean-shift model. In

[1] A first draft of this manuscript was prepared when I was a Humboldt research fellow in the University of Dortmund. The financial support of the Humboldt Foundation is gratefully acknowledged. Finally, I thank Professor Götz Trenkler, University of Dortmund, for his hospitality.

[2] Support by "The German Marshall Fund of the United States" is gratefully acknowledged. This author also wishes to thank Professor George Judge, University of Berkeley, for his kind hospitality.

0026–1335/93/5/263–269 $2.50

the following we generalize these results to the general regression model where some of the variables are dropped. Furthermore we provide a uniformly most powerful F-statistic for testing the MMSE-improvement.

2 The Linear Regression Model

Consider the linear regression model

$$\mathbf{y} = \mathbf{W}\boldsymbol{\theta} + \mathbf{u} \ , \tag{2.1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations on the dependent variable, $\mathbf{W}$ is an $n \times p$ matrix of observations on p independent variables of full column rank, $\mathbf{u}$ is an $\mathbf{n} \times 1$ vector of unobservable disturbances, and $\boldsymbol{\theta}$ is an unknown $p \times 1$ vector of parameters. We assume $E(\mathbf{u}) = 0$ and $E(\mathbf{uu}') = \sigma^2\mathbf{I}$, where $\sigma^2 > 0$ is unknown. If we allow a more general error process, e.g. an autoregressive process, then in principle model (2.1) could be transformed such that the transformed disturbances have the simple independence structure. In order to determine the transformation matrix we should know $\text{cov}(\mathbf{u}) = \boldsymbol{\Sigma}$, which is not available in practice. So, the transformation technique can be applied only conditionally, provided the covariance matrix $\boldsymbol{\Sigma}$ is given. If the unknown covariance can be estimated from the data, then also the transformed $\mathbf{W}$ depends on $\mathbf{y}$ and the distributional assumptions, for example in Section 3, do not hold.

Suppose that $\mathbf{W} = (\mathbf{X}, \mathbf{Z})$, where $\mathbf{X}$ and $\mathbf{Z}$ are $n \times p_1$ and $n \times p_2$ matrices, respectively. Then (2.1) may be alternatively written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \ , \tag{2.2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ and $\mathbf{v} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}$. If $\mathbf{Z}$ is available, then the LSE of $\boldsymbol{\beta}$ is

$$\bar{\boldsymbol{\beta}} = [\mathbf{X}'(\mathbf{I} - \mathbf{Z}\mathbf{Z}^+)\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{Z}\mathbf{Z}^+)\mathbf{y} \ , \tag{2.3}$$

where $\mathbf{Z}^+ = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ denotes the Moore-Penrose inverse of $\mathbf{Z}$. However if we drop the observation matrix $\mathbf{Z}$ from (2.2), the formal LSE of β becomes

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^+\mathbf{y}$$

which has bias

$$E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{X}^+\mathbf{Z}\gamma \ . \tag{2.4}$$

For the sake of convenience we introduce the following orthogonal projectors

$$\mathbf{P}_X = \mathbf{X}\mathbf{X}^+ \ , \qquad \mathbf{P}_Z = \mathbf{Z}\mathbf{Z}^+ \tag{2.5}$$

and

$$\mathbf{Q}_X = \mathbf{I} - \mathbf{P}_X \ , \qquad \mathbf{Q}_Z = \mathbf{I} - \mathbf{P}_Z \ . \tag{2.6}$$

To compare $\bar{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ in terms of their MMSE we need their dispersion matrices, which are given by

$$\operatorname{cov}(\bar{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{Q}_Z\mathbf{X})^{-1} \tag{2.7}$$

and

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \ .$$

Then we obtain as the corresponding MMSE

$$M(\bar{\boldsymbol{\beta}}) = \operatorname{cov}(\bar{\beta}) = \sigma^2(\mathbf{X}'\mathbf{Q}_Z\mathbf{X})^{-1} \tag{2.8}$$

and

$$M(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{X}^+\mathbf{Z}\gamma\gamma'\mathbf{Z}'\mathbf{X}^{+\prime} \ , \tag{2.9}$$

and the difference of MMSE's becomes

$$\begin{aligned} M(\bar{\boldsymbol{\beta}}) - M(\hat{\boldsymbol{\beta}}) &= \sigma^2[(\mathbf{X}'\mathbf{Q}_Z\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}] - \mathbf{X}^+\mathbf{Z}\gamma\gamma'\mathbf{Z}'\mathbf{X}^{+\prime} \\ &= (\mathbf{X}'\mathbf{Q}_Z\mathbf{X})^{-1}\mathbf{D}(\mathbf{X}'\mathbf{Q}_Z\mathbf{X})^{-1} \ , \end{aligned} \tag{2.10}$$

where

$$\mathbf{D} = \mathbf{A} - \mathbf{a}\mathbf{a}' \tag{2.11}$$

with

$$\mathbf{A} = \sigma^2\mathbf{X}'\mathbf{Q}_Z\mathbf{Q}_X\mathbf{Q}_Z\mathbf{X} \qquad \text{and} \qquad \mathbf{a} = \mathbf{X}'\mathbf{Q}_Z\mathbf{P}_X\mathbf{Z}\gamma \ .$$

Obviously $M(\bar{\boldsymbol{\beta}}) - M(\hat{\boldsymbol{\beta}})$ is nonnegative definite (n.n.d.) if and only if $\mathbf{D}$ is n.n.d.

Theorem 1: The least squares estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ which drops the observations from $\mathbf{Z}$ dominates the least squares estimator $\bar{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{Q}_Z X)^{-1}\mathbf{X}'\mathbf{Q}_Z\mathbf{y}$ with respect to MMSE if and only if

$$\gamma'\mathbf{Z}'\mathbf{P}_G\mathbf{Z}\gamma \leq \sigma^2 \ , \tag{2.12}$$

where

$$\mathbf{P}_G = \mathbf{G}(\mathbf{G}'\mathbf{G})^+\mathbf{G}' = \mathbf{G}\mathbf{G}^+$$

is the orthogonal projector on the column space of $\mathbf{G} = \mathbf{Q}_X\mathbf{Q}_Z\mathbf{X}$.

Proof: By our lemma from the appendix the matrix $\mathbf{D}$ is n.n.d. iff $\mathbf{A}$ is n.n.d., $\mathbf{a} \in \mathfrak{R}(\mathbf{A})$ and $\lambda_1(\mathbf{a}'\mathbf{A}^+\mathbf{a}) = \mathbf{a}'\mathbf{A}^+\mathbf{a} \leq 1$. Clearly $\mathbf{A}$ is n.n.d. and $\mathbf{a} = \mathbf{X}'\mathbf{Q}_Z\mathbf{P}_X\mathbf{Z}\gamma = -\mathbf{X}'\mathbf{Q}_Z\mathbf{Q}_X\mathbf{Z}\gamma$ belongs to $\mathfrak{R}(\mathbf{A}) = \mathfrak{R}(\mathbf{X}'\mathbf{Q}_Z\mathbf{Q}_X)$. Thus $\mathbf{a}'\mathbf{A}^+\mathbf{a} = \gamma'\mathbf{Z}'\mathbf{G}(\mathbf{G}'\mathbf{G})^+\mathbf{G}'\mathbf{Z}\gamma/\sigma^2$, and the third condition turns out to be

$$\gamma'\mathbf{Z}'\mathbf{P}_G\mathbf{Z}\gamma \leq \sigma^2 \ . \qquad \blacksquare$$

To illustrate the potential gains in precision we write (2.10) in the form

$$M(\bar{\boldsymbol{\beta}}) - M(\hat{\boldsymbol{\beta}}) = \mathbf{X}^+\mathbf{Z}[\mathrm{cov}(\hat{\gamma}) - \gamma\gamma']\mathbf{Z}'\mathbf{X}^{+\prime} \ ,$$

which can be proved straightforwardly (c.f. Hocking 1974). Thus $M(\bar{\boldsymbol{\beta}}) - M(\hat{\boldsymbol{\beta}})$ is n.n.d., if $\mathrm{cov}(\hat{\gamma}) - \gamma\gamma'$ is n.n.d. If only a single variable, $\mathbf{z}$, is excluded, $M(\bar{\boldsymbol{\beta}}) \geq M(\hat{\boldsymbol{\beta}})$ if $\mathrm{var}(\hat{\gamma}) \geq \gamma^2$, see also Rao (1971). So, we can improve the precision of our estimates by eliminating a variable provided that the absolute value of the corresponding coefficient is small compared with the standard deviation of the LSE $\hat{\gamma}$.

Example: Consider the mean-shift model investigated by Beckmann and Cook (1983), Cook and Weisberg (1982) and Boscher (1991), where $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$ and $\mathbf{Z} = (0, \mathbf{I}_k)'$ with $\mathbf{X}_1$ of type $(n-k) \times (p-k)$ being of full column rank and $\mathbf{X}_2$ of type $k \times (p-k)$. Then the "never-pool" estimator $\hat{\boldsymbol{\beta}}_* = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1$, where $\mathbf{y}_1$ results from $\mathbf{y}$ by deleting the last k observations, coincides with $\bar{\boldsymbol{\beta}}$ from (2.3).

Boscher (1991, Theorem 1) derived necessary and sufficient conditions for the superiority of $\hat{\beta}$ over $\hat{\beta}_* = \bar{\beta}$. We give an alternative criterion:

The least squares estimator $\hat{\beta} = \mathbf{X}^+\mathbf{y}$ which drops the matrix $\mathbf{Z}$ indicating the shift, dominates the least squares estimator $\bar{\beta} = \hat{\beta}_* = \mathbf{X}_1^+\mathbf{y}_1$ if and only if

$$\gamma'(\mathbf{I} - \mathbf{V}_2)\mathbf{X}_2[\mathbf{X}_2'(\mathbf{I} - \mathbf{V}_2)\mathbf{X}_2]^+\mathbf{X}_2'(\mathbf{I} - \mathbf{V}_2)\gamma \leq \sigma^2 \ , \tag{2.13}$$

where $\mathbf{V}_2 = \mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_2'$ is located in the southeast corner of $\mathbf{P}_X$. This follows immediately from Theorem 1 and the identities $\mathbf{G} = -\mathbf{Q}_X\mathbf{P}_Z\mathbf{X}$, $\mathbf{P}_Z = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{I}_k \end{pmatrix}$, $\mathbf{P}_Z\mathbf{Q}_X\mathbf{P}_Z = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{I} - \mathbf{V}_2 \end{pmatrix}$ and $\mathbf{G}'\mathbf{Z}\gamma = -\mathbf{X}_2'(\mathbf{I} - \mathbf{V}_2)\gamma$.

Note that in the mean shift model $\operatorname{cov}(\hat{y}) = \sigma^2(\mathbf{I} - \mathbf{V}_2)$, and for $k = 1$ we have $\operatorname{var}(\hat{y}) = \sigma^2(1 - v_n)$, where $v_n = \mathbf{x}_n'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_n$. If $\operatorname{var}(\hat{y}) > \gamma^2$, then one should consider to ignore the level shift. However, since σ^2 and γ are unknown, we need a test for detecting the unknown 'state of nature'.

Condition (2.12) can be simplified if $\mathbf{X}$ or $\mathbf{Z}$ is a column vector: If $\mathbf{X} = \mathbf{x}$ we may write $\mathbf{G} = \mathbf{g} = \mathbf{Q}_X\mathbf{Q}_Z\mathbf{x}$ and $\mathbf{G}\mathbf{G}^+ = \mathbf{g}\mathbf{g}'/\mathbf{g}'\mathbf{g}$. Hence (2.12) becomes $(\mathbf{g}'\mathbf{Z}\gamma)^2 \leq \sigma^2\mathbf{g}'\mathbf{g}$. If $\mathbf{Z} = \mathbf{z}$, $\mathbf{z}'\mathbf{P}_G\mathbf{z}$ is a scalar, and (2.12) turns out to be $(\mathbf{z}'\mathbf{P}_G\mathbf{z})\gamma'\gamma \leq \sigma^2$.

Let us now investigate the situation when $\bar{\beta}$ dominates $\hat{\beta}$.

Theorem 2: The least squares estimator $\bar{\beta} = (\mathbf{X}'\mathbf{Q}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}_Z\mathbf{y}$ dominates the least squares estimator $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ which drops the observations from $\mathbf{Z}$ with respect to MMSE if and only if $\mathfrak{R}(\mathbf{A}) \subset \mathfrak{R}(\mathbf{a})$ and $\mathbf{a}'\mathbf{A}\mathbf{a} \leq (\mathbf{a}'\mathbf{a})^2$ where $\mathbf{A} = \sigma^2\mathbf{X}'\mathbf{Q}_Z\mathbf{Q}_X\mathbf{Q}_Z\mathbf{X}$ and $\mathbf{a} = -\mathbf{X}'\mathbf{Q}_Z\mathbf{Q}_X\mathbf{Z}\gamma$.

Proof: By (2.10) and (2.11) $\bar{\beta}$ is superior to $\hat{\beta}$ iff $\mathbf{a}\mathbf{a}' - \mathbf{A}$ is n.n.d. From our lemma in the appendix this is the case iff $\mathfrak{R}(\mathbf{A}) \subset \mathfrak{R}(\mathbf{a}\mathbf{a}') = \mathfrak{R}(\mathbf{a})$ and $\lambda_1((\mathbf{a}\mathbf{a}')^+\mathbf{A}) \leq 1$. However, since $(\mathbf{a}\mathbf{a}')^+ = \mathbf{a}\mathbf{a}'/(\mathbf{a}'\mathbf{a})^2$ and $\lambda_1(\mathbf{a}\mathbf{a}'\mathbf{A}) = \mathbf{a}'\mathbf{A}\mathbf{a}$ we obtain the asserted equivalence. ■

Since $\mathfrak{R}(\mathbf{A}) \subset \mathfrak{R}(\mathbf{a})$ is equivalent to $\mathbf{A} = \mathbf{a}\mathbf{b}'$ for some vector $\mathbf{b}$ we can state that $\bar{\beta}$ dominates $\hat{\beta}$ if and only if $\mathbf{A} = \mathbf{a}\mathbf{b}'$ and $\mathbf{b}'\mathbf{a} \leq \mathbf{a}'\mathbf{a}$ for some vector $\mathbf{b}$.

Again the case $\mathbf{X} = \mathbf{x}$ or $\mathbf{Z} = \mathbf{z}$ deserves special attention. If $\mathbf{X} = \mathbf{x}$, then $\mathbf{a} = -\mathbf{x}'\mathbf{Q}_Z\mathbf{Q}_X\mathbf{Z}\gamma$ and $\mathbf{A} = \sigma^2\mathbf{x}'\mathbf{Q}_Z\mathbf{Q}_X\mathbf{Q}_Z\mathbf{x}$ are both scalars. Let now $\mathbf{a} \neq 0$. Then we may choose $\mathbf{b} = \mathbf{b}' = \mathbf{A}/\mathbf{a}$ and $\mathbf{b}'\mathbf{a} \leq \mathbf{a}'\mathbf{a}$ is equivalent to

$$\sigma^2\mathbf{x}'\mathbf{Q}_Z\mathbf{Q}_X\mathbf{Q}_Z\mathbf{x} \leq (\mathbf{x}'\mathbf{Q}_Z\mathbf{Q}_X\mathbf{Z}\gamma)^2 \ . \tag{2.14}$$

If $\mathbf{a} = 0$ we must have $\mathbf{A} = 0$ or, equivalently $\mathbf{P}_Z\mathbf{x} \in \mathfrak{R}(\mathbf{X})$. Since, vice versa, $\mathbf{P}_Z\mathbf{x} \in \mathfrak{R}(\mathbf{X})$ also implies $\mathbf{a} = 0$ we may say that $\bar{\beta}$ dominates $\hat{\beta}$ if either $\mathbf{a} \neq 0$ and (2.14) holds or $\mathbf{P}_Z\mathbf{x} \in \mathfrak{R}(\mathbf{X})$.

If $\mathbf{Z} = \mathbf{z}$, then γ is a scalar, $\mathbf{a} = \gamma \mathbf{X}'\mathbf{z}(\mathbf{z}'\mathbf{Q}_X\mathbf{z}/\mathbf{z}'\mathbf{z})$, $\mathbf{A} = \mathbf{X}'\mathbf{z}\mathbf{z}'\mathbf{X}$ $(\sigma^2\mathbf{z}'\mathbf{Q}_X\mathbf{z}/(\mathbf{z}'\mathbf{z})^2)$ and $\mathbf{b} = \mathbf{X}'\mathbf{z}\sigma^2/\gamma\mathbf{z}'\mathbf{z}$, provided that $\gamma \neq 0$. Then $\mathbf{b}'\mathbf{a} \leq \mathbf{a}'\mathbf{a}$ is equivalent to

$$\sigma^2 \leq \gamma^2\mathbf{z}'\mathbf{Q}_X\mathbf{z} \tag{2.15}$$

(cf. Boscher, 1991, formula (6)). If $\gamma = 0$, we must have $\mathbf{A} = 0$ which is satisfied iff $\mathbf{P}_X\mathbf{z} = 0$ or $\mathbf{P}_X\mathbf{z} = \mathbf{z}$. Since $\operatorname{rank}(\mathbf{W}) = \operatorname{rank}(\mathbf{X}, \mathbf{Z}) = p$, $\mathbf{P}_X\mathbf{z} = \mathbf{z}$ can be excluded for $\mathbf{z} \neq 0$. Hence $\bar{\boldsymbol{\beta}}$ dominates $\hat{\boldsymbol{\beta}}$ if either $\gamma \neq 0$ and (2.15) holds or $\gamma = 0$ and $\mathbf{P}_X\mathbf{z} = 0$ holds.

3 A Test for the MMSE Improvement

Let us assume that $\mathbf{u}$ follows a normal distribution, i.e. $\mathbf{u} \sim N(0, \sigma^2\mathbf{I})$. Then the quadratic form $\hat{\gamma}'\mathbf{Z}'\mathbf{P}_G\mathbf{Z}\hat{\gamma}/\sigma^2 = \mathbf{y}'\mathbf{P}_G\mathbf{y}/\sigma^2$ has a noncentral χ^2-distribution with noncentrality parameter $\lambda = \gamma'\mathbf{Z}'\mathbf{P}_G\mathbf{Z}\gamma/2\sigma^2$ and with $r = \operatorname{tr}[\mathbf{P}_G]$ degrees of freedom, where $\hat{\gamma} = (\mathbf{Z}'\mathbf{Q}_X\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Q}_X\mathbf{y}$. If we put $\mathbf{Q} = \mathbf{I} - \mathbf{P}$ where $\mathbf{P} = \mathbf{Q}_X\mathbf{Z}(\mathbf{Z}'\mathbf{Q}_X\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Q}_X$ is the orthogonal projector on $\mathfrak{R}(\mathbf{Q}_X\mathbf{Z})$, the quadratic form $\mathbf{y}'\mathbf{Q}_X\mathbf{Q}\mathbf{Q}_X\mathbf{y}/\sigma^2$ follows a central χ^2-distribution with $\operatorname{tr}(\mathbf{Q}_X\mathbf{Q}\mathbf{Q}_X) = \operatorname{tr}(\mathbf{Q}_X - \mathbf{P}) = \operatorname{tr}(\mathbf{I} - \mathbf{P}_X - \mathbf{P}) = n - p_1 - p_2$ degrees of freedom. Since $\mathbf{Q}_X\mathbf{P}_G = \mathbf{P}_G$ and $\mathbf{P}\mathbf{P}_G = \mathbf{P}_G$ (the latter identity follows since $\mathfrak{R}(\mathbf{Q}_X\mathbf{Q}_Z\mathbf{X}) = \mathfrak{R}(\mathbf{Q}_X\mathbf{P}_Z\mathbf{X}) \subset \mathfrak{R}(\mathbf{Q}_X\mathbf{Z})$) we obtain $\mathbf{Q}\mathbf{Q}_X\mathbf{P}_G = 0$, and thus we may state

Theorem 3: The statistic

$$F = \frac{\mathbf{y}'\mathbf{P}_G\mathbf{y}/r}{\mathbf{y}'\mathbf{Q}_X\mathbf{Q}\mathbf{Q}_X\mathbf{y}/(n - p_1 - p_2)} \tag{3.1}$$

follows a noncentral F-distribution with r and $n - p_1 - p_2$ degrees of freedom and with $\lambda = \gamma'\mathbf{Z}'\mathbf{P}_G\mathbf{Z}\gamma/2\sigma^2$ as noncentrality parameter.

Since $\delta = 2\lambda$ the hypotheses H_0: $\delta \leq 1$ vs. H_1: $\delta > 1$ can now be tested by using the F-test of Toro-Vizcarrondo and Wallace (1968). This test suggests deleting the observation matrix $\mathbf{Z}$ at level α if $\mathbf{F} \leq \mathbf{F}_{1-\alpha}(r, n - p_1 - p_2; 1/2)$, where $\mathbf{F}_{1-\alpha}(r, n - p_1 - p_2; 1/2)$ is the $(1 - \alpha)$-quantile of the noncentral F-distribution with r and $n - p_1 - p_2$ degrees of freedom, and with noncentrality parameter 1/2. This test results in a pre-test estimator (cf. Judge and Bock, 1978): Choose $\hat{\boldsymbol{\beta}}$ if H_0 is accepted, and $\bar{\boldsymbol{\beta}}$ otherwise. The pre-test estimator may be written as a weighted sum of $\hat{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\beta}}$, where the weights are indicator functions. The hypothesis and the level of the test determine the weights to be

allocated to $\hat{\beta}$ and $\bar{\beta}$. General properties of pre-test estimators are discussed in Judge and Bock (1978, Chapters 3–5). The MMSE comparisons of the above pre-test estimator with $\hat{\beta}$ and $\bar{\beta}$ can be done by applying straightforwardly the results of Judge and Bock (1978, Chapter 5) in the same way as in Boscher (1991, Theorem 2).

Appendix

Lemma: Let $\mathbf{A}$ and $\mathbf{B}$ be n.n.d. matrices. Then $\mathbf{A} \leq \mathbf{B}$ iff $\mathfrak{R}(\mathbf{A}) \subset \mathfrak{R}(\mathbf{B})$ and $\lambda_1(\mathbf{B}^+\mathbf{A}) \leq 1$.

This result was introduced by Stepniak (1985, Theorem 1). Various proofs of the lemma can be found in Baksalary, Liski and Trenkler (1988, Theorem 2), Liski and Puntanen (1989, Lemma) and in Baksalary, Schipp and Trenkler (1991).

The authors wish to express their gratitude to the referee for his insightful suggestions.

References

Baksalary JK, Liski EP, Trenkler G (1989) Mean square error matrix improvements and admissibility of linear estimators. *Journal of Statistical Planning and Inference* 23:313–325

Baksalary JK, Schipp B, Trenkler G (1991) Some Further Results on Hermitian-Matrix Inequalities, accepted for publication. In: *Linear Algebra and its Applications*

Beckmann DA, Cook RD (1983) Outlier ... s. *Technometrics* 25:118–163

Boscher H (1991) Contaminations in Linear Regression Models and its Influence on Estimators. *Statistica Neerlandica* 45:9–19

Cook RD, Weisberg S (1982) *Residuals and Influence in Regression.* Chapman and Hall, London

Hocking RR (1974) Misspecification in Regression. *The American Statistician* 28:39–40

Judge GG, Bock ME (1978) *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, North-Holland, Amsterdam

Judge GG, Griffiths WE, Hill RC, Lütkepohl H, Lee T (1985) *The Theory and Practice of Econometrics*, John Wiley, New York

Liski EP, Puntanen S (1989) A Further Note on a Theorem on the Difference of the Generalized Inverses of two Nonnegative Definite Matrices. *Communications in Statistics–Theory Methods* 18(5):1747–1751

Rao P (1971) Some Notes on Misspecification in Multiple Regression. *The American Statistican* 25:37–39

Stepniak C (1985) Ordering of nonnegative definite matrices with application to comparison of linear models. *Linear Algebra and Its Applications* 70:67–71

Toro-Vizcarrondo C, Wallace TD (1968) A Test of the Mean Square Error Criterion for Restrictions in Linear Regression. *Journal of the American Statistical Association* 63:558–572

Received 24.09.1992
Revised version 11.01.1993

Book Review

D. A. Williams: Probability with martingales, Cambridge University Press, 1991, XV/251 pp., $59.50/£10.95

The textbook under review was based on lecture notes for a third-year undergraduate course at Cambridge and serves as an introduction to measure-theoretic probability with as much probability and as little measure theory as possible. At the same time, the author tries to make the exposition as rigorous as necessary for beginners.

As motivation for an analytically serious approach to probability theory, the author discusses a very special branching process (from scratch) and points out explicitly where his first martingale appears. Martingales as very useful stochastic processes then receive priority in the present introduction. They fill the main (and middle) part B of the book, which is framed by part A concerning the foundations including product measures and part C discussing Fourier transformation and central limit theorem. In part B, conditional expectations are introduced, then the usual Doob's 'Forward' convergence theorem is proved; chapters on L^2-bounded and uniformly integrable martingales follow; a rich collection of applications (to option pricing, the sheep problem, noisy observations, and harnesses) appears as a sort of culmination to the book.

Clearly, incorporating that much material on approximately 250 pages, with an advanced emphasis on a central theme of stochastic processes, makes it difficult to accomplish consequent and complete proving. The numerous appendices at the end of the book (and also some of the exercises supplementing the text) demonstrate this dilemma. In particular, the existence of an infinite product of probability measures and hence the existence of an infinite sequence of independent random variables having prescribed distributions (the topic of the Appendix to Chap. 9) remains difficult to verify unless the "keen student who has read all previous appendices studies it with a tutor" (slightly modified quotation).

But despite those well-known problems any author of an introduction to probability has to face, the book under review is most original in setup and style and will definitely enjoy strictly positive reception. The reviewer himself has decided to test the proposed approach in one of his introductory courses, since he agrees with the author who produced "as lively an introduction as he could manage".

Tübingen H. Heyer

Metrika (1993) 40:271–281

Metrika

Testing Homogeneity of Ordered Variances

GOVIND S. MUDHOLKAR, MICHAEL P. MCDERMOTT and JOHANNE AUMONT
Department of Statistics, University of Rochester, Rochester, NY 14627, USA

Summary: Fujino (1979) studied several tests for homogeneity of nondecreasing variances and concluded that the modification of Bartlett's (1937) test, first proposed by Boswell & Brunk (1969), is generally superior to its competitors in terms of power. A weakness of this test, however, is that the null distribution of the test statistic has not been adequately determined for cases other than when the group sample sizes are equal. In this article a class of simple tests for equality of nondecreasing variances is proposed which can be used without special tables for arbitrary sample sizes. Some of these tests have operating characteristics which compare favorably to those of the modification of Bartlett's test. A prescription is also given for applying the tests in cases where the population variances are constrained by more general partial orders.

Key Words: Bartlett's test; Combining independent tests; Fisher's Combination method; Logit Combination method; Order-restricted inference.

1 Introduction

The assumption of equal error variances underlies many inferential procedures in the normal theory analysis of linear models. The well-known homoscedasticity tests of Bartlett (1937), Cochran (1941), and Hartley (1950) provide appropriate type I error control even if the population variances are *a priori* order-restricted, however their powers would be expected to be smaller than those of procedures which utilize the prior information about the ordering. In this spirit Fujino (1979) suggested modifying the classical tests by using the order-constrained maximum likelihood estimates of the variances in the test statistics in place of the sample variances. He considered the case where the variances are known to be in nondecreasing order and the sample sizes are equal. In an empirical study of the classical tests and their modifications, as well as a regression-type test due to Vincent (1961) and its modification, Fujino found that the modified tests have a substantial power advantage over the classical tests, and that the modification of Bartlett's test is preferable in this regard.

In this paper we employ the general strategy for order-constrained hypothesis testing proposed in Mudholkar & McDermott (1989) for constructing simple alternatives to the modified tests investigated by Fujino. In this approach the

0026–1335/93/5/271–281 $2.50

null hypothesis is decomposed into several nested component hypotheses, each of which can be tested using simple, well-known statistics. It can be easily shown that these statistics, and therefore the associated p-values, are mutually independent. These p-values are then pooled using such classical devices as Fisher's combination method to provide an overall test of the null hypothesis.

The modified versions of the classical tests of homoscedasticity proposed by Fujino (1979) are outlined in Section 2. In Section 3 the new tests are described for the case where the population variances are known to be in nondecreasing order. These tests are easy to implement for arbitrary sample sizes. A power study is undertaken in Section 4, using both exact computation and simulation, which compares the new tests with the modification of Bartlett's test. The new tests for homoscedasticity are extended in Section 5 to cases where the population variances are constrained by more general partial orders. The conclusions are summarized in Section 6.

2 Modifications of the Classical Tests

Let $\hat{\sigma}_i^2 = S_i^2$ $(i = 1, \ldots, k)$ be the unbiased estimates of the variances σ_i^2 of k normal populations, where the $\nu_i S_i^2/\sigma_i^2$ are independently distributed as χ^2 with ν_i degrees of freedom. It is of interest to test H_0: $\sigma_1^2 = \cdots = \sigma_k^2$ against the simple order alternative H_1: $\sigma_1^2 \leq \cdots \leq \sigma_k^2$, with at least one inequality strict.

Fujino (1979) assumed $\nu_1 = \cdots = \nu_k = \nu$ and developed the modifications M^*, $F_{\max}^*$, and G^*, G_* of the classical tests M due to Bartlett (1937), Hartley's (1950) $F_{\max}$, and G proposed by Cochran (1941) respectively for testing H_0 against H_1. These modifications were obtained by replacing the unrestricted maximum likelihood estimates of σ_i^2 in the classical statistics by the order-constrained maximum likelihood estimates $\hat{\sigma}_i^{*2}$:

$$M^* = k\nu \log \hat{\sigma}^2 - \nu \sum_{i=1}^{k} \log \hat{\sigma}_i^{*2} , \tag{2.1}$$

where $\hat{\sigma}^2$ is the maximum likelihood estimate of the common value σ^2 under H_0,

$$F_{\max}^* = \max_{1 \leq i \leq k} \hat{\sigma}_i^{*2} \Big/ \min_{1 \leq i \leq k} \hat{\sigma}_i^{*2} , \tag{2.2}$$

$$G^* = \max_{1 \leq i \leq k} \hat{\sigma}_i^{*2} \Big/ \sum_{i=1}^{k} \hat{\sigma}_i^{*2} , \qquad G_* = \min_{1 \leq i \leq k} \hat{\sigma}_i^{*2} \Big/ \sum_{i=1}^{k} \hat{\sigma}_i^{*2} . \tag{2.3}$$

Note that several algorithms are available for computing the constrained estimates $\hat{\sigma}_i^{*2}$ $(i = 1, \ldots, k)$; see Robertson, Wright & Dykstra (1988, Ch. 1) for details.

Boswell & Brunk (1969) first proposed M^* as a statistic for this problem and also obtained its large sample null distribution. Fujino (1979) tabulated the null distributions of M^* and $F^*_{\max}$. Fujino also performed a simulation study in order to evaluate the power properties of the above tests as well as a "regression type" test to due Vincent (1961), which is based on the statistic

$$V = \sum_{i=1}^{k} \{i - \tfrac{1}{2}(k+1)\}\hat{\sigma}_i^2 \Big/ \sum_{i=1}^{k} \hat{\sigma}_i^2 \ , \tag{2.4}$$

and a modified version based on the statistic

$$V' = \sum_{i=1}^{k} \{i - \tfrac{1}{2}(k+1)\} \log \hat{\sigma}_i^2 \ . \tag{2.5}$$

The test based on M^* was found to have the best overall performance in this investigation.

3 A New Class of Tests

Following the approach given in Mudholkar & McDermott (1989), one may view the problem of testing H_0: $\sigma_1^2 = \cdots = \sigma_k^2$ subject to the simple order constraint $\sigma_1^2 \leq \cdots \leq \sigma_k^2$ as the conjunction of $k-1$ nested problems of testing $H_{0(i)}$: $\sigma_1^2 = \cdots = \sigma_{i-1}^2 = \sigma_i^2$ against the alternative $H_{1(i)}$: $\sigma_1^2 = \cdots = \sigma_{i-1}^2 < \sigma_i^2$ for $i = 2, \ldots, k$. To ease notation, let

$$S^2_{[i-1]} = \sum_{j=1}^{i-1} \nu_j S_j^2 \Big/ \sum_{j=1}^{i-1} \nu_j \ ,$$

$$\nu_{[i-1]} = \sum_{j=1}^{i-1} \nu_j \qquad (i = 2, \ldots, k) \ . \tag{3.1}$$

It is well-known that the uniformly most powerful unbiased test of $H_{0(i)}$ against $H_{1(i)}$ is based on the statistic $F_i = S_i^2 / S^2_{[i-1]}$. The construction of the new tests of H_0 against H_1 depends on the following result:

Theorem 1: Under the null hypothesis the $(k-1)$ *test statistics* F_i *are mutually independently distributed as* F *with* ν_i *and* $\nu_{[i-1]}$ *degrees of freedom* $(i = 2, \ldots, k)$.

The theorem follows immediately from the following lemma:

Lemma 1: Let $V_1, \ldots, V_n$ *be independent gamma random variables with the same scale parameter, i.e.*

$$f_{V_i}(v_i) = \frac{\beta^{\alpha_i}}{\Gamma(\alpha_i)} v_i^{\alpha_i - 1} e^{-\beta v_i} , \qquad v_i > 0$$

for $i = 1, \ldots, n$. *Then the random variables*

$$W_2 = \frac{V_2}{V_1} , \qquad W_3 = \frac{V_3}{V_1 + V_2} , \ldots, W_n = \frac{V_n}{V_1 + V_2 + \cdots + V_{n-1}}$$

are mutually independent.

Proof: Let $V_1 = W_1$. It follows that

$$V_1 = W_1 , \qquad V_2 = W_1 W_2 , \qquad V_3 = W_1(1 + W_2)W_3, \ldots,$$

$$V_n = W_1(1 + W_2)\ldots(1 + W_{n-1})W_n .$$

The Jacobian matrix

$$\mathbf{J} = \left(\frac{\partial V_j}{\partial W_i}\right) , \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n ,$$

is upper-triangular, therefore

$$|\mathbf{J}| = \prod_{i=1}^{n} \frac{\partial V_i}{\partial W_i}$$

$$= W_1(W_1(1 + W_2))\cdots(W_1(1 + W_2)\cdots(1 + W_{n-1}))$$

$$= W_1^{n-1} \prod_{i=2}^{n-1} (1 + W_i)^{n-i} .$$

The probability density function of $\mathbf{W} = (W_1, \ldots, W_n)'$ is then easily shown to be

$$g_{\mathbf{W}}(\mathbf{w}) \propto w_1^{\sum_{i=1}^n \alpha_i - 1} \exp\left\{-\beta w_1 \prod_{i=2}^n (1 + w_i)\right\} \prod_{i=2}^n w_i^{\alpha_i - 1}(1 + w_i)^{\sum_{j=i+1}^n \alpha_j} .$$

Hence the joint density of $\mathbf{W}^* = (W_2, \ldots, W_n)'$, obtained by integrating out W_1, is

$$h_{\mathbf{W}^*}(\mathbf{w}^*) \propto \prod_{i=2}^n w_i^{\alpha_i - 1}(1 + w_i)^{\sum_{j=i+1}^n \alpha_j} \int_0^\infty w_1^{\sum_{i=1}^n \alpha_i - 1} \exp\left\{-\beta w_1 \prod_{i=2}^n (1 + w_i)\right\} dw_1$$

$$\propto \prod_{i=2}^n w_i^{\alpha_i - 1}(1 + w_i)^{-\sum_{j=1}^i \alpha_j} .$$

This is clearly the product of the individual densities of $W_2, \ldots, W_n$. Therefore $W_2, \ldots, W_n$ are mutually independent.

Now let P_i be the p-values associated with the test statistics F_i, which by Theorem 1 are mutually independent. The new tests of H_0 subject to the simple order constraint are based on various methods of combining independent p-values. In this article we study four such tests based on the combination statistics $\Psi_T = \min(P_i)$ due to Tippett, $\Psi_F = -2\sum \log P_i$ introduced by Fisher, Liptak's $\Psi_N = \sum \Phi^{-1}(1 - P_i)$, and the logit statistic $\Psi_L = -A^{-1/2}\sum \log\{P_i/(1 - P_i)\}$ proposed by Mudholkar & George (1979), where $A = \pi^2 m(5m + 2)/(15m + 12)$ and $m = k - 1$ is the number of p-values being combined. Small values of Ψ_T and large values of Ψ_F, Ψ_N, and Ψ_L are seen as evidence against the null hypothesis. Under H_0 Ψ_T is distributed as the minimum of m uniform variates, Ψ_F has a χ^2 distribution with $2m$ degrees of freedom, Ψ_N has a $N(0, m)$ distribution, and Ψ_L has a distribution that is very well approximated by Student's t with $5m + 4$ degrees of freedom. Note that Ψ_F and Ψ_L are asymptotically equivalent and optimal among all monotone combination methods, and in some cases among all tests based on the data, in terms of Bahadur's exact slopes; for example see Berk & Cohen (1979).

4 Power Comparisons

Consider the simple case where $k = 3$ and the degrees of freedom are all equal, i.e. $\nu_1 = \nu_2 = \nu_3 = \nu$. To calculate the power of the new tests for H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$ against the alternative $\sigma_1^2 \leq \sigma_2^2 \leq \sigma_3^2$, with at least one inequality strict, it is necessary to examine the distributions of the component p-values under the alternative hypothesis.

Consider the random variables

$$F_2' = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2}, \qquad F_3' = \frac{2S_3^2/\sigma_3^2}{(S_1^2/\sigma_1^2) + (S_2^2/\sigma_2^2)},$$

where σ_1^2, σ_2^2, and σ_3^2 cancel under the null hypothesis. Clearly these random variables may be expressed as $F_2' = \mathscr{F}_{\nu,\nu}^{-1}(U_1)$ and $F_3' = \mathscr{F}_{\nu,2\nu}^{-1}(U_2)$, where $\mathscr{F}_{\nu_1,\nu_2}(\cdot)$ is the cumulative distribution function of the F-distribution with ν_1 and ν_2 degrees of freedom and U_1 and U_2 are independent $U(0, 1)$ random variables. Hence, after some manipulation, it is easily seen that the random variables F_2 and F_3 satisfy

$$F_2 \doteq \frac{\sigma_2^2}{\sigma_1^2}\mathscr{F}_{\nu,\nu}^{-1}(U_1), \qquad F_3 \doteq \frac{\sigma_3^2\{1 + \mathscr{F}_{\nu,\nu}^{-1}(U_1)\}}{\sigma_1^2 + \sigma_2^2\mathscr{F}_{\nu,\nu}^{-1}(U_1)}\mathscr{F}_{\nu,2\nu}^{-1}(U_2),$$

where $\doteq$ denotes equivalence in law. Therefore the component p-values satisfy

$$P_2 \doteq 1 - \mathscr{F}_{\nu,\nu}\left\{\frac{\sigma_2^2}{\sigma_1^2}\mathscr{F}_{\nu,\nu}^{-1}(U_1)\right\}, \tag{4.1}$$

$$P_3 \doteq 1 - \mathscr{F}_{\nu,2\nu}\left[\frac{\sigma_3^2\{1 + \mathscr{F}_{\nu,\nu}^{-1}(U_1)\}}{\sigma_1^2 + \sigma_2^2\mathscr{F}_{\nu,\nu}^{-1}(U_1)}\mathscr{F}_{\nu,2\nu}^{-1}(U_2)\right]. \tag{4.2}$$

The power function of the new test based on $\Psi_F = -2\log P_2 + -2\log P_3$ may be found directly using (4.1) and (4.2). Note however that this power function is not a simple convolution of $-2\log P_2$ and $-2\log P_3$ because, in general, the component p-values are not mutually independent under the alternative hypothesis. However because P_2 is a function of U_1 alone, the power function may be obtained by conditioning on U_1:

$$\begin{aligned} pr(\Psi_F \geq C) &= pr(-2\log P_2 + -2\log P_3 \geq C) \\ &= pr(P_3 \leq e^{-C/2}/P_2) \\ &= E\{pr(P_3 \leq e^{-C/2}/P_2)|U_1\}, \end{aligned}$$

where C is the upper $100\alpha\%$ point of the χ_4^2 distribution. Expressions (4.1) and (4.2) may be substituted above for P_2 and P_3 respectively, leading to the power function

$$1 - \int_0^y \mathscr{F}_{\nu,2\nu}\left(\frac{\sigma_1^2 + \sigma_2^2\mathscr{F}_{\nu,\nu}^{-1}(u)}{\sigma_3^2\{1 + \mathscr{F}_{\nu,\nu}^{-1}(u)\}}\mathscr{F}_{\nu,2\nu}^{-1}\left[1 - \frac{e^{-C/2}}{1 - \mathscr{F}_{\nu,\nu}\{\mathscr{F}_{\nu,\nu}^{-1}(u)\sigma_2^2/\sigma_1^2\}}\right]\right)du, \tag{4.3}$$

where

$$y = \mathcal{F}_{\nu,\nu}\{\mathcal{F}_{\nu,\nu}^{-1}(1 - e^{-C/2})\sigma_1^2/\sigma_2^2\} .$$

The power functions of Ψ_T, Ψ_N, and Ψ_L may be obtained with some effort in a similar fashion.

The power function (4.3) of Ψ_F and analogous expressions for Ψ_T, Ψ_N, and Ψ_L were numerically evaluated using NAG (1981) fortran library subroutines D01AJF and D01AHF for quadrature. The significance level was taken to be 5% and values of $\nu = 5$, 10 were used. The power function of M^* was estimated using Monte-Carlo simulation. NAG fortran library subroutines G05CCF and G05DDF were used to generate $n_i = \nu + 1$ independent normal random variables with mean 0 and variance σ_i^2, $i = 1, 2, 3$, and using these M^* was calculated. The powers were estimated by calculating the percentage of times out of 100,000 repetitions that M^* exceeded the appropriate critical value. The results of the comparisons of these five power functions are presented in Table 1(a). Various configurations $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ of the variances were used and these are listed in Table 1(b).

Table 1(a). Powers at the 5% level of five competing tests for H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$ in the model $\sigma_1^2 \le \sigma_2^2 \le \sigma_3^2$

Configuration	ν	M^*	Ψ_F	Ψ_L	Ψ_N	Ψ_T
Null	5	0.050	0.050	0.050	0.050	0.050
	10	0.050	0.050	0.050	0.050	0.050
Step (1, 2)	5	0.404	0.416	0.411	0.396	0.369
	10	0.710	0.729	0.720	0.691	0.656
Step (2, 3)	5	0.454	0.434	0.391	0.348	0.453
	10	0.737	0.702	0.640	0.565	0.728
Linear	5	0.378	0.392	0.402	0.397	0.325
	10	0.644	0.664	0.679	0.674	0.556
Quadratic	5	0.378	0.392	0.401	0.395	0.328
	10	0.640	0.658	0.672	0.666	0.556
Logarithm	5	0.380	0.395	0.404	0.399	0.329
	10	0.654	0.677	0.689	0.682	0.570

Powers are based on 100,000 simulated replications for M^* and are exact for Ψ_F, Ψ_L, Ψ_N, and Ψ_T

Table 1(b). Configurations $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$

Null	(1.00, 1.00, 1.00)
Step (1, 2)	(1.00, 4.00, 4.00)
Step (2, 3)	(1.00, 1.00, 4.00)
Linear	(1.00, 2.50, 4.00)
Quadratic	(1.00, 2.25, 4.00)
Logarithm	(1.00, 2.89, 4.00)

Table 2(a). Estimated powers at the 5% level of five competing tests for H_0: $\sigma_1^2 = \cdots = \sigma_6^2$ in the model $\sigma_1^2 \leq \cdots \leq \sigma_6^2$

Configuration	ν	M^*	Ψ_F	Ψ_L	Ψ_N	Ψ_T
Null	5	0.050	0.050	0.051	0.050	0.051
	10	0.050	0.050	0.050	0.050	0.051
Step (1, 2)	5	0.367	0.375	0.359	0.344	0.282
	10	0.694	0.706	0.661	0.616	0.540
Step (3, 4)	5	0.707	0.699	0.628	0.571	0.609
	10	0.949	0.943	0.900	0.847	0.882
Step (5, 6)	5	0.501	0.415	0.331	0.259	0.481
	10	0.776	0.673	0.553	0.411	0.752
Linear	5	0.460	0.490	0.494	0.483	0.315
	10	0.742	0.773	0.790	0.781	0.508
Quadratic	5	0.497	0.512	0.504	0.485	0.357
	10	0.777	0.782	0.782	0.764	0.574
Logarithm	5	0.418	0.462	0.471	0.460	0.282
	10	0.707	0.758	0.774	0.767	0.470

Table 2(b). Configurations $(\sigma_1^2, \ldots, \sigma_6^2)$

Null	(1.00, 1.00, 1.00, 1.00, 1.00, 1.00)
Step (1, 2)	(1.00, 4.00, 4.00, 4.00, 4.00, 4.00)
Step (2, 3)	(1.00, 1.00, 1.00, 4.00, 4.00, 4.00)
Step (5, 6)	(1.00, 1.00, 1.00, 1.00, 1.00, 4.00)
Linear	(1.00, 1.60, 2.20, 2.80, 3.40, 4.00)
Quadratic	(1.00, 1.26, 1.69, 2.29, 3.06, 4.00)
Logarithm	(1.00, 2.16, 2.84, 3.32, 3.69, 4.00)

A second simulation study was done in order to compare the five power functions for $k = 6$, again at the 5% level of significance and for $\nu = 5, 10$. The simulation was performed as described above, except that $N = 50{,}000$ repetitions were used for this study. The results and the configurations $(\sigma_1^2, \ldots, \sigma_6^2)$ of the variances employed are listed in Tables 2(a) and 2(b) respectively. These configurations are taken to be the same as in Fujino (1979) to facilitate comparison with his results.

For most of the configurations considered, the test based on Fisher's combination method is superior to the modification of Bartlett's test. As seen in Mudholkar & McDermott (1989) the tests based on the Logit and Liptak combination methods perform quite well for some configurations but quite poorly for others. The test based on Tippett's combination method is generally unsatisfactory.

It should be noted that the application of these combination methods in this setting is not invariant. Indeed one could proceed by first testing equality of any two adjacent variances, then one could test equality of their common value with an adjacent variance, and continue in this fashion. Clearly the power of the overall test will depend on the way in which the procedure is carried out. This

in part explains why the tests based on combination methods outperform the modification of Bartlett's test for some configurations, but not for others.

5 Applications to Other Partial Orders

The new tests proposed in this article can be extended to cases where the variances are constrained by partial orders other than simple order. As an illustration of the necessary notation, consider the constraints

$$\sigma_1^2 \leq \sigma_2^2 \ ; \qquad \sigma_3^2 \leq [\sigma_4^2, \sigma_5^2, \sigma_6^2] \leq [\sigma_7^2, \sigma_8^2] \ .$$

Here the eight variances are divided into two *blocks*, where the separation is indicated by a semicolon, and it is understood that there are no order restrictions among the variances in different blocks. The inclusion of variances within a bracket, such as $[\sigma_4^2, \sigma_5^2, \sigma_6^2]$, implies a lack of any order restrictions among them. More generally, the k variances will be divided into blocks such that there are no order restrictions among the variances in different blocks, but the variances within each block will be restricted in some manner.

To test the overall null hypothesis, first conduct the following steps for each block separately:

Step 1: Use Bartlett's test for testing equality of the variances included within a bracket. Obtain the corresponding significance probability for each bracket.
Step 2: Assume that all variances within each bracket are equal, thus yielding a simple order structure for the block. If there are, say, r inequalities in the block, obtain the $r - 1$ significance probabilities using the method described above for the case of simple order.

Having treated each of the blocks in this manner, assume equality within each block and use Bartlett's statistic to test equality of the variances between blocks. The component p-values involved in the procedure can be easily shown to be mutually independent. They can therefore be combined as described above in order to test equality of the variances.

In the above example, the new testing procedure would result in six component p-values arising from the following tests: (1) Bartlett's test for H_0: $\sigma_4^2 = \sigma_5^2 = \sigma_6^2$; (2) Bartlett's test for H_0: $\sigma_7^2 = \sigma_8^2$; (3) F-test for H_0: $\sigma_1^2 = \sigma_2^2$; (4) F-test for equality of σ_3^2 and the assumed common value of σ_4^2, σ_5^2, and σ_6^2; (5) F-test for equality of the common value of $\sigma_3^2, \ldots, \sigma_6^2$ and the common value of σ_7^2 and σ_8^2; (6) Bartlett's test for equality of the common value of σ_1^2 and σ_2^2 and the common value of $\sigma_3^2, \ldots, \sigma_8^2$.

The independence of the component test statistics relies on the following result (e.g. see Johnson and Kotz (1970)): if $X_1, \ldots, X_k$ are independent gamma random variables having the same scale parameter, then the two random variables $\sum X_i$ and $X_j/\sum X_i$ are independent for each j ($j = 1, \ldots, k$). As an illustration of the use of this result in proving independence consider the case where the variances are subject to the simple tree order, $\sigma_1^2 \leq [\sigma_2^2, \ldots, \sigma_k^2]$. The Bartlett statistic used for testing equality of $\sigma_2^2, \ldots, \sigma_k^2$ at the first stage is easily seen to be a simple function of

$$Q_1 = \prod_{i=2}^{k} \left\{ \sum_{j=2}^{k} \nu_j S_j^2 / S_i^2 \right\}^{\nu_i} .$$

The above property of gamma random variables can be applied directly to show that Q_1 is independent of the F-statistic used at the second stage, which is a simple function of

$$Q_2 = \sum_{i=2}^{k} \nu_i S_i^2 / S_1^2 .$$

A combination statistic Ψ may then be applied to the two associated component p-values in order to test equality of the k variances.

When testing equality of unrestricted variances in the above procedure, it is apparent that one may apply either Hartley's or Cochran's test instead of Bartlett's without sacrificing independence of the test statistics. This follows from the fact that, for independent gamma random variables having the same scale parameter, $\sum X_i$ will be independent of $\max X_i/\sum X_i$, $\min X_i/\sum X_i$, and their ratio $\max X_i/\min X_i$.

The new procedures may be extended further to more complex situations where the variances are subject to more general restrictions such as $\sigma_1^2 \leq [\sigma_3^2; \sigma_7^2; \sigma_2^2 \leq [\sigma_4^2, \sigma_5^2, \sigma_6^2]]$. Here groups of variances separated by semicolons are understood to have no order restrictions among them. In this case equality of the seven variances may be tested by first comparing $\sigma_2^2, \ldots, \sigma_7^2$ subject to the constraint to the right of σ_1^2, and then using an F-test to compare the common value of $\sigma_2^2, \ldots, \sigma_7^2$ with σ_1^2. An example of an ordering for which the new procedures may not be applied is $\sigma_1^2 \leq [\sigma_2^2, \sigma_4^2]$; $\sigma_3^2 \leq \sigma_4^2$, where σ_4^2 appears in both blocks. In general this new approach may be applied whenever a particular variance appears in one and only one block, not for general partial orders.

6 Conclusions

Until now satisfactory procedures for testing homogeneity of normal variances subject to order constraints have existed only for the simple order restriction and

equal sample sizes. The new approach to the problem outlined in this paper is easy to implement for arbitrary sample sizes and may be applied to a wide variety of order restrictions. The power comparisons show that, in particular, the proposed test based on Fisher's combination method compares favorably with the best of the existing tests, namely the modification of Bartlett's test.

Acknowledgement: The authors wish to thank an anonymous referee for some very helpful comments which substantially improved the manuscript.

References

Bartlett MS (1937) Properties of sufficiency and statistical tests. *Proc. Roy. Soc. A* 160:268–282

Berk RH, Cohen A (1979) Asymptotically optimal methods of combining tests. *J. Amer. Statist. Assoc.* 74:812–814

Boswell MT, Brunk HD (1969) Distribution of likelihood ratio in testing against trend. *Ann. Math. Statist.* 40:371–380

Cochran WG (1941) The distribution of the largest of a set of estimated variances as a fraction of their total. *Ann. Eugen.* 11:47–52

Fujino Y (1979) Tests for the homogeneity of a set of variances against ordered alternatives. *Biometrika* 66:133–139

Hartley HO (1950) The maximum *F*-ratio as a short-cut test for heterogeneity of variance. *Biometrika* 37:308–312

Johnson NL, Kotz S (1970) *Distributions in Statistics: Continuous Univariate Distributions*—1. New York: John Wiley

Mudholkar GS, George EO (1979) The logit statistic for combining probabilities—an overview. In *Optimizing Methods in Statistics*, ed. J. S. Rustagi, pp. 345–365. New York: Academic Press

Mudholkar GS, McDermott MP (1989) A class of tests for equality of ordered means. *Biometrika* 76:161–168

NAG (1981) *Numerical Algorithms Group Library, Mark 8.* Downers Grove, Illinois: Numerical Algorithms Group

Robertson T, Wright FT, Dykstra RL (1988) *Order Restricted Statistical Inference.* New York: John Wiley

Vincent SE (1961) A test of homogeneity for ordered variances. *J. Roy. Statist. Soc. Ser B* 23:195–206

Received 15.06.1992
Revised version 10.11.1992

Book Review

V. L. Girko: Theory of Random Determinants, Kluwer Academic Publishers, Dordrecht, Boston, London, 1990, XXV/677 pp., Dfl. 420.00/US$245.00/ UK£147.00

The book is monumental. It contains 677 numbered pages and 28 chapters, covering a wide range of topics. It was first published in Russian. The English version is not simply a translation, the material has been increased by about 50%.

Chapters 1–4 contain proofs of assertions concerning the probability distribution of random determinants; methods to find their moments; distributions of eigenvalues and eigenvectors of random matrices and inequalities for random determinants. Chapters 5–16 present various limit theorems concerning random determinants. Since a determinant is the signed volume of a parallelepiped, it is not surprising that a typical theorem states that the logarithm of the absolute value of a random determinant is normally distributed. Using suitable normalization, however, laws of large numbers can be proved. The more than 280 pages devoted to elaborate limit theorems on various special topics such as: accompanying infinitely divisible laws; determinants of random Gram, Toeplitz, Hankel, Jacobi and Fredholm matrices and some limit theorems connected with them; asymptotic properties of solutions of random linear algebraic and integral equations. The remaining chapters deal with applications of the theory presented in the former chapters. These include applications in statistical analysis, control theory, stochastic linear programming, pattern recognition, experimental design, Physics and numerical analysis.

The book is comprehensive and very useful. However, when reading it, one regrets that not enough effort has been put into its preparation. Names such a "Erdesh" and "Loran" should have been written as Erdös and Laurent. The theorems or methods are not always attributed to the right authors and the presentations of some areas are one-sided, some of the important results are disregarded.

In spite of the above critical remarks, Girko's book is a good one and can be used by many researchers working on various problems of stochastics.

New Brunswick A. Prekopa

Metrika (1993) 40:283–298

Metrika

On Least Favourable Two Point Priors and Minimax Estimators Under Absolute Error Loss

Wolfgang Bischoff and Werner Fieger

Institut für Mathematische Stochastik, Universität Karlsruhe, Englerstraße 2, 7500 Karlsruhe 1, Germany

Summary: In case of absolute error loss we investigate for an arbitrary class of probability distributions, if or if not a two point prior can be least favourable and a corresponding Bayes estimator can be minimax when the parameter is restricted to a closed and bounded interval of $\mathbb{R}$. The general results are applied to several examples, for instance location and scale parameter families are considered. We give examples for which, independent of the length of the parameter interval, no two point priors exist. On the other hand examples are given having a least favourable two point prior when the parameter interval is sufficiently small.

1 Introduction

In classical minimax estimation problems no restriction is made on the parameter space. In applications however, parameter spaces are usually restricted which has been taken into account by many statistical papers within the last decade. Most of these contributions have treated squared error loss; only a few papers have considered more general loss functions; cf. Bischoff (1992), Bischoff and Fieger (1992), Eichenauer-Hermann and Fieger (1992) and Eichenauer-Herrmann and Ickstadt (1992).

In this paper in case of absolute error loss we investigate for an arbitrary class of probability distributions, if or if not a two point prior can be least favourable and a corresponding Bayes estimator can be minimax. (See Section 2 for the definition of least favourable priors and their relation to minimax estimation).

It is worth mentioning that usually a two point prior is least favourable when the parameter interval is sufficiently small; cf. Zinzius (1979, 1981), Casella and Strawderman (1981), DasGupta (1985), and the papers cited above. Recently, Eichenauer-Herrmann and Ickstadt (1992) have shown among other results that, independent of the parameter interval, no two point prior is least favourable in case of some location parameter families under absolute error loss.

In this note we give general conditions which must be fulfilled if a two point prior and a corresponding Bayes estimator form a saddle point. We use these results to decide for some examples if or if not two point priors are least favourable and a corresponding Bayes estimator is minimax. In particular we investigate location and scale parameter families. By this means we can general-

0026–1335/93/5/283–298 $2.50

ize the result given by Eichenauer-Herrmann and Ickstadt (1992). On the other hand we give examples (Result 4 and Result 5) where two point priors are least favourable; these examples have the same behaviour as "expected", namely that a least favourable two point prior exists when the parameter interval is sufficiently small. The examples are chosen such that the used ideas together with the general results may be usefully applied to other examples as well.

Section 2 contains the statistical background and results of some examples. The proofs of these results are given in Section 4 using some general statements which are handled in Section 3.

2 Notation and Examples

Firstly, we give some statistical background useful for the following. Let $\{P_\theta | \theta \in \Theta\}$ be a family of probability measures on a sample space $(\mathcal{X}, \mathcal{A})$ where the parameter space $\Theta = [c, d]$ is an arbitrary interval of $\mathbb{R}$ equipped with the Borel σ-field $\mathcal{B}(\Theta)$. Let Δ denote the set of all (nonrandomized) estimators, i.e. measurable functions $\delta: \mathcal{X} \to \Theta$. Supposing absolute error loss the risk function of an estimator δ is given by

$$R(\theta, \delta) = \int_{\mathcal{X}} |\theta - \delta(x)| P_\theta(dx)$$

for $\theta \in \Theta$. Let Π be the set of all priors, i.e., probability measures on $(\Theta, \mathcal{B}(\Theta))$. The Bayes risk of an estimator δ with respect to $\pi \in \Pi$ is given by

$$r(\pi, \delta) = \int_{\Theta} R(\theta, \delta) \pi(d\theta) \ .$$

An estimator δ_π is called Bayes estimator with respect to a prior π, if δ_π minimizes $r(\pi, \cdot)$; an estimator δ^* is called minimax estimator if δ^* minimizes $\sup_{\pi \in \Pi} r(\pi, \cdot)$. A prior π^* is said to be least favourable, if π^* maximizes $\inf_{\delta \in \Delta} r(\cdot, \delta)$.

The statistical game (Π, Δ, r) is definite, if

$$\inf_{\delta \in \Delta} \sup_{\pi \in \Pi} r(\pi, \delta) = \sup_{\pi \in \Pi} \inf_{\delta \in \Delta} r(\pi, \delta) \ .$$

If in a definite statistical game (Π, Δ, r) a least favourable prior π^* and a minimax estimator δ^* exist, then (π^*, δ^*) forms a saddle point in the game (Π, Δ, r). On the other hand if (π^*, δ^*) is a saddle point in (Π, Δ, r), then (Π, Δ, r) is definite and π^* is least favourable and the estimator δ^* is minimax. Therefore, if in a

definite statistical game (Π, Δ, r) a least favourable prior π^* exists, then each minimax estimator δ^* is a Bayes estimator with respect to π^*.

Next, we give results on some illuminating examples. For that let ε_t be the one point measure on $\{t\}$.

The first result generalizes the results of Section 5 of Eichenauer-Herrmann and Ickstadt (1992).

Result 1: Let λ be the Lebesgue measure on $\mathbb{R}$, and let $X_1, \ldots, X_n$ be n independent random variables each of them having the λ-density $f_\theta(\cdot) = f(\cdot - \theta)$ where the location parameter θ is an element of $\Theta = [0, c]$, $c > 0$ fixed. Suppose that for each $a \in \mathbb{R}$ with $0 < a \leq c$ and for every $\beta \in (0, 1)$ the conditions

$$\lambda(\{x \in \mathbb{R} | \beta \cdot f_0(x) = (1 - \beta) \cdot f_a(x), f_0(x) \neq 0\}) = 0 \tag{2.1}$$

and

$$\lambda(\{x \in \mathbb{R} | \beta \cdot f_0(x) = (1 - \beta) \cdot f_a(x), f_{a/2}(x) \neq 0\}) = 0 \tag{2.2}$$

hold true. Then no two point prior can be least favourable.

Remarks: (a) The above conditions are obviously satisfied for most of the distributions used in statistics.

(b) It is worth mentioning that the assumptions supposed by Eichenauer-Herrmann and Ickstadt (1992) for their result are much more stringent than ours.

(c) A similar result holds true for scale parameter families:

Let λ be the Lebesgue measure on $\mathbb{R}$, and let $X_1, \ldots, X_n$ be n independent random variables each of them having the λ-density $f_\theta(x) = \theta \cdot g(\theta x)$ on $\mathbb{R}$ or on $(0, \infty)$ where the scale parameter θ is an element of $\Theta = [s, (1 + m)s]$ where s, $m \in \mathbb{R}$ with $s > 0$ and $m > 0$ are fixed. Provided that for each pair $a, b \in \Theta$ with $s \leq a < b \leq (1 + m) \cdot s$ and for every $\beta \in (0, 1)$ the conditions

$$\lambda(\{x \in \mathbb{R} | \beta \cdot a \cdot g(ax) = (1 - \beta) \cdot b \cdot g(bx), g(ax) \neq 0\}) = 0$$

and

$$\lambda\left(\left\{x \in \mathbb{R} | \beta \cdot a \cdot g(ax) = (1 - \beta) \cdot b \cdot g(bx), g\left(\frac{a + b}{2} x\right) \neq 0\right\}\right) = 0$$

hold true then no two point prior can be least favourable. Especially, the above conditions are fulfilled for a class of generalized gamma distributions. The proof is similar as for Result 1 and therefore omitted.

Let us have a look to a similar phenomenon for some families of normal distributions. In literature it is usually assumed that only the expectation or only the variance of a normal distribution depend on the unknown parameter. In practice this assumption is often violated. In the following result both expectation and variance depend on the unknown parameter.

Result 2: Let $s > 0$, $q \in \mathbb{R}$ be fixed, and let $\Theta = [-s, s]$. Let $X_1, \ldots, X_n$ be n independent, normally distributed random variables each of them having the mean θ and the variance $|\theta|^q$, if $\theta \neq 0$. For $\theta = 0$ we distinguish three cases depending on q:

a) $X_1, \ldots, X_n$ are uniformly distributed on $[-c, c]$ where $c > 0$ is arbitrarily fixed, if $q < 0$,
b) $X_1, \ldots, X_n$ are normally distributed with mean 0 and variance 1, if $q = 0$,
c) $X_1, \ldots, X_n$ are distributed according to ε_0, if $q > 0$.

Then no two point prior is least favourable.

Remark: Result 2 does not depend on the distribution of the random variable X_i for the parameter $\theta = 0$.

In Result 3 and Result 4 least favourable priors exist, if the parameter interval Θ is sufficiently small; that is caused by the fact that the set $C^{=}_{-s,s;1/2}$ has positive probability for some parameter $\theta \in \Theta$. (For the definition of $C^{=}_{-s,s;1/2}$ see Section 3.) An example that is partly treated in Eichenauer-Herrmann and Fieger (1989) is completely solved in Result 3.

Result 3: For fixed $s > 0$ let $\Theta = [-s, s]$, and let $X_1, \ldots, X_n$ be n independent random variables distributed according to the uniform distribution on $[-1 + \theta, 1 + \theta]$ where $\theta \in \Theta$. Let s_n denote the only positive zero of

$$h_n(s) := 2 \cdot (1 - s)^n - 1 - \tfrac{1}{2} \cdot s \cdot n - \tfrac{1}{2} \cdot s \cdot n(1 - s)^{n-1} \ .$$

Then in case of $s > s_n$ there does not exist a two point prior $\pi_{a,b;\beta}$ which is least favourable. If $0 < s \leq s_n$ then $\pi^* = \tfrac{1}{2}\varepsilon_{-s} + \tfrac{1}{2}\varepsilon_s$ is the only least favourable prior and

$$\delta^*(x_1, \ldots, x_n) = \begin{cases} -s \text{ for } \max\limits_{1 \leq i \leq n} x_i \leq -s + 1 \text{ and } -s - 1 \leq \min\limits_{1 \leq i \leq n} x_i < s - 1 \\ 0 \text{ else} \\ s \text{ for } \min\limits_{1 \leq i \leq n} x_i \geq s - 1 \text{ and } s + 1 \geq \max\limits_{1 \leq i \leq n} x_i > -s + 1 \end{cases}$$

is minimax.

Remarks: (a) For $n = 1$ the above result means:
No two point prior is least favourable, if $s > \frac{1}{3}$. But in case of $0 < s \leq \frac{1}{3}$ the prior $\pi^* = \frac{1}{2}\varepsilon_{-s} + \frac{1}{2}\varepsilon_s$ is least favourable and the estimator

$$\delta^*(x) = \begin{cases} -s & \text{for } x \in [-s-1, s-1) \\ 0 & \text{for } x \in (-\infty, -s-1) \cup [s-1, 1-s] \cup (s+1, \infty) \\ s & \text{for } x \in (1-s, s+1] \end{cases}$$

is minimax; note the risk function for $n = 1$ is given by

$$R(\theta, \delta^*) = s^2 - \theta^2 + (1-s)\cdot|\theta| \ .$$

(b) By methods similar to those of Eichenauer and Fieger (1989), pp. 370, and Eichenauer-Herrmann and Fieger (1992), p. 36, one can show that

$$s_n = \frac{1}{n}\cdot\alpha_1 + o(n^{-1}) \ ,$$

where α_1 is the positive zero of

$$h(\alpha) = 2e^{-\alpha} - 1 - \tfrac{1}{2}\alpha - \tfrac{1}{2}\alpha e^{-\alpha} \ .$$

The interesting problem of estimating $\theta \in \Theta = [0, s]$, $s > 0$, based on n independent random variables uniformly distributed on $[0, 1+\theta]$ can partially be solved by using Corollary 2. The complete solution may be obtained in a similar way as Result 4 where a "symmetrization" of that problem is treated.

Result 4: For fixed $s > 0$ let $\Theta = [-s, s]$, and let $X_1, \ldots, X_n$ be n independent random variables. Suppose that X_i, $i = 1, \ldots, n$, is uniformly distributed on $[-1+\theta, 1]$ if $\theta \in [-s, 0]$ and on $[-1, 1+\theta]$ if $\theta \in (0, s]$, respectively. Let s_n denote the only positive zero of

$$h_n(s) := -1 + 2^n\cdot(2+s)^{-n-1}\cdot[4-(n-2)s] \ .$$

Then in case of $s > s_n$ there does not exist a two point prior $\pi_{a,b;\beta}$ which is least favourable. If $0 < s \leq s_n$ then $\pi^* = \frac{1}{2}\varepsilon_{-s} + \frac{1}{2}\varepsilon_s$ is the only least favourable prior and

$$\delta^*(x_1, \ldots, x_n) = \begin{cases} -s & \text{for } \max\limits_{1 \le i \le n} x_i \le 1 \text{ and } -s-1 \le \min\limits_{1 \le i \le n} x_i \le -1 \\ 0 & \text{else} \\ s & \text{for } -1 \le \min\limits_{1 \le i \le n} x_i \text{ and } 1 \le \max\limits_{1 \le i \le n} x_i \le s+1 \end{cases}$$

is minimax.

Remark: The above result means for $n = 1$:
If $s \le \sqrt{5} - 1$, then the prior $\pi^* = \frac{1}{2}\varepsilon_{-s} + \frac{1}{2}\varepsilon_s$ is least favourable and the estimator

$$\delta^*(x) = \begin{cases} -s & \text{for } x \in [-s-1, -1) \\ 0 & \text{for } x \in (-\infty, -s-1) \cup [-1, 1] \cup (s+1, \infty) \\ s & \text{for } x \in (1, s+1) \end{cases}$$

is minimax. In case of $s > \sqrt{5} - 1$ no two point prior can be least favourable.

3 General Results

We use the notations and assumptions given at the very beginning of Section 2. Further we suppose:

$a, b \in \Theta$ with $a < b$,

P_a, P_b are absolutely continuous with respect to a measure μ, say, on $(\mathcal{X}, \mathcal{A})$.

Let f_a and f_b denote the corresponding Radon-Nikodym derivatives of P_a and P_b with respect to μ, respectively. For $\beta \in (0, 1)$ we define:

$$C^{>}_{a,b;\beta} := \{x \in \mathcal{X} \mid \beta \cdot f_a(x) - (1-\beta) \cdot f_b(x) > 0\} ,$$

$$C^{=}_{a,b;\beta} := \{x \in \mathcal{X} \mid \beta \cdot f_a(x) - (1-\beta) \cdot f_b(x) = 0\} ,$$

$$C^{<}_{a,b;\beta} := \{x \in \mathcal{X} \mid \beta \cdot f_a(x) - (1-\beta) \cdot f_b(x) < 0\} .$$

Consider the two point prior $\pi_{a,b;\beta} = \beta \cdot \varepsilon_a + (1-\beta) \cdot \varepsilon_b$. Then a Bayes estimator $\delta^* = \delta^*_{a,b;\beta}$ with respect to the prior $\pi_{a,b;\beta}$ fulfills the condition

$$\delta^*(x) \in [a, b] \quad P_a - \text{a.s.} \quad \text{and} \quad P_b - \text{a.s.} \tag{3.1}$$

Thus we get

$$r(\pi_{a,b;\beta}, \delta^*) = \int (\beta f_a(x) - (1-\beta) f_b(x)) \cdot \delta^*(x) \mu(dx) - \beta a + (1-\beta) b \ .$$

Therefore an estimator $\delta^* \in \Delta$ is a Bayes estimator with respect to $\pi_{a,b;\beta}$, if and only if

$$\delta^*(x) = \begin{cases} a & \text{for } x \in C^{>}_{a,b;\beta} \\ b & \text{for } x \in C^{<}_{a,b;\beta} \end{cases} \qquad P_a - \text{a.s.} \quad \text{and} \quad P_b - \text{a.s.} \tag{3.2}$$

and for $x \in C^{=}_{a,b;\beta}$ the value $\delta^*(x)$ may be chosen arbitrarily subject to the condition (3.1). Thus the risk function of any Bayes estimator δ^* with respect to $\pi_{a,b;\beta}$ is given by

$$R(\theta, \delta^*) = |\theta - a| \cdot P_\theta(C^{>}_{a,b;\beta}) + |b - \theta| \cdot P_\theta(C^{<}_{a,b;\beta}) + \int\limits_{C^{=}_{a,b;\beta}} |\theta - \delta^*(x)| P_\theta(dx) \ .$$

In particular, we investigate the case that $P_a(C^{=}_{a,b;\beta}) = 0$ which is equivalent to $P_b(C^{=}_{a,b;\beta}) = 0$.

In the following result $(\pi_{a,b;\beta}, \delta^*)$ is supposed to be a saddle point which implies that δ^* is a Bayes estimator with respect to $\pi_{a,b;\beta}$, i.e. δ^* fulfills (3.1) and (3.2).

Theorem 1: Let $P_a(C^{=}_{a,b;\beta}) = 0$ (or equivalently $P_b(C^{=}_{a,b;\beta}) = 0$) be fulfilled. Then each of the following conditions is necessary for $(\pi_{a,b;\beta}, \delta^*)$ being a saddle point:

a) $P_a(C^{>}_{a,b;\beta}) = P_b(C^{<}_{a,b;\beta})$,
b) $P_a(C^{<}_{a,b;\beta}) = P_b(C^{>}_{a,b;\beta})$,
c) $P_a(C^{<}_{a,b;\beta}) < \min\{\beta, 1-\beta\}$ and $P_b(C^{>}_{a,b;\beta}) < \min\{\beta, 1-\beta\}$,
d) $|\theta - a| \cdot P_\theta(C^{>}_{a,b;\beta}) + |b - \theta| \cdot P_\theta(C^{<}_{a,b;\beta}) \leq (b-a) \cdot P_a(C^{<}_{a,b;\beta})$ for all $\theta \in \Theta$.

Proof: Let $(\pi_{a,b;\beta}, \delta^*)$ be a saddle point. This implies

$$(b-a) \cdot P_a(C^{<}_{a,b;\beta}) = R(a, \delta^*) = R(b, \delta^*) = (b-a) \cdot P_b(C^{>}_{a,b;\beta}) \ ;$$

thus we have

$$P_a(C^{<}_{a,b;\beta}) = P_b(C^{>}_{a,b;\beta}), \qquad P_a(C^{>}_{a,b;\beta}) = P_b(C^{<}_{a,b;\beta}) \ .$$

Hence condition a) and condition b) are shown.

Furthermore these results and the definitions of $C^{<}_{a,b;\beta}$ and $C^{>}_{a,b;\beta}$ provide

$$1 - P_a(C^{<}_{a,b;\beta}) = P_b(C^{<}_{a,b;\beta}) > \frac{\beta}{1-\beta} P_a(C^{<}_{a,b;\beta}) \Rightarrow P_a(C^{<}_{a,b;\beta}) < 1 - \beta \ ,$$

$$1 - P_b(C^{>}_{a,b;\beta}) = P_a(C^{>}_{a,b;\beta}) > \frac{1-\beta}{\beta} P_b(C^{>}_{a,b;\beta}) \Rightarrow P_b(C^{>}_{a,b;\beta}) < \beta \ ;$$

whence we obtain

$$P_a(C^{<}_{a,b;\beta}) = P_b(C^{>}_{a,b;\beta}) < \min\{\beta, 1-\beta\} \ .$$

Thus assertion c) is proved.

Next, for fixed $\theta \in \Theta$ we consider the estimator

$$\tilde{\delta}^*(x) = \begin{cases} a & \text{for } x \in C^{>}_{a,b;\beta} \\ b & \text{for } x \in C^{<}_{a,b;\beta} \\ \theta & \text{for } x \in C^{=}_{a,b;\beta} \end{cases}$$

which is Bayes with respect to $\pi_{a,b;\beta}$. Then we get

$$R(\theta, \delta^*) \geq R(\theta, \tilde{\delta}^*) = |\theta - a| \cdot P_\theta(C^{>}_{a,b;\beta}) + |b - \theta| \cdot P_\theta(C^{<}_{a,b;\beta})$$

for every Bayes estimator δ^* with respect to $\pi_{a,b;\beta}$. Thus, for δ^* being minimax the following condition is necessary:

$$(b-a) \cdot P_a(C^{<}_{a,b;\beta}) = R(a, \delta^*) \geq R(\theta, \delta^*) \geq R(\theta, \tilde{\delta}^*) \ .$$

Hence d) follows. □

The above result provides

Corollary 2: Let $P_a(C^{=}_{a,b;\beta}) = 0$ (or equivalently $P_b(C^{=}_{a,b;\beta}) = 0$) be fulfilled. Then each of the following conditions is necessary for $(\pi_{a,b;\beta}, \delta^*)$ being a saddle point:

a) $P_a(C^{>}_{a,b;\beta}) > \max\{\beta, 1-\beta\}$,
b) $P_b(C^{<}_{a,b;\beta}) > \max\{\beta, 1-\beta\}$,
c) $P_{(a+b)/2}(C^{=}_{a,b;\beta}) > 0$,
d) $\frac{1}{2} P_{(a+b)/2}(C^{>}_{a,b;\beta} + C^{<}_{a,b;\beta}) \leq P_a(C^{<}_{a,b;\beta})$,
e) $\frac{1}{2} P_{(a+b)/2}(C^{>}_{a,b;\beta} + C^{<}_{a,b;\beta}) \leq P_b(C^{>}_{a,b;\beta})$.

Proof: Theorem 1 c) implies a) and b) since $P_a(C^{=}_{a,b;\beta}) = 0$ and $P_b(C^{=}_{a,b;\beta}) = 0$, respectively.

In order to prove c) and d) put $\theta = \dfrac{a+b}{2}$. Using Theorem 1 c) and d) we obtain

$$\frac{b-a}{2} \cdot (1 - P_\theta(C^{=}_{a,b;\beta})) = \frac{b-a}{2} \cdot P_\theta(C^{<}_{a,b;\beta} + C^{>}_{a,b;\beta}) < (b-a) \cdot \min\{\beta, 1-\beta\}$$

$$\leq \frac{b-a}{2}$$

and therefore

$$P_\theta(C^{=}_{a,b;\beta}) > 0 \ ;$$

i.e. condition c) holds true.

Furthermore by Theorem 1 b) and d) we get

$$(b-a) \cdot P_a(C^{<}_{a,b;\beta}) = (b-a) \cdot P_b(C^{>}_{a,b;\beta}) \geq \frac{b-a}{2} \cdot P_{(a+b)/2}(C^{>}_{a,b;\beta} + C^{<}_{a,b;\beta}) \ . \qquad \square$$

The following result is useful for determining triplets (a, b, β) which are not in question that $\pi_{a,b;\beta}$ and any Bayes estimator δ^* with respect to $\pi_{a,b;\beta}$ form a saddle point $(\pi_{a,b;\beta}, \delta^*)$.

Proposition 3: Let a be an interior point of Θ, suppose that $P_\theta(C^{=}_{a,b;\beta}) = 0$ for all θ of a neighbourhood of a, and that $\dfrac{d}{d\theta} P_\theta(C^{<}_{a,b;\beta})$ exists for $\theta = a$. (This implies that $\dfrac{d}{d\theta} P_\theta(C^{>}_{a,b;\beta})$ exists for $\theta = a$ as well). Then $(\pi_{a,b;\beta}, \delta^*)$ is not a saddle point.

An analogous result holds true, if b is an interior point of Θ.

Proof: For $(\pi_{a,b;\beta}, \delta^*)$ being a saddle point the left-hand derivative $\frac{d}{d(\theta-)}R(\theta, \delta^*)$ and the right-hand derivative $\frac{d}{d(\theta+)}R(\theta, \delta^*)$ of $R(\cdot, \delta^*)$ must fulfill the following conditions:

$$0 \leq \frac{d}{d(\theta-)}R(\theta, \delta^*)|_{\theta=a} = -P_a(C^{>}_{a,b;\beta}) - P_a(C^{<}_{a,b;\beta}) + (b-a)\cdot\frac{d}{d\theta}P_\theta(C^{<}_{a,b;\beta})|_{\theta=a}$$

and

$$0 \geq \frac{d}{d(\theta+)}R(\theta, \delta^*)|_{\theta=a} = P_a(C^{>}_{a,b;\beta}) - P_a(C^{<}_{a,b;\beta}) + (b-a)\cdot\frac{d}{d\theta}P_\theta(C^{<}_{a,b;\beta})|_{\theta=a}\ .$$

Thus $P_a(C^{>}_{a,b;\beta}) = 0$; whence the assertion follows by Corollary 2 a). □

4 Proofs of the Results of Section 2

In the sequel let $X_1, \ldots, X_n$ be n independent, identically distributed random variables whose distributions are specified in the handled results. Note that $\{P_\theta : \theta \in \Theta\}$ denotes the family of probability measures belonging to $X = (X_1, \ldots, X_n)$.

Proof of Result 1: Note that by Fubini's Theorem for each $a \in \mathbb{R}$ with $0 < a \leq c$ and for every $\beta \in (0, 1)$ the conditions (2.1) and (2.2) imply

$$\lambda^n\left(\left\{x = (x_1, \ldots, x_n) | \beta\cdot\prod_{i=1}^n f_0(x_i) = (1-\beta)\cdot\prod_{i=1}^n f_a(x_i), \prod_{i=1}^n f_0(x_i) \neq 0\right\}\right) = 0$$

and

$$\lambda^n\left(\left\{x = (x_1, \ldots, x_n) | \beta\cdot\prod_{i=1}^n f_0(x_i) = (1-\beta)\cdot\prod_{i=1}^n f_a(x_i), \prod_{i=1}^n f_{a/2}(x_i) \neq 0\right\}\right) = 0$$

where λ^n is the Lebesgue measure on $\mathbb{R}^n$. Then the assertion follows directly by Corollary 2 c). □

Proof of Result 2: By a straightforward calculation one can verify that $\frac{d}{d\theta} P_\theta(C^<_{a,b;\beta})$ exists for $a, b \in \Theta$ with $a < b$. Proposition 3 implies that only $(\pi, \delta) \in \Pi \times \Delta$ with

$$\pi \in \{\pi_{-s,s;\beta}, \pi_{-s,0;\beta}, \pi_{0,s;\beta} : \beta \in (0, 1)\}$$

and δ suitable chosen is in question for being a saddle point. According to Corollary 2 c) only the prior $\pi_{-s,s;1/2}$ and a suitable estimator δ can form a saddle point.

a) In case of $q \leq 0$ the above prior cannot be least favourable according to Corollary 2 c).

b) In case of $0 < q$ consider the estimator

$$\delta^*(x) = \begin{cases} -s & \text{for } x \in \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : \sum x_i < 0\} = C^>_{-s,s;1/2} \\ 0 & \text{for } x \in \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : \sum x_i = 0\} = C^=_{-s,s;1/2} \\ s & \text{for } x \in \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : \sum x_i > 0\} = C^<_{-s,s;1/2} \end{cases}$$

which is uniformly better than each other Bayes estimator. Thus we get for each Bayes estimator δ with respect to $\pi_{-s,s;1/2}$:

$$R(\theta, \delta) = (\theta + s) \cdot P_\theta(C^>_{-s,s;1/2}) + (s - \theta) \cdot P_\theta(C^<_{-s,s;1/2}) \geq s - \theta \quad \text{for } \theta > 0 \ .$$

Therefore Theorem 1 d) implies that $\pi_{-s,s,1/2}$ is not least favourable. □

Note that in the following two estimation problems a two point prior π^* is least favourable, if and only if π^* and a suitable Bayes estimator with respect to π^* form a saddle point.

Proof of Result 3: (i) Let $-s \leq a < b \leq s$. Corollary 2c) provides that $\pi_{a,b;\beta}$ can only be least favourable if $b - a < 2$ and $\beta = \frac{1}{2}$.

(ii) In order to show that the prior $\pi_{-s,s;1/2}$ is least favourable if and only if $s \leq s_n$ we may assume $s < 1$ since (i). By similar considerations as in Section 5 of Eichenauer-Herrmann and Fieger (1992) one gets that $(\pi_{-s,s;1/2}, \delta^*)$ is a saddle point if $s \leq s_n$.

On the other hand $\pi_{-s,s;1/2}$ is not least favourable for $1 > s > s_n$ because

$$\frac{d}{d\theta} R(\theta, \delta^*)\Big|_{\theta=s} < 0 \quad \text{for } s > s_n$$

and $R(\theta, \delta^*) \leq R(\theta, \delta)$ for all $\theta \in [-s, s]$ and every Bayes estimator δ with respect to $\pi_{-s,s;1/2}$.
(iii) Finally, we have to show that no other prior $\pi_{a,b;\beta}$ can be least favourable. Because of (i) and (ii) we may assume without loss of generality that

$$-s < a < b \leq s \ , \qquad b - a < 2 \ , \qquad \beta = \frac{1}{2} \ .$$

In the sequel we need the set

$$C_0^= := C_{a,b;1/2}^= \cap \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n | f_a(x_i) > 0 \text{ for } i = 1, \ldots, n\} \ .$$

Now we define for each $\theta \in [-s, b]$ the particular Bayes estimator δ_θ^* with respect to $\pi_{a,b;1/2}$. Firstly for $\theta \in [a, b]$ let be

$$\delta_\theta^*(x) = \begin{cases} a & \text{for } x \in C_{a,b;1/2}^> \\ \dfrac{a+b}{2} & \text{for } x \in C_0^= \\ b & \text{for } x \in C_{a,b;1/2}^< \\ \theta & \text{for } x \in C_{a,b;1/2}^= \backslash C_0^= \end{cases} \ .$$

Further choose $\gamma \in \mathbb{R}$ such that

$$P_a\left(C_0^= \cap \left\{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \,\middle|\, \max_{1 \leq i \leq n} x_i \leq \gamma\right\}\right) = \frac{1}{2} P_a(C_0^=)$$

and use the sets

$$C_1^= := C_0^= \cap \left\{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \,\middle|\, \max_{1 \leq i \leq n} x_i \leq \gamma\right\}$$

and

$$C_2^= := C_0^= \backslash C_1^= \ .$$

Secondly for each $\theta \in [-s, a)$ we define

$$\delta_\theta^*(x) = \begin{cases} a & \text{for } x \in C^{>}_{a,b;1/2} \cup C^{=}_1 \\ b & \text{for } x \in C^{<}_{a,b;1/2} \cup C^{=}_2 \\ \theta & \text{for } x \in C^{=}_{a,b;1/2} \backslash C^{=}_0 \end{cases} .$$

Then it can easily be seen for all $\theta \in [-s, b]$ and any Bayes estimator δ^* with respect to $\pi_{a,b;1/2}$ fulfilling $R(a, \delta^*) = R(b, \delta^*)$ that the following inequality holds true:

$$R(\theta, \delta_\theta^*) \leq R(\theta, \delta^*) .$$

Note, that

$$R(\theta, \delta_\theta^*) = (\theta - a) \cdot P_\theta(C^{>}_{a,b;1/2}) + (b - \theta) \cdot P_\theta(C^{<}_{a,b;1/2}) + \left(\frac{a+b}{2} - \theta\right) \cdot P_\theta(C^{=}_0)$$

for $\theta \in \left[a, \frac{a+b}{2}\right]$ and

$$R(\theta, \delta_\theta^*) = (a - \theta) \cdot (P_\theta(C^{>}_{a,b;1/2}) + P_\theta(C^{=}_1)) + (b - \theta) \cdot (P_\theta(C^{<}_{a,b;1/2}) + P_\theta(C^{=}_2))$$

for $\theta \in [-s, a)$. Necessary conditions for $\pi_{a,b;1/2}$ being least favourable are

$$0 \geq \frac{d}{d(\theta+)} R(\theta, \delta_\theta^*)\Big|_{\theta=a} = P_a(C^{>}_{a,b;1/2}) - P_a(C^{<}_{a,b;a/2})$$

$$+ (b - a) \cdot \frac{d}{d(\theta+)} P_\theta(C^{<}_{a,b;1/2})\Big|_{\theta=a} - P_a(C^{=}_0) + \frac{b-a}{2} \cdot \frac{d}{d(\theta+)} P_\theta(C^{=}_0)\Big|_{\theta=a}$$

and

$$0 \leq \frac{d}{d(\theta-)} R(\theta, \delta_\theta^*)\Big|_{\theta=a} = -P_a(C^{>}_{a,b;1/2}) - P_a(C^{=}_1) - P_a(C^{<}_{a,b;1/2}) - P_a(C^{=}_2)$$

$$+ (b - a) \cdot \left(\frac{d}{d(\theta-)} P_\theta(C^{<}_{a,b;1/2})\Big|_{\theta=a} + \frac{d}{d(\theta-)} P_\theta(C^{=}_2)\Big|_{\theta=a}\right) .$$

Because of

$$P_a(C^{=}_0) = P_a(C^{=}_1) + P_a(C^{=}_2) ,$$

and

$$\frac{d}{d(\theta+)} P_\theta(C^{<}_{a,b;1/2})\Big|_{\theta=a} = -\frac{d}{d(\theta-)} P_\theta(C^{=}_2)\Big|_{\theta=a} > 0 ,$$

$$\frac{d}{d(\theta+)} P_\theta(C^{=}_0)\Big|_{\theta=a} = \frac{d}{d(\theta-)} P_\theta(C^{<}_{a,b;1/2}) = 0$$

we obtain

$$P_a(C^{>}_{a,b;1/2}) < 0 .$$

Hence $\pi_{a,b;1/2}$ can not be least favourable. □

Proof of Result 4: (i) We assume that $\pi_{a,b;\beta}$ with $-s \le a < b \le s$ is least favourable and δ^* is a minimax estimator. This implies that $\pi_{-b,-a;1-\beta}$ is least favourable as well, and δ^* is Bayes estimator with respect to $\pi_{a,b;\beta}$ and $\pi_{-b,-a;1-\beta}$. Using (3.2) one gets $a = -b$. Therefore a necessary condition for $\pi_{a,b;\beta}$ being least favourable is $a = -b$ and $\beta = 1/2$ by Corollary 2c).
(ii) Next we show that $\pi_{-b,b;1/2}$ can not be least favourable for $0 < b < s$. Consider for fixed $\theta \in [-s, s]$ the particular Bayes estimator

$$\delta^*_\theta(x) = \begin{cases} b & \text{for } x \in C^{>}_{-b,b;1/2} \\ 0 & \text{for } x \in [-1, 1]^n \\ -b & \text{for } x \in C^{<}_{-b,b;1/2} \\ \theta & \text{for } C^{=}_{-b,b;1/2} \backslash [-1, 1]^n \end{cases}$$

with respect to $\pi_{-b,b;1/2}$. Then for all $\theta \in [-s, s]$ and any Bayes estimator δ^* with respect to $\pi_{-b,b;1/2}$ the following inequality holds true:

$$R(\theta, \delta^*_\theta) \le R(\theta, \delta^*) .$$

Following the idea of the proof of Result 3 part (iii) one gets that $\pi_{-b,b;1/2}$ is not least favourable for $0 < b < s$.
(iii) In order to show that the prior $\pi_{-s,s;1/2}$ is least favourable and δ^* is minimax if and only if $s \le s_n$ we consider $R(\theta, \delta^*)$ only for $\theta \ge 0$; for the risk function $R(\cdot, \delta^*)$ of δ^* is symmetrical with respect to 0. It holds for $\theta \ge 0$:

$$R(\theta, \delta^*) = \theta \cdot 2^n \cdot (2 + \theta)^{-n} + (s - \theta) \cdot (1 - 2^n \cdot (2 + \theta)^{-n}) \ .$$

Thus we get

$$\frac{d}{d\theta} R(\theta, \delta^*) = -1 + 2^n \cdot (2 + \theta)^{-n-1} \cdot [4 + s \cdot n - 2 \cdot \theta \cdot (n - 1)] \ ,$$

$$\frac{d}{d\theta} R(\theta, \delta^*)\Big|_{\theta=s} = -1 + 2^n \cdot (2 + s)^{-n-1} \cdot [4 - (n - 2)s]$$

and

$$\frac{d}{d(\theta+)} R(\theta, \delta^*)\Big|_{\theta=0} = -1 + 2 + \frac{s \cdot n}{2} > 0 \ .$$

A necessary condition for (π^*, δ^*) being a saddle point is $\frac{d}{d\theta} R(\theta, \delta^*)\Big|_{\theta=s} \geq 0$ which implies

$$s \leq \frac{4}{n - 2} \quad \text{if } n \geq 3 \ .$$

Let

$$D_n = \begin{cases} (0, \infty) & \text{if } n = 1, 2 \\ \left(0, \frac{4}{n-2}\right) & \text{if } n \geq 3 \end{cases} \ .$$

Then we have on the other hand

$$\frac{d^2}{d\theta^2} R(\theta, \delta^*) < 0 \ , \qquad \theta \in (0, s) \ ,$$

for arbitrary $s \in D_n$. Therefore $R(\cdot, \delta^*)$ is strictly concave on $(0, s)$ where $s \in D_n$. Thus $R(\cdot, \delta^*)$ is strictly isotonic on $[0, s]$, if and only if $s \leq s_n$ where s_n is the unique positive solution of

$$h_n(s) = \frac{d}{d\theta} R(\theta, \delta^*)\Big|_{\theta=s} = -1 + 2^n \cdot (2 + s)^{-n-1} \cdot [4 - (n - 2)s] = 0 \ .$$

Therefore we have

$$\left\{\tilde{\theta} \in \Theta \,\middle|\, R(\tilde{\theta}, \delta) = \max_{\theta \in \Theta} R(\theta, \delta^*)\right\} = \{-s, s\} \quad \text{if } s \leq s_n \ .$$

Thus by a well-known argument $(\pi_{-s,s;1/2}, \delta^*)$ forms a saddle point if $s \leq s_n$. It follows analogously as for Result 3 that $\pi_{-s,s;1/2}$ is not least favourable if $s > s_n$. □

References

Bischoff W (1992) Minimax and Γ-minimax estimation for functions of the bounded parameter of a scale parameter family under "L_p-loss". Statistics & Decisions 10:45–61

Bischoff W, Fieger W (1992) Minimax estimators and Γ-minimax estimators for a bounded normal mean under the loss $l_p(\theta, d) = |d - \theta|^p$. Metrika 39:185–197

Casella G, Strawderman WE (1981) Estimating a bounded normal mean. Ann. Statist. 9:870–878

DasGupta A (1985) Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set. Sankhya A 47:326–332

Eichenauer-Herrmann J, Fieger W (1992) Minimax estimation under convex loss when the parameter interval is bounded. Metrika 39:27–43

Eichenauer-Herrmann J, Fieger W (1989) Minimax estimation in scale parameter families when the parameter interval is bounded. Statistics & Decisions 7:363–376

Eichenauer-Herrmann J, Ickstadt K (1992) Minimax estimators for a bounded location parameter. Metrika 39:227–237

Zinzius E (1979) Beiträge zur Theorie der nichtsequentiellen Parameterschätzprobleme. Docteral thesis, Universität Karlsruhe

Zinzius E (1981) Minimaxschätzer für den Mittelwert θ einer normalverteilten Zufallsgröße mit bekannter Varianz bei vorgegebener oberer und unterer Schranke für θ. Math. Operationsforsch. Statist., Ser. Statist. 12:551–557

Received 24.07.1992
Revised version 29.10.1992

Metrika (1993) 40:299–318

Metrika

An Outlier Test for Linear Processes

THOMAS FLAK AND WOLFGANG SCHMID

Universität Ulm, Abteilung Stochastik, Helmholtzstraße 18, 7900 Ulm, Germany

Abstract: In this paper we introduce an outlier test for linear processes. It is assumed that an upper bound for the number of outliers is known which is not too big in relation to the sample size. The test statistic bases on the comparison of the observations with certain predictors.

We discuss the asymptotical behaviour of the test statistic under the null hypothesis 'no outlier' and derive the asymptotic distribution for the case that the distribution of the squared white noise process belongs to a certain subset of the domain of attraction of the Gumbel distribution. Especially the most important case in applications, the Gaussian white noise is included.

Keywords: Outlier, ARMA-process, linear process, outlier test, robust estimates

AMS-Classification: 62M10, 62F03

1 Introduction

In practice it is often important to know whether a data set contains outliers or not.

The answer to this question affects the choice of the statistical procedure for data analysis. Shall we apply a robust or a classical method? It is known that robust techniques must be favoured, if outliers may occur, while frequently classical procedures have advantages in outlier free situations.

Sometimes one is also directly interested in the outlying observations.

There is little published work on outlier tests in time series, e.g. *Fox (1972)*, *Martin and Zeh (1977)*, *Abraham and Yatawara (1988)*, *Schmid (1990 a/b/c)*. These tests base on a comparison of the observations $x_1, \ldots, x_n$ of the process $\{X_t\}$ with predictors $\hat{Y}_t$, $t = 1, \ldots, n$ of the undisturbed process $\{Y_t\}$.

Since in general we have no prior information on the appearance of outliers, it is advisable to use robust estimators for the model parameters. This procedure was proposed by *Martin and Zeh (1977)* in the framework of time series analysis.

A popular assumption of many robust procedures is to require that an upper bound s_n for the total number of outliers is known. In this paper we make use of this additional information.

0026–1335/93/5/299–318 $2.50

This leads to a test statistic of the following form

$$T_n := \max_{1 \le \mu(1) < \cdots < \mu(s_n) \le n} \sum_{i=1}^{s_n} (X_{\mu(i)} - \hat{Y}_{\mu(i)})^2 \ .$$

Test statistics of that kind were introduced by *Schmid (1990c)*. They base on the idea that the residuals are large at the outlier positions. Since all possible arrangements of the residuals are taken into account these quantities use maximal information.

The case $s_n \to s < \infty$ was studied by *Schmid (1990c)*. Here we discuss the more natural case that $\{s_n\}$ converges to infinity. Furthermore we assume $s_n/n \to 0$, i.e. the maximum degree of contamination shall be small, but not too small, in relation to the sample size.

All investigations are made under the null hypothesis 'no outlier'. For this reason it is not necessary to introduce an outlier generating model.

The main result of our paper is Theorem 3.1. We determine the asymptotic distribution T_n for linear processes with white noise process $\{\varepsilon_t\}$, where $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \sigma_0^2$. It is assumed that the distribution function of $(\varepsilon_t/\sigma_0)^2$ is in the domain of attraction of the Gumbel distribution. Our result includes the family of distributions considered by *Csörgö and Mason (1985)*, especially the most important case in applications $\varepsilon_t \sim N_{0,\sigma_0^2}$, i.e. $F = \chi_1^2$.

Now, in practice, we don't know the whole sequence $\{s_n\}$, we know just one value for a fixed n. Theorem 3.1 shows that in many cases the asymptotic distribution does not depend on $\{s_n\}$, provided that $s_n \to \infty$ and $s_n/n \to 0$. This ensures the applicability of the implied outlier test.

The paper is organized as follows. In *Section 2* we describe the underlying model. The main result on the asymptotic distribution of T_n together with the proof, is given in *Section 3*. The proof of the theorem bases essentially on several lemmata, which are given in an appendix at the end of the paper, in order to improve the clarity.

2 The Model

Frequently the occurrence of outliers is declared by the influence of a contamination process on the interesting system, i.e. we obtain a realization $x_1, \ldots, x_n$ of a process $\{X_t\}$, instead of one of the undisturbed process $\{Y_t\}$. Since in this paper we only analyse the behaviour of the test statistic under the null hypothesis, we may assume $X_t = Y_t$, $t = 1, \ldots, n$.

In the following let $\{Y_t\}$ be a MA(∞) process,

$$Y_t = \sum_{i=0}^{\infty} c_i(\vartheta_0)\varepsilon_{t-i} \tag{2.1}$$

with $c_0(\vartheta_0) := 1$, $\{c_i(\vartheta_0)\} \in l_1$ and $\vartheta_0 \in \Theta$. The random variables ε_t, $t \in \mathbb{Z}$ are assumed to be independent and identically distributed with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \sigma_0^2 > 0$. The underlying probability space is denoted by $(\Omega, \Sigma, \mathcal{P})$.

We suppose that the parameter set Θ is a subset of a normed space with norm $\|\cdot\|$. Let ϑ_0 be an inner point of Θ.

In addition we need some assumptions on the coefficients $\{c_i(\vartheta)\}$. Suitable conditions (A) can be obtained by restricting the induced power series P:
(A)

$(A1)$ There is $R > 1$ and an open neighbourhood $\mathcal{V} \subseteq \Theta$ of ϑ_0 such that the mapping $P: \mathcal{U} \times \mathcal{V} \to \mathbb{C}$ given by

$$P(z; \vartheta) := \sum_{i=0}^{\infty} c_i(\vartheta) z^i$$

exists and is continuous on its domain. We set $c_0(\vartheta) := 1$ and define $\mathcal{U} := \{z \in \mathbb{C}: |z| < R\}$.

$(A2)$ P is assumed to be holomorphic in z on $\mathcal{U}$ for all $\vartheta \in \mathcal{V}$ and

$$0 < \inf_{z \in \mathcal{U}} \inf_{\vartheta \in \mathcal{V}} |P(z; \vartheta)| \leq \sup_{z \in \mathcal{U}} \sup_{\vartheta \in \mathcal{V}} |P(z; \vartheta)| < \infty \ .$$

$(A3)$ There is $L > 0$ such that for all $z \in \mathcal{U}$ the series P satisfies a Lipschitz condition in ϑ, i.e.

$$|P(z; \vartheta_1) - P(z; \vartheta_2)| \leq L\|\vartheta_1 - \vartheta_2\| \quad \forall z \in \mathcal{U}, \vartheta_i \in \mathcal{V} \quad (i = 1, 2) \ .$$

Remarks: Since $c_0(\vartheta_0)$ can be regarded as a normalizing constant of ε_t, we may assume $c_0(\vartheta_0) = 1$.

If P satisfies (A), then $1/P$ does it too. Let

$$Q(z; \vartheta) := 1/P(z; \vartheta) = \sum_{i=0}^{\infty} d_i(\vartheta) z^i$$

for $z \in \mathcal{U}$, $\vartheta \in \mathcal{V}$ with coefficients $\{d_i(\vartheta)\}$. Especially we have $d_0(\vartheta) = 1$.

The assumption (A) is not very restrictive. Each stationary, causal and invertible ARMA process $\{Y_t\}$, satisfies (A). This can be seen by straightforward calculations.

The best predictor of Y_t in terms of X_τ, $\tau < t$, in the undisturbed case is given by

$$\tilde{Y}_t := \mathcal{E}(Y_t | X_\tau, \tau < t) \ . \tag{2.2}$$

For a MA(∞) process which satisfies condition (A) the one-step predictor can easily be calculated. Since P is not equal to zero, $1/P$ is holomorphic in z on $\mathcal{U}$ for each fixed $\vartheta \in \mathcal{V}$. Thus we find

$$\varepsilon_t = \sum_{i=0}^{\infty} d_i(\vartheta_0) X_{t-i}$$

(see e.g. *Brillinger (1975, p. 76/77)*). Hence it follows under the null hypothesis that

$$\tilde{Y}_t = \sum_{i=1}^{\infty} c_i(\vartheta_0)\varepsilon_{t-i} = X_t - \varepsilon_t = -\sum_{i=1}^{\infty} d_i(\vartheta_0) X_{t-i}$$

Since in practice only a finite realization $x_1, \ldots, x_n$ is present, we use the following truncated quantity as a predictor of Y_t

$$\hat{Y}_t := -\sum_{i=1}^{t-1} d_i(\vartheta_0) X_{t-i} \ .$$

For a discussion of this predictor for an ARMA process we refer to *Brockwell and Davis (1987, p. 177)*. $\hat{Y}_t$ is an approximation of $\tilde{Y}_t$ for large t.

In general ϑ_0 and σ_0 are both unknown. Let $\hat{\vartheta}_n$ and $\hat{\sigma}_n^2$ denote arbitrary estimates of ϑ_0 and σ_0^2. Since outliers may be present it is advisable to use robust estimators. E.g. GM estimators are suitable for an AR process *(Martin and Zeh (1977))* and RA estimators for an ARMA process *(Bustos and Yohai (1986))*. A review on robust estimation procedures for the parameters of an ARMA process is given by *Martin and Yohai (1985)*.

Thus a natural predictor in the case of unknown parameters is

$$\hat{Y}_t^{(n)} := -\sum_{i=1}^{t-1} d_i(\hat{\vartheta}_n) X_{t-i} \ .$$

This quantity is well-defined, provided that $\hat{\vartheta}_n \in \mathcal{V}$.

The residuals are given by

$$\hat{\varepsilon}_t^{(n)} := X_t - \hat{Y}_t^{(n)} = \varepsilon_t + \sum_{i=1}^{t-1} (d_i(\hat{\vartheta}_n) - d_i(\vartheta_0)) X_{t-i} - \sum_{i=t}^{\infty} d_i(\vartheta_0) X_{t-i} \tag{2.3}$$

and are no longer independent.

We consider the test statistic

$$\hat{T}_n = \max_{1 \le \mu(1) < \cdots < \mu(s_n) \le n} \sum_{i=1}^{s_n} (X_{\mu(i)} - \hat{Y}_{\mu(i)}^{(n)})^2 / \hat{\sigma}_n^2$$

$$= \sum_{i=1}^{s_n} \eta_{n+1-i:n} \ ,$$

where $\eta_{1:n}, \ldots, \eta_{n:n}$ denote the order statistics of $(\hat{\varepsilon}_t^{(n)}/\hat{\sigma}_n)^2$, $t = 1, \ldots, n$. Note that the test statistic is equal to the sum of the s_n largest order statistics.

In the following $\{s_n\}$ is assumed to be a sequence satisfying

$$1 \le s_n \le n, \ s_n \to \infty, \ s_n/n \to 0 \quad \text{as } n \to \infty \ . \tag{S}$$

i.e. we suppose that the maximum degree of contamination is not too big.

3 The Asymptotic Distribution of the Test Statistic

Let F be any right-continuous distribution function and

$$Q(s) := \inf\{x \in \mathbb{R}: F(x) \ge s\} \qquad \text{for all} \quad 0 < s \le 1$$

will denote its *quantile function.* Furthermore let $\omega := Q(1)$ denote the *right endpoint* of F.

The notation $\mathcal{D}(\Lambda)$ is used for the *domain of attraction of the Gumbel distribution* Λ. For $F \in \mathcal{D}(\Lambda)$ and $0 < s \le 1$ let

$$c(s) := \frac{1}{s} \int_{1-s}^{1} (1 - u) \, dQ(u) \ .$$

We will use the same integral convention as in *Csörgö et al. (1986).*

Note that the residuals $X_t - \tilde{Y}_t$, $t = 1, \ldots, n$ are independent and identically distributed.

The asymptotic distribution of the s_n largest order statistics of a sequence of independent and identically distributed random variables with distribution function F was analysed in the case (S) and $F \in \mathcal{D}(\exp(-x^{-1/a}))$ (domain of

attraction of the Fréchet distribution) by *Csörgö and Mason (1985)*, while *Lo (1989)* studied the situation $F \in \mathcal{D}(\Lambda)$.

Since $X_t - \tilde{Y}_t = \varepsilon_t$, for known parameters the asymptotic distribution of the test quantity can be obtained with the result of *Lo (1989)*.

Using the same normalizing constants, $\hat{T}_n$ is also asymptotically normally distributed. This is a consequence of Proposition 3.1.

For small t the estimator $\hat{Y}_t$ of Y_t may behave badly and thus $\hat{\varepsilon}_t^{(n)}$ is big. To ensure that the influence of the first observations is not too strong, we have to demand that the scale quantity tends to infinity. This assumption is satisfied for a large class of distribution functions belonging to the domain of the Gumbel distribution.

Proposition 3.1: Let $\{Y_t\}$ *be a* $MA(\infty)$ *process,* $(\varepsilon_t/\sigma_0)^2 \sim F$ *and* $F \in \mathcal{D}(\Lambda)$. *Suppose that the conditions* (A) *and* (S) *are satisfied. Let* $\hat{\vartheta}_n$ *be an estimator of* ϑ_0 *with* $\sqrt{n}\|\hat{\vartheta}_n - \vartheta_0\| = O_P(1)$ *and* $c(s) = c_F(s)$ *such that*

$$\sqrt{s_n}c(s_n/n) \to \infty \quad (n \to \infty) \; . \tag{3.1}$$

Denoting

$$J_n := \{\mu = (\mu(i))_{i=1,\ldots,s_n} \in \mathbb{N}^{s_n} \colon 1 \leq \mu(1) < \cdots < \mu(s_n) \leq n\}$$

then as $n \to \infty$

$$\frac{1}{\sqrt{s_n}c(s_n/n)} \left| \max_{\mu \in J_n} \sum_{i=1}^{s_n} \hat{\varepsilon}_{\mu(i)}^{(n)2} - \max_{\mu \in J_n} \sum_{i=1}^{s_n} \varepsilon_{\mu(i)}^2 \right| \xrightarrow{P} 0 \; .$$

Remark: The assumptions on the estimators in Proposition 3.1 are very weak and are satisfied both of nearly all classical estimators (e.g. Yule–Walker, ML, CLS) and the robust GM- and RA- estimators. Note that we are working under the null hypothesis 'no outliers'. For the reasons cited above robust estimators must be favoured.

Proof of Proposition 3.1: If $\hat{\vartheta}_n$ lies in the neighbourhood of ϑ_0, Lemma 4.1 can be applied on $c_i(\vartheta)$ and $d_i(\vartheta)$. A short consideration shows that this can be assumed without any restriction.

We denote by C_n the left-hand side of the assertion. It must be proved that $C_n \xrightarrow{P} 0$.

Let $\mathcal{M}_n := \{\omega \in \Omega \colon \|\hat{\vartheta}_n - \vartheta_0\| < r\}$ for some $r > 0$ such that $K_r := \{\vartheta \colon \|\vartheta - \vartheta_0\| < r\} \subseteq \mathcal{W}$ with $\mathcal{W}$ as in Lemma 4.1.

Since $\lim_{n\to\infty} \mathscr{P}(\mathscr{M}_n) = 1$, it follows that for all $\varepsilon > 0$

$$\mathscr{P}(|C_n| > \varepsilon) = \mathscr{P}(\{|C_n(\omega)| > \varepsilon\} \cap \mathscr{M}_n) + \mathscr{P}(\{|C_n(\omega)| > \varepsilon\} \cap \mathscr{M}_n^c)$$

$$= \mathscr{P}(\{|C_n(\omega)| > \varepsilon\} \cap \mathscr{M}_n) + o(1) \ .$$

Thus we can assume $\omega \in \mathscr{M}_n$, i.e. $\hat{\vartheta}_n \in \mathscr{W}$.

Let $p_n := s_n/n$ and

$$M_n := \frac{1}{\sqrt{s_n c(p_n)}} \max_{\mu \in J_n} \sum_{i=1}^{s_n} \varepsilon_{\mu(i)}^2 \ ,$$

$$\hat{M}_n := \frac{1}{\sqrt{s_n c(p_n)}} \max_{\mu \in J_n} \sum_{i=1}^{s_n} \hat{\varepsilon}_{\mu(i)}^{(n)2} \ .$$

Furthermore let

$$U_n := M_n - \sigma_0^2 \sqrt{s_n}(Q(1 - p_n) + c(p_n))/c(p_n) \tag{3.2}$$

for $1 < \tilde{\delta} < \delta$ (δ as in Lemma 4.1)

$$H_t := \sum_{i=0}^{\infty} \tilde{\delta}^{-i} \left(\frac{\varepsilon_{t-i}}{\sigma_0}\right)^2, \qquad V := \sum_{i=1}^{\infty} \frac{|\varepsilon_i|}{\delta^i} \ .$$

We have

$$|M_n - \hat{M}_n| \le \frac{1}{\sqrt{s_n c(p_n)}} \max_{\mu \in J_n} \sum_{i=1}^{s_n} |\hat{\varepsilon}_{\mu(i)}^{(n)2} - \varepsilon_{\mu(i)}^2| \ . \tag{3.3}$$

Let $\mu \in J_n$ be arbitrary, but fixed. Because of (2.3), we get $\hat{\varepsilon}_t^{(n)} = \varepsilon_t + R_{1t} - R_{2t}$ with

$$R_{1t} := \sum_{i=1}^{t-1} (d_i(\hat{\vartheta}_n) - d_i(\vartheta_0)) X_{t-i} \qquad R_{2t} := \sum_{i=t}^{\infty} d_i(\vartheta_0) X_{t-i} \ ,$$

consequently

$$\sum_{i=1}^{s_n} |\hat{\varepsilon}_{\mu(i)}^{(n)2} - \varepsilon_{\mu(i)}^2| \le \sum_{i=1}^{s_n} (R_{1\mu(i)} - R_{2\mu(i)})^2 + 2 \sum_{i=1}^{s_n} |\varepsilon_{\mu(i)}|(|R_{1\mu(i)}| + |R_{2\mu(i)}|)$$

$$\le 2(I_n + II_n + III_n + IV_n) \tag{3.4}$$

with

$$I_n := \sum_{i=1}^{s_n} (R_{1\mu(i)})^2 \ , \qquad II_n := \sum_{i=1}^{s_n} (R_{2\mu(i)})^2 \ ,$$

and

$$III_n := \sum_{i=1}^{s_n} |\varepsilon_{\mu(i)}||R_{1\mu(i)}| \ , \qquad IV_n := \sum_{i=1}^{s_n} |\varepsilon_{\mu(i)}||R_{2\mu(i)}| \ .$$

Applying Lemma 4.2 it can be found that

$$II_n \le K_2^2 H_0 \sum_{i=1}^{s_n} \delta^{-2\mu(i)} \le \frac{K_2^2}{\delta^2 - 1} H_0 \quad \text{resp} \ .$$

$$IV_n \le K_2 H_0^{1/2} \sum_{i=1}^{s_n} \delta^{-\mu(i)} |\varepsilon_{\mu(i)}| \le K_2 H_0^{1/2} V \ .$$

Thus we have an upper bound of II_n and IV_n, which does neither depend on the index set nor on n. This is the only case where we need (3.1). It follows that

$$\max_{\mu \in J_n} \frac{II_n}{\sqrt{s_n} c(p_n)} = o_P(1) \ , \qquad \max_{\mu \in J_n} \frac{IV_n}{\sqrt{s_n} c(p_n)} = o_P(1) \ . \tag{3.5}$$

Using Lemma 4.2, the inequality of Cauchy-Schwarz and Lemma 4.4, we obtain

$$\begin{aligned} I_n &\le K_1^2 \|\hat{\vartheta}_n - \vartheta_0\|^2 \sum_{i=1}^{s_n} H_{\mu(i)} \le K_1^2 \|\hat{\vartheta}_n - \vartheta_0\|^2 \max_{\mu \in J_n} \sum_{i=1}^{s_n} H_{\mu(i)} \\ &\le K_3 \|\hat{\vartheta}_n - \vartheta_0\|^2 \left(\max_{\mu \in J_n} \sum_{i=1}^{s_n} \varepsilon_{\mu(i)}^2 + \frac{1}{\tilde{\delta}} H_0 \right) \\ &\le K_3 \sqrt{s_n} c(p_n) \|\hat{\vartheta}_n - \vartheta_0\|^2 M_n + K_4 \|\hat{\vartheta}_n - \vartheta_0\|^2 H_0 \end{aligned}$$

and in the same way

$$\begin{aligned} III_n &\le \left(\sum_{i=1}^{s_n} \varepsilon_{\mu(i)}^2 \sum_{j=1}^{s_n} R_{1\mu(j)}^2 \right)^{1/2} = \left(\sum_{i=1}^{s_n} \varepsilon_{\mu(i)}^2 I_n \right)^{1/2} \\ &\le (K_5 (\sqrt{s_n} c(p_n) \|\hat{\vartheta}_n - \vartheta_0\| M_n)^2 + K_6 \sqrt{s_n} c(p_n) \|\hat{\vartheta}_n - \vartheta_0\|^2 M_n H_0)^{1/2} \ . \end{aligned}$$

Both upper bounds do not depend on the partition, thus

$$\max_{\mu \in J_n} \frac{I_n}{\sqrt{s_n c(p_n)}} = o_P(1) \quad \text{and} \quad \max_{\mu \in J_n} \frac{III_n}{\sqrt{s_n c(p_n)}} = o_P(1) \ , \tag{3.6}$$

if

$$i) \ \|\hat{\vartheta}_n - \vartheta_0\| M_n = o_P(1) \quad \text{and} \quad ii) \ \sqrt{n s_n} c(p_n) \to \infty (n \to \infty) \ .$$

Since Theorem 4.1 implies $U_n = O_P(1)$ (see (3.2)), we conclude

$$\|\hat{\vartheta}_n - \vartheta_0\| M_n = \frac{M_n O_P(1)}{\sqrt{n}} = \frac{U_n O_P(1)}{\sqrt{n}} + \frac{\sqrt{s_n}(Q(1-p_n) + c(p_n))}{\sqrt{n c(p_n)}} O_P(1)$$

$$= o_P(1) + \left(\sqrt{p_n} + \frac{\sqrt{p_n} Q(1-p_n)}{c(p_n)} \right) O_P(1) \ .$$

Now p_n tends to 0. Lemma 4.3 *iii)* implies that the right-hand term in the above equality is converging to 0 too, and *i)* is proved.

With some $0 < \gamma < 1/2$ and $K(\gamma) > 0$ as in Lemma 4.3, we obtain that, as $n \to \infty$.

$$\sqrt{s_n n} c(p_n) = n \sqrt{p_n} c(p_n) \geq n K(\gamma) p_n^{\gamma + 1/2} = K(\gamma) n^{1/2 - \gamma} s_n^{1/2 - \gamma} \to \infty$$

and thus *ii)* is true.

Using (3.5) and (3.6) the proof is finished. □

This result immediately implies

Theorem 3.1: Let $\{Y_t\}$ *be a* $MA(\infty)$ *process,* $(\varepsilon_t/\sigma_0)^2 \sim F$ *and* $F \in \mathcal{D}(A)$. *Suppose that the assumptions* (A) *and* (S) *are satisfied. Let* $\hat{\vartheta}_n$ *and* $\hat{\sigma}_n^2$ *be estimators of* ϑ_0 *and* σ_0^2 *with*

$$\sqrt{n} \|\hat{\vartheta}_n - \vartheta_0\| = O_P(1) \quad \textit{and}$$

$$\sqrt{n} |\hat{\sigma}_n^2 - \sigma_0^2| = O_P(1) \ .$$

If furthermore $c(s) = c_F(s)$ *with* $p_n := s_n/n$ *satisfies*

$$\sqrt{s_n} c(p_n) \to \infty \quad (n \to \infty) \ ,$$

then as $n \to \infty$

$$\frac{1}{\sqrt{2s_n}c(p_n)\hat{\sigma}_n^2} \max_{1 \le \mu(1) < \cdots < \mu(s_n) \le n} \sum_{i=1}^{s_n} (\hat{\varepsilon}_{\mu(i)}^{(n)2} - \hat{\sigma}_n^2(Q(1-p_n) + c(p_n))) \xrightarrow{d} X \ ,$$

with $X \sim \Phi$, *where* Φ *denotes the standard normal distribution.*

Proof: We use the same notation as in proof of Proposition 3.1 and

$$R_n := \frac{1}{\sqrt{2}\sigma_0^2} M_n - \frac{\sqrt{s_n}(Q(1-p_n) + c(p_n))}{\sqrt{2}c(p_n)} \ ,$$

$$\hat{R}_n := \frac{1}{\sqrt{2}\hat{\sigma}_n^2} \hat{M}_n - \frac{\sqrt{s_n}(Q(1-p_n) + c(p_n))}{\sqrt{2}c(p_n)} \ .$$

Furthermore we define $S_n := \sigma_0^2 R_n$, $\hat{S}_n := \hat{\sigma}_n^2 \hat{R}_n$.

Combining Proposition 3.1 and Lemma 4.3 *iii)* we have

$$\hat{S}_n - S_n = \frac{1}{\sqrt{2}}(\hat{M}_n - M_n) - \frac{\sqrt{p_n}(Q(1-p_n) + c(p_n))}{\sqrt{2}c(p_n)} \sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)$$

$$= o_P(1) \ .$$

Theorem 4.1 gives $R_n = O_P(1)$ and $R_n \xrightarrow{d} X \sim \Phi$. Since

$$\hat{R}_n = \frac{1}{\hat{\sigma}_n^2} \hat{S}_n = \frac{\hat{S}_n - S_n}{\hat{\sigma}_n^2} + \frac{\sigma_0^2}{\hat{\sigma}_n^2} R_n = R_n + o_P(1)$$

the assertion follows. □

Remarks: 1. (3.1) is satisfied, if one of the following assumptions is true:

i) There exists $c > 0$ such that

$$\liminf_{s \to 0+} c(s) = c \ .$$

ii) There are $0 < \beta < 1$ and $c > 0$ so that

$$\liminf_{s \to 0+} s_n / n^{\beta} = c \ .$$

Proof: i) is evident; *ii)* is an immediate consequence of Lemma 4.3 with $\delta := \beta/2$. □

Thus Theorem 3.1 can be applied to all distribution functions discussed in *Csörgö and Mason (1985, Theorem 1.5)*.

2. (3.1) is not always satisfied.

Consider e.g. the following distribution function with $\sigma^2 := \pi/2$

$$F(x) := \begin{cases} 2N_{0,\sigma^2}(x) - 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Hence $f(x) = 2n_{0,\sigma^2}(x)$, $x \geq 0$ is a density of F and we have

$$Q(s) = N^{-1}_{0,\sigma^2}\left(\frac{1+s}{2}\right) \qquad \text{for all } 0 \leq s < 1$$

The special choice of σ^2 ensures the first moment to be equal to 1. The upper tail probability of F and N_{0,σ^2} are of same order of magnitude.

Using a *von Mises condition*, (see e.g. *Resnick (1987, Proposition 1.17)*), which is easy to calculate for F and all normal distribution functions, we obtain $F \in \mathscr{D}(\Lambda)$.

Expansion of $1 - F(u)$ in terms u for $u \to \infty$, the von Mises condition and using Lemma 4.3 *i)* lead to

$$c(s) \sim \frac{\sigma^2}{Q(1-s)} \ , \qquad Q(1-s) \sim \sqrt{-2\sigma^2 \log s} \quad \text{as } s \to 0+$$

Thus we get in (3.1) as $n \to \infty$

$$\sqrt{s_n} c(p_n) \sim \frac{\sigma^2 \sqrt{s_n}}{Q(1-p_n)} \sim K\left(\frac{s_n}{-\log p_n}\right)^{1/2} .$$

The last quantity converges to infinity if and only if

$$\lim_{n \to \infty} \frac{\log n}{s_n} = 0 \ .$$

3. In applications it is often assumed that the white noise is normally distributed, i.e. $(\varepsilon_t/\sigma_0)^2 \sim \chi_1^2$.

Applying Lemma 4.3 *i)* to a gamma distribution function F we find that

$$\lim_{s\to 0+} c_F(s) = c_F > 0 \ , \qquad \text{especially} \qquad \lim_{s\to 0+} c_{\chi_1^2}(s) = 2 \ .$$

Because of remark 1, (3.1) is satisfied. Since gamma distributions belong to $\mathscr{D}(\Lambda)$, Theorem 3.1 can be applied. Note that we don't need any restrictions on $\{s_n\}$ except of (*S*).

4. By means of a simulation study we have determined the empirical power functions of some outlier tests of type T_n.

We confined ourselves to the case that the undisturbed process is an AR process with a Gaussian white noise. To describe the appearance of outliers an additive outlier model is used. We compared the test φ_{s_n} basing on T_n with the test φ basing on the maximum of the squared one-step residuals. It turns out that φ_{s_n} provides better results than φ, if several outliers occur, i.e. it is recommendable to make use of the information on the upper bound s_n. Furthermore the test φ_{s_n} shows to be rather robust with regard to the choice of s_n, provided that s_n does not exceed the exact number of outliers too much. The results of the simulation study are given in *Flak (1992)*.

A closer investigation of the proof of Proposition 3.1 and Theorem 3.1 leads to some interesting consequences.

Corollary 3.1: Let $\{Y_t\}$ denote a causal and invertible AR process with $(\varepsilon_t/\sigma_0)^2 \sim F$ and $F \in \mathscr{D}(\Lambda)$. Suppose (S) is satisfied.

Let $\hat{\vartheta}_n$ and $\hat{\sigma}_n^2$ be estimators of ϑ_0 and σ_0^2 with

$$\sqrt{n}\|\hat{\vartheta}_n - \vartheta_0\| = O_P(1) \qquad \textit{and}$$

$$\sqrt{n}|\hat{\sigma}_n^2 - \sigma_0^2| = O_P(1) \ .$$

Then, denoting $p_n := s_n/(n-p)$, as $n \to \infty$

$$\frac{1}{\sqrt{2s_n c(p_n)\hat{\sigma}_n^2}} \max_{p+1 \le \mu(1) < \cdots < \mu(s_n) \le n} \sum_{i=1}^{s_n} (\hat{\varepsilon}_{\mu(i)}^{(n)2} - \hat{\sigma}_n^2(Q(1-p_n) + c(p_n))) \xrightarrow{d} X \ ,$$

with $X \sim \Phi$.

Proof: The proof is similar to that of Theorem 3.1. The only difference is that Proposition 3.1 may not be applied, since (3.1) is not required. $\hat{\varepsilon}_t^{(n)}$ is of rather simple form for all $t > p$, since $d_i(\vartheta_0) = 0$ for all $i \geq p + 1$.

We have to examine the $n - p$ residuals $\hat{\varepsilon}_t^{(n)}$, $t = p + 1, \ldots, n$. Note that $R_{2t} = 0$ for all $t > p$, consequently we have $II_n = 0$ and $IV_n = 0$. The remaining terms I_n, III_n can be handled as above. This completes the proof. □

In the case of an AR(p) model the processes $\tilde{Y}_t$ and $\hat{Y}_t$ are identical for all $t > p$. Therefore the loss of information we suffered by using the truncated one-step predictor (see *Brockwell and Davis (1987, p. 117)*) is not crucial.

In section 2) we stated that all causal and invertible ARMA processes satisfy (*A*). The moving average component causes $\tilde{Y}_t$ to depend on all X_τ, $\tau < t$, i.e. the truncated approximations $\hat{Y}_t$ will behave badly for small t. The loss of information may be important if (3.1) is not satisfied (see e.g. remark 2).

We now state a variant of Theorem 3.1 without using (3.1). One can proceed as follows:

Examine the time series realizations $x_{r(n)+1}, \ldots, x_n$ for some $r(n) < n$, but use the preceding (undisturbed) $x_1, \ldots, x_{r(n)}$ to improve $\tilde{Y}_t^{(n)}$, $t > r(n)$. Choose $r(n)$ with $r(n) \to \infty (n \to \infty)$. On the other hand $r(n)$ should not grow too fast, because we want to examine as many data as possible.

Corollary 3.2 gives some proper choice of $r(n)$.

Corollary 3.2: Let $\{Y_t\}$, $\hat{\vartheta}_n$ *and* $\hat{\sigma}_n^2$ *be as in Theorem 3.1. Suppose* (*A*) *and* (*S*) *are satisfied. If* $r(n) \sim \alpha \log n$ *for some* $\alpha > 0$, *then we have with* $p_n := s_n/(n - r(n))$ *as* $n \to \infty$

$$\frac{1}{\sqrt{2s_n}c(p_n)\hat{\sigma}_n^2} \max_{r(n)+1 \leq \mu(1) < \cdots < \mu(s_n) \leq n} \sum_{i=1}^{s_n} (\hat{\varepsilon}_{\mu(i)}^{(n)2} - \hat{\sigma}_n^2(Q(1 - p_n) + c(p_n))) \xrightarrow{d} X \ ,$$

with $X \sim \Phi$.

Proof: Since $n - r(n) \sim n$, p_n tends to 0. We replace n by $n - r(n)$ in M_n and R_n. For reasons of simplicity the same notation is used for the new quantities.

The behaviour of I_n, III_n is still the same. Furthermore we get

$$II_n \leq K_2^2 H_0 \sum_{i=1}^{s_n} \delta^{-2\mu(i)} \leq \frac{K_2^2}{\delta^2 - 1} \delta^{-2r(n)} H_0 \qquad \text{and}$$

$$IV_n \leq \tilde{K}_2 H_0^{1/2} V_{r(n)} \delta^{-r(n)}$$

with the stationary process $V_n = \sum_{i=1}^{\infty} \delta^{-i} |\varepsilon_{i+n}|$.

Both upper bounds converge in probability to zero. Adding the scaling constants we only have to ensure that instead of (3.1) the weaker condition

$$\delta^{r(n)}\sqrt{s_n}c(p_n) \to \infty \qquad \text{as} \quad n \to \infty$$

is satisfied.

But since $r(n) \sim \alpha \log n$, this follows at once (Lemma 4.3). □

4 Appendix

Lemma 4.1: Let P be a power series satisfying condition (A). Then for each $1 < \delta < R$ exist a $K = K(\delta) > 0$ and an open neighbourhood $\mathcal{W} = \mathcal{W}(\delta)$ of ϑ_0 such that

i) *for all $i \in \mathbb{N}_0$, $\vartheta_1 \in \mathcal{W}$, $\vartheta_2 \in \mathcal{W}$*

$$|c_i(\vartheta_1) - c_i(\vartheta_2)| \leq \frac{K}{\delta^i}\|\vartheta_1 - \vartheta_2\| \ .$$

ii) *for all $\vartheta \in \mathcal{W}$*

$$|c_i(\vartheta)| \leq \frac{K}{\delta^i} \qquad \text{for all } i = 0, 1, \ldots$$

Proof: Let $K_r := \{\vartheta \in \Theta: \|\vartheta - \vartheta_0\| < r\}$, $r > 0$.

Since $\mathcal{V}$ is open, there exists some $r > 0$ such that

$$\mathcal{W} := K_r \subset K_{2r} \subseteq \mathcal{V}.$$

P is a holomorphic function in z for fixed ϑ. Applying Cauchy's integral formula, we have for all j

$$c_j(\vartheta) = \frac{1}{2\pi i}\oint_{\mathcal{L}} \frac{P(z;\vartheta)}{z^{j+1}}\,dz \ , \tag{4.1}$$

where $\mathscr{L}$ represents a circle of radius $1 < \delta < R$ centered at 0. Hence the second inequality follows with

$$K_1 = \sup_{|z| \le \delta} \sup_{\vartheta \in \mathscr{W}} |P(z; \vartheta)| \ .$$

Combining (*A*3) and (4.1) we get *i)*. □

Lemma 4.2: Suppose that the assumptions of Proposition 3.1 are satisfied and that $\hat{\vartheta}_n \in \mathscr{W}$ *(see Lemma 4.1), there exist* $K_1 > 0$, $K_2 > 0$ *such that*

$$|R_{1t}| \le K_1 \|\hat{\vartheta}_n - \vartheta\| H_t^{1/2} \qquad \text{and} \qquad |R_{2t}| \le \frac{K_2}{\delta^t} H_0^{1/2} \ ,$$

where R_{1t} *and* R_{2t} *are defined as in proof of Proposition 3.1.*

Proof: Given $1 < \delta < \tilde{\delta} < R$ there is some $c > 0$ such that $m/\delta^m \le c/\tilde{\delta}^m$ for all $m = 1, 2, \ldots$. Applying Lemma 4.1 to P and $Q = 1/P$, the Cauchy-Schwarz inequality and the above inequality, we find

$$|R_{1t}| \le K^2 \|\hat{\vartheta}_n - \vartheta_0\| \sum_{i=1}^{t-1} \sum_{j=0}^{\infty} \delta^{-i-j} |\varepsilon_{t-i-j}| \le \tilde{K}_0 \|\hat{\vartheta}_n - \vartheta_0\| \sum_{m=1}^{\infty} m\delta^{-m} |\varepsilon_{t-m}|$$

$$\le \tilde{K}_1 \|\hat{\vartheta}_n - \vartheta_0\| \sum_{m=1}^{\infty} \tilde{\delta}^{-m} |\varepsilon_{t-m}| \le K_1 \|\hat{\vartheta}_n - \vartheta_0\| H_t^{1/2} \ .$$

Moreover, we obtain that

$$|R_{2t}| \le K^2 \sum_{i=t}^{\infty} \sum_{j=0}^{\infty} \delta^{-i-j} |\varepsilon_{t-i-j}| = \frac{K^2}{\delta^t} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \delta^{-i-j} |\varepsilon_{-i-j}|$$

$$\le \frac{\tilde{K}_2}{\delta^t} \sum_{m=0}^{\infty} \tilde{\delta}^{-m} |\varepsilon_{-m}| \le \frac{K_2}{\delta^t} H_0^{1/2} \ .$$

and thus the proof is finished. □

A fundamental auxiliary result for the proof of Theorem 3.1 and Proposition 3.1 is due to *Lo (1989)*.

Theorem 4.1: Let $X_{i:n}$, $i = 1, \ldots, n$ denote the order statistics of an independent random sample of size n with distribution function $F \in \mathcal{D}(\Lambda)$. Let $\{s_n\}$ satisfy (S) and define $p_n := s_n/n$. Then as $n \to \infty$

$$\frac{1}{\sqrt{2s_n}c(p_n)} \sum_{i=1}^{s_n} (X_{n-s_n+i:n} - (Q(1-p_n) + c(p_n))) \xrightarrow{d} X \ ,$$

with $X \sim \Phi$.

Proof: See *Lo (1989)*. □

The crucial normalizing constant $c(s)$ needs some further study. In the following we prefer, for technical reasons, to work with functionals of F instead of Q, as *Lo (1989)* did.

Given a distribution function F with right endpoint ω we define for all $u < \omega$

$$m(u) := \frac{1}{1-F(u)} \int_u^\omega t \, dF(t) \qquad \text{and} \qquad r(u) := m(u) - u$$

A sufficient condition for the existence of $m(u)$, $r(u)$ is $\int_u^\omega |t|^\gamma \, dF(t) < \infty$ for some $\gamma > 1$. The integral exists for all $\gamma \geq 0$, if $F \in \mathcal{D}(\Lambda)$ (*de Haan (1970, Corollary 2.5.3)*).

Partial integration leads to

$$r(u) = \frac{1}{1-F(u)} \int_u^\omega (1 - F(t)) \, dt \ .$$

The first part of Lemma 4.3 connects $c(s)$ with $r(u)$, while property *ii)* provides a lower bound for $c(s)$.

Lemma 4.3: Assume that $F \in \mathcal{D}(\Lambda)$.

i) Let $\Delta(s) := 1 - F(Q(1-s))$ for all $0 < s < 1$, then

$$\lim_{s \to 0+} \frac{\Delta(s)}{s} = 1 \qquad \text{and} \qquad c(s) = \frac{\Delta(s)}{s} r(Q(1-s)) \ .$$

ii) For all $\delta > 0$ there exist $0 < s_0(\delta) < 1$ and $K(\delta) > 0$ such that

$$c(s) \geq K(\delta) s^\delta \qquad \text{for all } 0 < s \leq s_0(\delta) \ .$$

iii) For all $\delta > 0$ we have

$$\lim_{s \to 0+} \frac{Q(1-s)s^{\delta}}{c(s)} = 0 \ .$$

Proof:

i): Recalling $1 - F(Q(1-s)) \leq s \leq 1 - F(Q(1-s)-)$ and using $\lim_{x \to \omega-}(1 - F(x))/(1 - F(x-)) = 1$ *(Resnick (1987, Corollary 1.6))* we get the first part of *i)*.
Applying the integral transformation to $m(Q(1-s))$ gives

$$\begin{aligned} &\Delta(s)m(Q(1-s)) \\ &\quad = \int_{Q(1-s)}^{\omega} t \, dF(t) = E(I_{(Q(1-s),\omega)}(X)X) = E(I_{(F(Q(1-s)),1)}(U)Q(U)) \\ &\quad = E(I_{(1-s,1)}(U)Q(U)) - E(I_{(1-s,F(Q(1-s))]}(U)Q(U)) \\ &\quad = \int_{1-s}^{1} Q(u)\, du - Q(1-s)(s - \Delta(s)) \ . \end{aligned}$$

since $Q(u) > Q(1-s)$ is equivalent to $u > F(Q(1-s))$ and since $F(Q(1-s)) \geq 1 - s$ we have for $1 - s \leq t \leq F(Q(1-s))$:

$$Q(1-s) = Q(F(Q(1-s))) \geq Q(t) \geq Q(1-s) \quad \text{i.e.} \quad Q(t) = Q(1-s)$$

(see *Shorack and Wellner (1986, p. 5)*). Using $r(Q(1-s)) = m(Q(1-s)) - Q(1-s)$ and noting that partial integration of $c(s)$ leads to $c(s) = -Q(1-s) + \int_{1-s}^{1} Q(u)\, du$ we obtain the required result.
ii): For $u < \omega$ we define

$$h(u) := \int_u^{\omega} \int_t^{\omega} (1 - F(s))\, ds\, dt \ .$$

Resnick (1987, Prop. 1.9), resp. *de Haan (1970, Theorem 2.5.2)* proved with the preceding notation

$$\lim_{u \to \omega-} \frac{h(u)}{(1 - F(u))r^2(u)} = 1 \ . \tag{4.2}$$

De Haan (1970, Theorem 2.5.2) has given the following representation for F

$$1 - F(u) = b(u)\, exp\left(-\int_{x_1}^{u} \frac{a(t)}{f(t)}\, dt \right) \qquad \text{for all } u < \omega \ ,$$

where $a(t)$, $b(t)$, $f(t)$ are real functions with

$$\lim_{t \to \omega -} a(t) = 1 \ , \qquad \lim_{t \to \omega -} b(t) = b > 0 \ , \qquad f(t) = \frac{h(t)}{(1 - F(t))r(t)} \ .$$

This means that for each $\varepsilon > 0$ exists a $x_1 \leq \omega_1 < \omega$ such that

$$a(t) \geq 1 - \varepsilon \ , \qquad b(t) \leq 2b \qquad \text{for all } \omega_1 \leq u < \omega \ .$$

Since $\int_{\omega_1}^{u} 1/f(t)\, dt = c - \log(h(u))$ for all $u \geq \omega_1$, we have for some L_0, $L > 0$

$$1 - F(u) \leq L_0\, exp\left(-(1 - \varepsilon) \int_{\omega_1}^{u} 1/f(t)\, dt \right) \leq L(h(u))^{1-\varepsilon} \ .$$

Using (4.2) (enlarge ω_1 if necessary) we get for all $\omega_1 \leq u < \omega$ (choose $K > 0$ small enough)

$$r(u) \geq K(1 - F(u))^{\varepsilon/(2-2\varepsilon)} \ ,$$

i.e. for all $\delta > 0$ there are $\omega_1(\delta) < \omega$ and $K(\delta) > 0$ such that

$$r(u) \geq K(\delta)(1 - F(u))^{\delta} \qquad \text{for all} \quad \omega_1 \leq u < \omega \ . \tag{4.3}$$

Note that $Q(1 - s) \geq \omega_1(\delta)$ for all $0 < s \leq s_0(\delta)$ ($s_0(\delta)$ small). Replacing in inequality (4.3) for all $u = Q(1 - s)$, $0 < s \leq s_0(\delta)$ (reduce $K(\delta)$ if necessary) $r(Q(1 - s))$ by $c(s)$ and $1 - F(Q(1 - s))$ by s (see part *i)*) completes the proof.
iii): Let $\delta > 0$. Applying *ii)* with $\delta/2$, it follows that for all sufficiently small $s > 0$

$$0 \leq \frac{|Q(1 - s)|s^{\delta}}{c(s)} \leq \frac{|Q(1 - s)|s^{\delta/2}}{K(\delta/2)} \to 0 \qquad (s \to 0+) \ ,$$

since $\lim_{u \to \omega -} u^{\gamma}(1 - F(u)) = 0$ for all $\gamma \geq 0$ (see *de Haan (1970, Corollary 2.5.3)*). □

Lemma 4.4: Let ε_t, $t \in \mathbb{Z}$ *be a random process with* $\sup_t E(\varepsilon_t^2) < \infty$. *For some* $0 \le \vartheta < 1$ *let* $H_t := \sum_{i=0}^{\infty} \vartheta^i \varepsilon_{t-i}^2$. *Then for all* $1 \le k \le n$

$$\max_{1 \le \mu(1) < \cdots < \mu(k) \le n} \sum_{i=1}^{k} H_{\mu(i)} \le \frac{1}{1-\vartheta} \left(\max_{1 \le \mu(1) < \cdots < \mu(k) \le n} \sum_{i=1}^{k} \varepsilon_{\mu(i)}^2 + \vartheta H_0 \right)$$

Proof: We use the elementary inequality

$$\max_{0 \le \mu(1) < \cdots < \mu(r) \le m-1} \sum_{i=1}^{r} a_{\mu(i)} \le a_0 + \max_{1 \le \mu(1) < \cdots < \mu(r) \le m} \sum_{i=1}^{r} a_{\mu(i)} , \tag{4.4}$$

valid for $a_i \ge 0$, $i = 0, 1, \ldots, m$, $1 \le r \le m$.

Let

$$M(k) := \max_{1 \le \mu(1) < \cdots < \mu(k) \le n} \sum_{i=1}^{k} H_{\mu(i)} ,$$

$$MRes(k) := \max_{1 \le \mu(1) < \cdots < \mu(k) \le n} \sum_{i=1}^{k} \varepsilon_{\mu(i)}^2 .$$

and fix an arbitrary partition $0 \le \mu(1) < \cdots < \mu(k) \le n-1$.

Applying (4.4) we get

$$\begin{aligned} \sum_{i=1}^{k} H_{\mu(i)} &= \sum_{i=1}^{k} (\vartheta H_{\mu(i)-1} + \varepsilon_{\mu(i)}^2) = \vartheta \sum_{i=1}^{k} H_{\mu(i)-1} + \sum_{i=1}^{k} \varepsilon_{\mu(i)}^2 \\ &\le \vartheta \max_{0 \le \mu(1) < \cdots < \mu(k) \le n-1} \sum_{i=1}^{k} H_{\mu(i)} + MRes(k) \\ &\le \vartheta(H_0 + M(k)) + MRes(k) . \end{aligned}$$

Maximizing the left-hand side of the inequality gives the result. □

References

Abraham B and Yatawara N (1988) *A score test for detection of time series outliers.* J. Time Series Analysis 9(2):109–119

Brillinger DR (1975) *Time Series: Data Analysis and Theory.* Holt, Rhinehart and Winston Inc., New York

Brockwell PJ and Davis RA (1987) *Time Series: Theory and Methods.* Springer, New York

Bustos OH and Yohai VJ (1986) *Robust estimation for ARMA-models.* J. American Statist. Assoc. 81:155–168

Csörgö S, Csörgö M, Horváth L and Mason DM (1986) *Normal and stable convergence of integral functions of the empirical distribution function.* Ann. Probab., 14(1):86–118

Csörgö S and Mason DM (1985) *Central limit theorems for sums of extreme values.* Math. Proc. Cambridge Phil. Soc. 98:547–558

de Haan L (1970) *On Regular Variation and its Application to the Weak Convergence of Sample Mean.* Mathematisch Centrum, Amsterdam

Flak T (1992) *Ausreißertests basierend auf Extremsummen und Methoden zur Ausreißeridentifikation bei Zeitreihen.* Dissertation Universität Ulm, Ulm (Germany)

Fox AJ (1972) *Outliers in time series.* JR Statist. Soc. (B) 34(3):350–363

Lo GS (1989) *A note on the asymptotic normality of sums of extreme values.* Journal of Statistical Planning and Inference 22:127–136

Martin RD and Yohai VJ (1985) *Robustness in time series and estimating ARMA-models.* Eds.: Hannan, EJ, Krishnaiah, PR and Rao, MM: Handbook of Statistics Vol. 5 p. 119–155. Elsevier Science Publishers BV

Martin RD and Zeh JE (1977) *Determining the character of time series outliers.* ASAProBuEc 818–823

Resnick SI (1987) *Extreme Values, Regular Variation and Point Processes.* Springer, New York

Schmid W (1990a) *How to locate outliers in a time series, if a starting-block is present.* To appear in Sankhya, Ser. B

Schmid W (1990b) *Discussion of a LR test for detecting outliers in time series data.* Statistics & Decisions 8:271–294

Schmid W (1990c) *Ausreißertests und Ausreißeridentifikation bei Zeitreihen.* Habilitationsschrift, Ulm (Germany)

Shorack GR and Wellner JA (1986) *Empirical Processes with Applications to Statistics.* Wiley, New York

Received: 02.09.1991
Revised version: 01.12.1992

Book Review

S. Fienberg/J. Olkin (Eds.): Log Linear Models by Ronald Christensen, Springer Texts in Statistics, Springer-Verlag, 1990, 419 pp., DM 98,—

This book is written for advanced Master's level statistics students. Broadly, it discusses the use of log-linear models to analyze discrete multivariate data with nominal-valued variates (contingency tables data), the use of logistic regression and logistic discrimination to analyze binary response data with covariates, and generalized linear models (GLIMs). The authors' intention is to "fill a niche between basic introductory texts such as Fienberg (1980) and Everitt (1977) and advanced books such as Bishop, Fienberg, and Holland (1975), Haberman (1974), and Santner and Duffy (1989)."

The chapter titles are as follows.

Chapter 1 Introduction
Chapter 2 Two-Dimensional Tables
Chapter 3 Three-Dimensional Tables
Chapter 4 Higher Dimensional Tables
Chapter 5 Models for Factors with Qualitative Levels
Chapter 6 The Matrix Approach to Log-linear Models
Chapter 7 Response Factors and Logistic Discrimination
Chapter 8 The Matrix Approach to Logit Models
Chapter 9 Fixed and Random Zeros
Chapter 10 Generalized Linear Models
Chapter 15 Maximum Likelihood Theory for Log-linear Models

Chapter 1 provides some motivating examples and reviews the binomial, multinomial, and Poisson distributions.

Chaps. 2–4, 6 and 9 study log-linear models for the joint distribution of discrete multivariate data, each of whose variates is a finite valued nominal responses. These chapters form the heart of the book. Sections 3.5 and 6.3 also briefly consider product-multinomial sampling, which is a prelude to the formal distinction between response and explanatory variables emphasized beginning in Chap. 7. Chap. 9 discusses fixed and random zeroes.

Among the book's strengths is its inclusion of the rather recent literature on graphical models and their use in interpreting certain log-linear models. The book also contains an extensive discussion of model selection algorithms which includes a nice case study illustrating the algorithms. There is also considerable material on residual analysis and the identification of influential points. In order to completely understand the discussion of maximum likelihood theory in Chap. 6, the student should have a good course in linear algebra and be familiar with the notion of column space. Similarly, the asymptotic theory discussed in 6.2 may be difficult for some applied students.

Chapter 5 describes the uniform association model for tables formed from one or quantitative dimensions. Sections 5.1–5.2 consider the case of log-linear

models which arise when the factor scores are known while Sect. 5.3 discusses the case of unknown factor scores.

The study of response variables and their differentiation from 'explanatory' variables is introduced in Chap. 7 and Chap. 8 uses matrices to state the models and associated inference procedures. Most applied M.S. students will have difficulty thoroughly mastering the material in 8.2 on asymptotic results (projections, sups, norms and the like are prerequisites.)

The primary case considered in Chap. 7 and 8 is that of a binary response with multivariate explanatory variables. However the extension to a multinomial response is considered in 7.2. Discrimination and classification based on the logistic model are introduced in 7.4 and 8.6. Chapter 7 concludes with a discussion of 'recursive casual models' which are useful when external considerations suggest a hierarchy of the variables in which variables lower on the hierarchy are considered as potential explanatory variables for variables higher in the hierarchy.

Chapter 10 is brief 14-page introduction to generalized linear models.

The final chapter is essentially an appendix on large sample properties of maximum likelihood estimators. Its mysterious numbering arises from the fact that it is a reprinting of Chap. 15 from the author's earlier book on linear models (*Plane Answers to Complex Questions: The Theory of Linear Models*, 1987).

The choice of 'discretionary' topics covered in this book are, perhaps, of more interest to social scientists than to biostatisticians; the latter group would be more interested in nuisance parameter models and procedures which arise naturally in the stratification of retrospective and prospective data. For example, biostatisticians would find the discussion of conditional procedures such as conditional maximum likelihood estimation limited or missing.

Generating interesting problem sets is one of the most difficult parts of book writing and some instructors would find the set of problems a bit thin.

In sum, this is a well-written text with many light touches which will keep students interested. It uses many recent innovations from the statistical literature to help interpret the models and provide complete data analyses. However some advanced mathematical tools are required of the students at certain points to completely master the material; most applied M.S. students will not have this background. Lastly, some instructors will want to supplement the basic text material if they wish to consider specialized applications areas, such as biostatistics.

Columbus Th. Santner

R. V. Ambartzumian: Factorization Calculus and Geometric Probability,[1] Cambridge University Press New York, ISBN 0-521-34535-9, 1990, XI/286 pp., $59.00

Stochastic geometry deals with mathematical models for describing random geometrical structures, practical examples of which abound in all facets of everyday life. It has emerged from the theory of geometric probability and is closely related to integral geometry. The roots can be traced back to Buffon's famous needle problem in 1733.

The author of the present book, R. V. Ambartzumian, is one of the leading experts in the mentioned fields. He has influenced the development of integral and stochastic geometry by numerous, interesting results, new methods, and problems, a great part of which is presented in the monograph.

Ambartzumian has realized that many of the tools and notions of integral and stochastic geometry can be explained by the unifying concept of Haar factorization of invariant measures on product spaces with transformation group. A corresponding program is carried out in the book. The principle of Haar factorization is fully described and consequently employed for a systematic exposition of integral geometry and for the treatment of geometric random processes with laws which are invariant under the action of transformation groups.

Most of the explanations refer to the two- or three-dimensional Euclidean space, even if the treatment of higher-dimensional cases seems to be possible. In this way, a better understanding is achieved, and the reader has an intuitively geometric access to the theory. The methods used appeal to the geometric intuition.

In the chapters devoted to integral geometry, Haar measures on groups of Euclidean motions and on group of affine transformations are considered. Starting from this, invariant measures on sets of lines, planes, segments, and other elementary geometric objects are determined. Moreover, the main features of combinatorial integral geometry are developed. This discipline was established by Ambartzumian in the 1970s and was more exhaustively represented in his book "Combinatorial Integral Geometry" (Wiley, 1982). All the provided tools are used for the treatment of problems within the framework of geometric probability, and they are needed later for the investigation of geometric random processes.

Many of these processes may be interpreted as point processes on spaces of geometric objects. Accordingly, the theory of point processes including the theory of Palm distributions plays a central role in the book. Stationary point processes on Euclidean spaces as well as stationary or motion invariant line, plane, and segment processes, Boolean models, and random mosaics are treated. Models derived from Poisson point processes are of special interest. Finally, sterological problems for special processes in the plane are regarded.

---

[1] This review will also be published in "SIAM Review"

The book is highly recommended to mathematicians interested in integral and stochastic geometry. It is very stimulating because of the abundance of interesting results, models, and problems.

Jena J. Mecke

I. M. Bomze: A functional analytic approach to statistical experiments, Longman Scientific & Technical, Harlow, Essex, UK, 1990, 116 pp., £13.00

This research note is concerned with sufficiency, optimal unbiased (bounded) estimation, and invariance using the setup of L- and M-spaces introduced by Le Cam. It is essentially self-contained, but a good knowledge in Banach lattices is presumed. Although there are comprehensive books on mathematical statistics using abstract L- and M-spaces written by Le Cam (1986), Strasser (1985) and Torgersen (1991), there are many new things in Bomze's booklet, for example, most of the results are stated for generalized Banach-space-valued random elements. The research note will be of interest to researchers or graduate students who are working in this abstract setup or want to get to know something about it.

In Chap. 1, the L- and M-spaces of a statistical experiment are introduced. If the experiment is dominated by a σ-finite measure μ, L can be identified with $L^1(\mu)$ and M with $L^\infty(\mu)$. In general, L is a vector lattice of measures, and M its dual. The elements of M are "generalized (bounded) real-valued random elements". For a Banach space B, the elements of the space M^B of all continuous linear mappings from L to B are "generalized (bounded) B-valued random elements". They may also be considered as generalized bounded B-valued estimators. In order to compare such estimators by means of a convex loss function, a composition of convex functions on B and elements of M^B has to be defined which yields elements of M^B.

In the dominated case, sub-σ-algebras and conditional expectations can be identified with certain subalgebras of M and projections on M, respectively. In Chap. 2, these subalgebras and projections are defined in the general case. An inequality which replaces Jensen's inequality for conditional expectations and Banach space valued random elements is given.

In Chap. 3, sufficient sub-σ-algebras and the corresponding conditional expectations are replaced by "sufficient subalgebras" of M and "sufficient projections" on M. Among other things, in this setup a minimal sufficient subalgebra A_m always exists, and for a sufficient projection there is a Jensen-type inequality, which is useful in unbiased estimation.

Optimal unbiased estimation using estimators from M or M^B is investigated in Chap. 4. It turns out that an estimator from M is optimal iff it is contained in the "optimal algebra" $A_0 \subset M$. It is shown, for example, that the following assertions are equivalent: A_0 is sufficient; $A_m = A_0$; A_m is (boundedly) complete;

there is a complete and sufficient subalgebra in M; every (generalized) unbiased estimable parameter pocesses a (generalized) UMVA-estimator. A criterion for the existence of nontrivial optimal estimators is given.

In Chap. 5, experiments are studied which are, roughly speaking, invariant under a group G of isometric lattice isomorphismus ("transitions") on L. The algebra $A_G \subset M$ of elements which are invariant under duals of transitions from G is sufficient. Among other things, a criterion for the equality $A_0 = A_m = A_G$ is given, and it is shown how to calculate the sufficient projeciton on A_G for groups satisfying a kind of amenability property.

In an appendix, a representation of M according to Torgersen is discussed.

Hanover D. Mussmann

H. Bauer: Wahrscheinlichkeitstheorie, Walter de Gruyter, 1991, 520 pp., DM 98,—

The present book is a completely revised new edition of the well known textbook "Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie" (which appeared in English under the title "Probability Theory and Elements of Measure Theory"). The previous editions gave an introduction both to measure theory and to probability within about 400 pages. This meant that rather little space remained for probability, and perhaps the name "Measure Theory and Elements of Probability Theory" would have been more appropriate. Now, the sections on measure theory have been turned into a separate monograph of 250 pages ("Maß-und Integrationstheorie", de Gruyter, 1991), and the volume on probability now presupposes knowledge of measure theory. It can therefore treat several additional topics and present others in more depth.

After some fundamentals, chapter III discusses the law of large numbers and the almost sure convergence of series of independent random variables. It includes the Hewitt-Savage 0–1 law and the Chung-Fuchs theorem on recurrence of 1-dimensional random walks.

Chapter IV gives a modern introduction to martingales. A nice feature is a proof of the Radon-Nikodym theorem via martingales with semiordered index sets.

Chapter V develops the theory of Fourier transforms of probability measures on $\mathbb{R}^d$, the basic tool for the proof of the central limit theorem in chapter VI (Lindeberg-Feller theorem, infinitely divisible distributors, Gauß measures in $\mathbb{R}^d$). Chapter VII is devoted to the law of the iterated logarithm and includes a portion of Strassens work and Fellers proof of Strassens result that the finiteness of the variance is necessary in the law of the iterated logarithm.

Chapter VIII treats the construction of stochastic processes. Alter the general theory of semigroups of kernels and of processes with stationary independent

increments, there are sections on Brownian motion, the Poisson process, on general Markov processes (Ornstein-Uhlenbeck process) and on Gaussian processes. The final chapter IX studies Brownian motion in more depth (path properties, strong Markov property) and gives a quick introduction to stochastic integrals.

The aim of the author has been to give a guided tour which provides an overview of the field and enables the student to begin reading more advanced literature on special topics, or to explore new territories. This has been accomplished with the usual clarity of exposition for which the author is well known. The arguments are given with complete detail, and some redundancy is used on purpose to help understanding connections between the various results. Also, the structure of proofs is outlined before the steps of the proofs are carried out. Therefore, reading these 500 pages may take less time than reading a shorter but more concise presentation.

Apparently, the book does not try to give an overview over all of the principal directions of probability. For example, important topics like stationary processes, limit theorems for Markov chains, renewal theory, or central limit theorems for martingales or other dependent processes are completely absent. The choice of the material is mostly classical and biased towards results of interest to pure mathematicians (existence theorems and characterization theorems). Comparatively little is said about applications and about probability models of real phenomena. On the other hand, the new edition includes various examples of interesting probabilistic estimates, some of which not available in most other texts (Cramér-Chernoff estimates on large deviations). The section on stochastic integrals gives a very illuminating exposition of some of the fundamental ideas behind this notion, not burdened by technicalities.

On some occasions, the general existence theorems would not be needed for the examples to which they are applied. There are shorter direct arguments showing the existence of Brownian motion and of the Poisson process.

A little error occurs in the section on normal number: The author calls a number normal when the relative frequency of all digits in the decimal expansion tends to 1/10, but for one of the results he would need the more common definition requiring this for arbitrary blocks of digits.

There are numerous interesting historical remarks. A few of them are not quite precise: Borel did not propose σ-additivity (and would not have been the first to do so), and his "proof" of the strong law in the Bernoulli case was incomplete. Kolmogorov stated his strong law for variables with finite expectations in 1933; de Moivre knew Stirling's formula prior to Stirling, and Herglotz' role in the Herglotz-Bochner theorem should not be overlooked.

I propose that in future editions the author marks those sections not needed in the sequel by a star.

This book is a very remarkable and original addition to the existing textbooks on probability, and it is fun to read. It shall surely gain many friends.

Göttingen U. Krengel

Metrika

International Journal for Theoretical and Applied Statistics

Volume 40 **1993** **ISSN 0026-1335**

Physica-Verlag
A Springer-Verlag Company

© Physica-Verlag GmbH, Tiergartenstraße 17, 69121 Heidelberg, FRG, 1993
Printed in Germany
ISSN 0026-1335

Typesetting: Asco Trade Typesetting Limited, Hong Kong
Printing: Druckerei Schmidt & Sohn GmbH, Mannheim-Seckenheim
Binding: T. Gansert GmbH, Weinheim-Sulzbach

Contents

I. Publications

II. Book Reviews

Metrika (1993) 40:325–332

Metrika

Accuracy of Proportion Estimators: a Simple Rule

EUGENE F. SCHUSTER

Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, Texas 79968-0514, USA

Abstract: We consider the sample survey type problem of estimating the proportion p of a finite population of size N having a given attribute by the proportion $\hat{p}$ of successes in a random sample (with or without replacement) of size r from the population. Our main result indicates that $\hat{p} \pm 1/\sqrt{r}$ is always at least a 91.0% confidence interval (C.I.) for the parameter p. We show that $h(p) = h(p; r, N) = Prob\{|\hat{p} - p| \leq 1/\sqrt{r}\}$ is at least as large under the hypergeometric model of simple random sampling without replacement as it is under the corresponding binomial model of random sampling with replacement. The significance of our main result is that it is a good, easily stated accuracy rule, holding for all r, N, and p, which can easily be understood by the layman when assessing accuracy of the estimator $\hat{p}$ and discussing the relationship between accuracy and sample size.

Key words and phrases: binomial; hypergeometric; confidence interval; finite population

1 Introduction and Summary

In Schuster (1978), we gave a good easily stated rule regarding the accuracy of probability estimates:

Theorem 1.1: Let A be an event associated with a random experiment ξ, where $p = Prob(A)$. Let the random variable X record the number of occurrences of A among r independent trials of ξ and let $\hat{p} = X/r$. Then with probability at least 0.910 (*i.e. at least* 91.0% *of the time in the long run*) *the estimator $\hat{p}$ of p will be in the interval* $[p - 1/\sqrt{r}, p + 1/\sqrt{r}]$.

Noting that it is sufficient to consider $p \leq 0.5$, we proved Theorem 1.1 via three cases: Chebyshev's inequality easily handled $p \in [0, 0.1]$; published results on the accuracy of the normal approximation to the binomial inferred the theorem for $p \in (0.1, 0.5]$ with samples sizes $r > 200$; finally, numerical methods were used to analyze the remaining possibilities, $0.1 < p \leq 0.5$ and $1 \leq r \leq 200$.

0026–1335/93/5/325–332 $2.50

Of course, the same theorem holds when estimating a binomial proportion p by the proportion of successes $\hat{p}$ which occur in r trials, i.e., $\hat{p} \pm 1/\sqrt{r}$ is always at least a 91.0% confidence interval (C.I.) for a binomial parameter p.

Several people, working with sample survey type data, have inquired if a corresponding theorem holds in the hypergeometric case of estimating the proportion p, having a given attribute (say success) in a finite population of N, by the proportion $\hat{p}$ of successes in a simple random sample without replacement from the population. Our main result indicates that the answer to this question is in the affirmative:

Theorem 1.2: Let the random variable X record the number of successes in a random sample (with or without replacement) of size r, $1 \leq r \leq N$, from a population of size N containing n successes and let $p = n/N$, $\hat{p} = X/r$. Then with probability at least 0.910 *(i.e. at least* 91.0% *of the time in the long run) the estimator $\hat{p}$ of p will be in the interval $[p - 1/\sqrt{r}, p + 1/\sqrt{r}]$.*

Our proof of Theorem 1.2, given in Section 2, will show that $h(p) = h(p; r, N) = Prob(|\hat{p} - p| \leq 1/\sqrt{r})$ is at least as large under the hypergeometric model of simple random sampling without replacement as it is in the corresponding binomial model of random sampling with replacement. In Schuster (1978), we showed that the lowest confidence level under the binomial case of Theorem 1.1 occurs when $r = 6$ with p approaching $5/6 - 1/\sqrt{6} \doteq 0.425085$ from the left. Here the confidence level percentage dips to nearly 91.011%. Thus the 91.0% is also the best overall confidence level percentage one can achieve in both cases of Theorem 1.2. The lowest percentages again occur when $r = 6$ with N large and p approaching $5/6 - 1/\sqrt{6}$ from the left, i.e. when the hypergeometric model approaches the binomial model having the lowest confidence level.

How conservative is the 91.0% confidence bound? The Central Limit Theorem tells us that for large r and (usually unknown) p not too close to 0 or 1, $\hat{p} \pm 1/\sqrt{r}$ is about a 95% C.I. for p. In a related paper, Schuster (1992), we give a new proof of the numerical part of the proof of Theorem 1.1 which enables the exact computation of the (minimum) probability coverage

$$H_r = inf\{h(p; r) = Prob(|\hat{p} - p| \leq 1/\sqrt{r}): 0 \leq p \leq 1\}$$

in the binomial case. In the following we discuss Tables 1 and 2 from this paper. $\{H_r\}$ is not monotone in r as one might expect. There are only 2, 4, and 18 sample sizes r where the probability coverage H_r falls below 0.92, 0.93, and 0.94, respectively. Six decimal digit approximations to these 18 cases are given in Table 1. For the (1992) referenced paper, we computed values of H_r for $r \leq 700$. Except for the degenerate case $r = 1$, all values of H_r were below 0.954500 (the asymptotic value of $h(1/2; r)$ given by the normal approximation to the binomial). Moreover, we observed that picking a sample size r to be a

Table 1. Probability Coverages Below 0.94

Sample size r	$\inf_p h(p; r)$
6	0.910113
2	0.914214
12	0.922074
3	0.924501
8	0.927673
20	0.929369
5	0.930495
15	0.932670
11	0.932957
30	0.934054
24	0.935862
19	0.935884
42	0.937285
4	0.937500
29	0.938338
35	0.938548
56	0.939641
7	0.939671

Table 2. Coverages for Perfect Squares

r	$\inf h(p; r)$	r	$\inf h(p; r)$
1	1.0	196	0.954223
4	0.937500	225	0.954260
9	0.948609	289	0.954313
16	0.950958	324	0.954333
25	0.952342	361	0.954350
36	0.952969	400	0.954365
49	0.953398	441	0.954377
64	0.953647	484	0.954388
81	0.953833	529	0.954398
100	0.953956	576	0.954406
121	0.954053	625	0.954413
144	0.954123	676	0.954420
169	0.954180	∞	0.954500

perfect square, i.e. the square of an integer, elevates the probability coverage quickly to the asymptotic value. For r a perfect square larger than one, Table 2 shows that the probability coverage increases with r (at least to 700), has lowest value of 0.9375 at $r = 4$, a value of 0.948609 at $r = 9$, and is above 0.95 for squares $r \geq 16$. We believe that the pattern of Table 2 continues for larger squares r and that the minimum probability coverage is larger than 0.95 for all perfect squares larger than 9, but have no proof. For the present paper, we have used the algorithm in our 1992 paper to verify this assertion for perfect square sample sizes r to 2500 (which covers most applications). When considering theoretical models to approximate real world problems, there is often little practical difference between confidence levels 91% and 95%. Thus the significance of Theorem 1.2 is that it is a good, simple, easily stated accuracy rule, holding for all r, N, and p, which can easily be understood by the layman when assessing accuracy of estimates and discussing the relationship between accuracy and sample size.

We prove Theorem 1.2 from Theorem 1.1 and Lemmas 2.1–2.3 of Section 2 which follow from Uhlmann's (1966) relations between the binomial and hypergeometric cumulative distribution functions (c.d.f.'s). Lemma 2.2 indicates that (for fixed r, p, and N) the hypergeometric c.d.f. starts out and remains below the corresponding binomial c.d.f. until it crosses at an integer close to the common mean and then remains above until both c.d.f.'s reach one at r. This observation sheds further light on the relationship between binomial and hypergeometric probabilities, C.I.'s, and c.d.f.'s. Let $\alpha(\hat{p})$ be some function of $\hat{p}$. Then Lemma 2.3 indicates that

$$Prob(|\hat{p} - p| \leq \alpha(\hat{p})) = Prob(|X - rp| \leq r\alpha(X/r))$$

is at least as large under the hypergeometric model as under the corresponding binomial model whenever

$$\{x: |x - rp| \leq r\alpha(x/r)\}$$

contains both $[rp]$ and $[rp] + 1$ ($[\cdot]$ is the greatest integer function). This is easily seen to be the case for every p when $\alpha(\hat{p}) = 1/\sqrt{r}$.

2 Proof of Main Result

Let the random variable X record the number of successes in a random sample of r, $1 \leq r \leq N$, from a population of N containing n successes and let $p = n/N$. If sampling is with replacement, then X has the binomial distribution with

parameters r and p. If sampling is without replacement, X has the hypergeometric distribution with parameters r, n, and $m = N - n = N(1 - p)$. We will use the notation $F_B(c; r, n, N)$ and $F_H(c; r; n, N)$ for the c.d.f. of X ($(Prob(X \leq c))$ in the binomial and hypergeometric cases, respectively. Let $D(c) = D(c; r, n, N) = F_B(c; r, n, N) - F_H(c; r; n, N)$. Utilizing the relation between X and $r - X$ one can see that $D(c; r, n, N) = -D(r - c - 1; r, N - n, N)$. We first note the trivial cases:

- For $r = 1$, $D(c; r, n, N) = 0$ for every c, n and N.
- For $p = 0$ or 1, i.e., $n = 0$ or $n = N$, $D(c; r, n, N) = 0$ for every c, r and N.
- $D(r; r, n, N) = 0$ for every r, n and N.
- $D(0; r, n, N) > 0$ for every $r > 1$, and $0 < n < N$.
- $D(r - 1; r, n, N) = -D(0; r, N - n, N) < 0$ for every $r > 1$, and $0 < n < N$.

The following lemma will be used to show that the sign of $D(c)$ can be determined for all integers c except possibly at $[\mu]$ where $\mu = E(X) = rp = rn/N$.

Lemma 2.1: Assume $1 < r \leq N$ *and* $0 < n < N$. *Let* $b = (r - 1)n(N + 1)/N^2$ *and* $a = b - (r - 1)/N$. *Then for integer* c, $0 \leq c < r$,

$$D(c) = F_B(c; r, n, N) - F_H(c; r; n, N) \begin{cases} > 0 & \text{for } 0 \leq c \leq a \\ < 0 & \text{for } b \leq c < r \end{cases}.$$

Proof: Noting the trivial cases $c = 0$ and $c = r - 1$ above, the proof follows easily by inverting Uhlmann's (1966, Satz 5, p. 156) relations (restated in Johnson and Kotz (1969), p. 151) which imply that if $p = n/N$ and c is an integer, $0 < c < r - 1$ then

$$F_H(c; r, n, N) - F_B(c; r; n, N) \begin{cases} > 0 & \text{for } 0 < p \leq \dfrac{c}{r-1} - \dfrac{c}{(r-1)(N+1)} \\ < 0 & \text{for } \dfrac{c}{r-1} + \dfrac{r-c-1}{(r-1)(N+1)} \leq p < 1 \end{cases}.$$

□

Under the assumptions and notation of Lemma 2.1, one can use simple algebra to observe:

(i) $0 < b - a = (r - 1)/N < 1$.

(ii) $\mu - b = (N - r + 1)n/N^2 = p\left(1 - \dfrac{r-1}{N}\right) > 0$.

(iii) $a - (\mu - 1) = (N - r + 1)(N - n)/N^2 = q\left(1 - \dfrac{r-1}{N}\right) > 0.$

(iv) $\mu - 1 < a < b < \mu.$

Now (i) says that the interval $[a, b]$ contains at most one integer. But then (iv) indicates that the only possible integer in $[a, b]$ is $[\mu]$.

We can now use Lemma 2.1 to analyze the sign of $D(c)$ depending upon the three cases for the possible position of $[\mu]$ with respect to the interval $[a, b]$ to arrive at:

Lemma 2.2: Assume r, n, N are integers and c is any real number with $1 < r \leq N$, $0 < n < N$ and $0 \leq c < r$.

Case 1: If $[\mu] \geq b$, then the sign of $D(c)$ is completely determined with

$D(c) > 0$ *for* $c < [\mu]$ *and* $D(c) < 0$ *for* $c \geq [\mu]$.

(*Note that this case includes the cases where μ is an integer.*)

Case 2: If $[\mu] \leq a$, then the sign of $D(c)$ is completely determined with

$D(c) > 0$ *for* $c < [\mu] + 1$ *and* $D(c) < 0$ *for* $c \geq [\mu] + 1$.

Case 3: If $a < [\mu] < b$, then $D([\mu])$ can be positive, negative, or zero, but

$D(c) > 0$ *for* $c < [\mu]$ *and* $D(c) < 0$ *for* $c \geq [\mu] + 1$.

(*Examples where $D([\mu])$ is positive, negative, and zero are given in Table 3.*)

The symmetry in the cases of Lemma 2.2 is easily explained by the identity $D(c; r, n, N) = -D(r - c - 1; r, N - n, N)$. The following Table gives examples of the various cases in Lemma 2.2.

Table 3.

r	n	N	a	b	μ	$[\mu]$	$D([\mu])$	Case
8	5	10	3.15	3.85	4.0	4	−0.1411	Case 1, $\mu = [\mu]$
8	9	10	6.23	6.93	7.2	7	−0.2305	Case 1, $\mu > [\mu] > b$
8	1	10	0.07	0.77	0.8	0	0.2305	Case 2, $[\mu] < a$
8	6	10	3.92	4.62	4.8	4	0.0726	Case 3, $a < [\mu] < b$
8	4	10	2.38	3.08	3.2	3	−0.0726	Case 3, $a < [\mu] < b$
5	5	10	1.80	2.2	2.5	2	0	Case 3, $a < [\mu] < b$

Lemma 2.3: Assume l and u are any real numbers with $l < u$. If $[\mu]$ and $[\mu] + 1$ are both in the (half-open) interval $(l, u]$, then

$$F_H(u; r, n, N) - F_H(l; r; n, N) \geq F_B(u; r, n, N) - F_B(l; r; n, N) \ .$$

Proof: For $1 < r \leq N$, $0 < n < N$ and $0 \leq c < r$ Lemma 2.2 indicates that

$$D(c) = F_B(c; r, n, N) - F_H(c; r; n, N) \begin{cases} > 0 & \text{for } c < [\mu] \\ < 0 & \text{for } [\mu] + 1 \leq c < r \ . \end{cases}$$

The proof is easily completed by noting the trivial cases discussed before the statement of Lemma 2.1. □

Proof of Theorem 1.2: Since $r \geq 1$, the proof of Theorem 1.2 follows from Lemma 2.3 by noting that

$$Prob(|X/r - p| \leq 1/\sqrt{r}) = Prob(|X - rn/N| \leq \sqrt{r})$$

and that the interval

$$\{x: |x - rp| \leq \sqrt{r}\} = [rp - \sqrt{r}, rp + \sqrt{r}]$$

will always contain $[\mu] = [rp]$ and $[\mu] + 1 = [rp] + 1$ for every choice of r, N, and $p = n/N$.

Remark: Note that Lemma 2.3 indicates that the hypergeometric probability of any interval I containing both $[\mu]$ and $[\mu] + 1$ is at least as large as the corresponding binomial probability. The question then arises: does the common textbook confidence interval for p, $\hat{p} \pm z\sqrt{\hat{p}(1 - \hat{p})/r}$ have at least as large a hypergeometric probability as it does a binomial probability? If it does not, then there must exist r, N, p and number z such that the interval

$$I = \{x: |x - rp| \leq z\sqrt{x(r - x)/r}\}$$

contains exactly one of $[\mu]$ and $[\mu] + 1$, say a, and the binomial probability of a is larger than the hypergeometric probability. Our *Mathematica* program did not find any such cases for $N \leq 100$. We do not believe that there are any such cases for $N > 100$, however we have been unable to prove this.

Acknowledgement: The author would like to thank the referee for constructive comments which clarified the presentation in this paper.

References

[1] Johnson NL, Kotz S (1969) Discrete Distributions. Houghton Mifflin, Boston
[2] Schuster EF (1978) On the accuracy of probability estimates. Mathematics Magazine 51:227–229
[3] Schuster EF (1992) On the accuracy of binomial and proportion estimators: an absolute rule. To appear in Computing Science and Statistics Proceedings of the 24th Symposium on the Interface, Interface Foundation of North America, Fairfax Station, VA
[4] Uhlmann W (1966) Vergleich der hypergeometrischen mit der Binomial-Verteilung. Metrika 10:145–158

Received 11.02.1992
Revised version 11.12.1992

Metrika (1993) 40:333–338

Metrika

Equidistribution Properties of Nonlinear Congruential Pseudorandom Numbers

JÜRGEN EICHENAUER–HERRMANN

Fachbereich Mathematik, Technische Hochschule Darmstadt, Schloßgartenstr. 7, 64289 Darmstadt, Germany

Abstract: Equidistribution properties of a general class of nonlinear congruential methods for generating uniform pseudorandom numbers are established. The results that are obtained are essentially best possible and show that the generated pseudorandom numbers have an equidistribution behaviour like true random numbers.

1 Introduction and Main Results

Recently, several nonlinear congruential methods of generating uniform pseudorandom numbers in the interval [0, 1) have been introduced and analysed, which show more favourable properties than the classical methods. A review of the development of this area is given in the survey articles of Niederreiter (1991, 1992a, 1992c) and of Eichenauer–Herrmann (1992) and in the excellent monograph of Niederreiter (1992b).

In the present paper the following general class of nonlinear congruential generators is considered, which has been introduced and analysed with respect to Marsaglia's lattice test in Eichenauer et al. (1988) and in Niederreiter (1988a). Let $p \geq 5$ be a prime and identify $\mathbb{Z}_p := \{0, 1, \ldots, p-1\}$ with the finite field of order p. Let $\gamma \in \mathbb{Z}_p^* := \mathbb{Z}_p \setminus \{0\}$ and $g: \mathbb{Z} \to \mathbb{Z}_p$ be a monic permutation polynomial of $\mathbb{Z}_p$ with degree s as a polynomial over $\mathbb{Z}_p$, where $3 \leq s \leq p-2$ is assumed in order to avoid uninteresting cases. A *nonlinear congruential sequence* $(y_n)_{n \geq 0}$ of elements of $\mathbb{Z}_p$ is generated by

$$y_n \equiv \gamma g(n) \pmod{p}, \qquad n \geq 0 .$$

A sequence $(x_n)_{n \geq 0}$ of *nonlinear congruential pseudorandom numbers* in the interval [0, 1) is obtained by the normalization $x_n = y_n/p$ for $n \geq 0$. Obviously, the sequence $(y_n)_{n \geq 0}$ is purely periodic with maximal period length p, i.e., $\{y_0, y_1, \ldots, y_{p-1}\} = \mathbb{Z}_p$.

A basic condition for the quality of the generated pseudorandom numbers is an appropriate equidistribution of parts of the period, which can be assessed

0026–1335/93/5/333–338 $2.50

by the discrepancy of the corresponding point sets. For N arbitrary points t_0, $t_1, \ldots, t_{N-1} \in [0, 1)$ the *discrepancy* is defined by

$$D_N(t_0, t_1, \ldots, t_{N-1}) := \sup_{0 \le \alpha < \beta \le 1} |F_N([\alpha, \beta)) - (\beta - \alpha)| ,$$

where $F_N([\alpha, \beta))$ is N^{-1} times the number of points among $t_0, t_1, \ldots, t_{N-1}$ falling into $[\alpha, \beta)$. For a sequence $(x_n)_{n \ge 0}$ of nonlinear congruential pseudorandom numbers the abbreviation

$$D_N := D_N(x_0, x_1, \ldots, x_{N-1})$$

is used. The following upper bound for D_N slightly improves upper bounds resulting from Theorem 2 in Niederreiter (1988b) and Theorem 2 in Eichenauer–Herrmann and Niederreiter (1992). In the third section the proof of Theorem 1 is sketched for the sake of completeness.

Theorem 1: Let $1 \le N < p$. Then the discrepancy D_N for any nonlinear congruential generator satisfies

$$D_N < (s-1)\frac{p^{1/2}}{N}\left(\frac{4}{\pi^2}\log p + 0.38 + \frac{0.608}{p} + \frac{0.116}{p^2}\right)^2 + \frac{1}{p} .$$

In the following main result of the present paper a lower bound for D_N is established. Its proof is given in the third section, too. For $1 \le N < p$ and $0 < t \le \frac{1}{2}\sqrt{\frac{N(p-N)}{p(p-1)}}$ let

$$A_N(t) := \frac{\frac{N(p-N)}{p(p-1)} - 4t^2}{(s-1)^2\left(\frac{4}{\pi^2}\log p + 0.38 + \frac{0.608}{p} + \frac{0.116}{p^2}\right)^2 - 4t^2} .$$

Theorem 2: Let $1 \le N < p$ and $0 < t \le \frac{1}{2}\sqrt{\frac{N(p-N)}{p(p-1)}}$. Then for any monic permutation polynomial g there are more than $A_N(t)(p-1)$ values of $\gamma \in \mathbb{Z}_p^*$ such that the discrepancy D_N of the corresponding nonlinear congruential generator satisfies

$$D_N \ge t\frac{p^{1/2}}{N} .$$

Theorem 1 shows that $D_N = O(N^{-1}p^{1/2}(\log p)^2)$ for any nonlinear congruential generator, where the implied constant is absolute. It follows from Theorem 2 that the upper bound in Theorem 1 is in general best possible up to the logarithmic factor, since there exist nonlinear congruential generators with discrepancy D_N of the order of magnitude at least $N^{-1}p^{1/2}$.

If $N \approx \alpha p$ with $0 < \alpha < 1$, then Theorems 1 and 2 imply that for a positive proportion of the nonlinear congruential generators the discrepancy D_N is of an order of magnitude between $p^{-1/2}$ and $p^{-1/2}(\log p)^2$. It is in this range of magnitudes where one also finds the discrepancy of true random numbers (cf. Chung 1949).

2 Auxiliary Results

From now on, the abbreviations $e(t) = e^{2\pi it}$ for real numbers t and $\chi(z) = e(z/p)$ for integers z are used. Subsequently, two known results are stated. The first lemma is a special case of a classical result of Weil (1948) (see also Theorem 5.38 in Lidl and Niederreiter 1983) and the second lemma follows from Theorem 1 in Cochrane (1987).

Lemma 1: Let $Q: \mathbb{Z}_p \to \mathbb{Z}_p$ be a polynomial with $1 \leq deg(Q) < p$. Then

$$\left| \sum_{n \in \mathbb{Z}_p} \chi(Q(n)) \right| \leq (deg(Q) - 1)p^{1/2} .$$

Lemma 2: Let N be an integer. Then

$$\frac{1}{p} \sum_{h \in \mathbb{Z}_p^*} \left| \frac{\sin(\pi hN/p)}{\sin(\pi h/p)} \right| < \frac{4}{\pi^2} \log p + 0.38 + \frac{0.608}{p} + \frac{0.116}{p^2} .$$

A crucial role in the following presentation is played by certain exponential sums. Let $1 \leq N < p$ and g be a given monic permutation polynomial as above. Then define

$$S_{N,g}(\gamma) := \sum_{n=0}^{N-1} \chi(\gamma g(n))$$

for integers γ.

Lemma 3: Let $\gamma \in \mathbb{Z}_p^*$. *Then*

$$|S_{N,g}(\gamma)| < (s-1)p^{1/2}\left(\frac{4}{\pi^2}\log p + 0.38 + \frac{0.608}{p} + \frac{0.116}{p^2}\right).$$

Proof: First, it follows as in the proof of Theorem 2 in Niederreiter (1988b) that

$$\begin{aligned} S_{N,g}(\gamma) &= \sum_{n \in \mathbb{Z}_p} \chi(\gamma g(n)) \sum_{j=0}^{N-1} \frac{1}{p} \sum_{h \in \mathbb{Z}_p} \chi(h(j-n)) \\ &= \frac{1}{p} \sum_{h \in \mathbb{Z}_p^*} \left(\sum_{n \in \mathbb{Z}_p} \chi(\gamma g(n) - hn) \right) \left(\sum_{j=0}^{N-1} \chi(hj) \right), \end{aligned}$$

where in the last step $h = 0$ can be omitted since g is a permutation polynomial. Again, as in Niederreiter (1988b), by applying Lemmas 1 and 2 one obtains

$$\begin{aligned} |S_{N,g}(\gamma)| &\le \frac{1}{p} \sum_{h \in \mathbb{Z}_p^*} \left| \sum_{n \in \mathbb{Z}_p} \chi(\gamma g(n) - hn) \right| \left| \sum_{j=0}^{N-1} \chi(hj) \right| \\ &\le (s-1)p^{-1/2} \sum_{h \in \mathbb{Z}_p^*} \left| \frac{\sin(\pi h N/p)}{\sin(\pi h/p)} \right| \\ &< (s-1)p^{1/2}\left(\frac{4}{\pi^2}\log p + 0.38 + \frac{0.608}{p} + \frac{0.116}{p^2}\right). \end{aligned}$$

□

Lemma 4: There holds

$$\sum_{\gamma \in \mathbb{Z}_p^*} |S_{N,g}(\gamma)|^2 = N(p-N).$$

Proof: Since g is a permutation polynomial, it follows at once that

$$\begin{aligned} \sum_{\gamma \in \mathbb{Z}_p^*} |S_{N,g}(\gamma)|^2 &= \sum_{\gamma \in \mathbb{Z}_p} |S_{N,g}(\gamma)|^2 - N^2 \\ &= \sum_{\gamma \in \mathbb{Z}_p} \sum_{n,m=0}^{N-1} \chi(\gamma(g(n) - g(m))) - N^2 \\ &= \sum_{n,m=0}^{N-1} \sum_{\gamma \in \mathbb{Z}_p} \chi(\gamma(g(n) - g(m))) - N^2 \\ &= pN - N^2 = N(p-N). \end{aligned}$$

□

3 Proof of the Theorems

Proof of Theorem 1: First, it follows from (3.12) in the proof of Theorem 3.10 in Niederreiter (1992b) that

$$D_N < \frac{1}{p} + \frac{1}{Np} \max_{\substack{a,b \in \mathbb{Z}_p \\ a \le b}} \sum_{h \in \mathbb{Z}_p^*} \left| \sum_{j=a}^{b} \chi(hj) \right| \left| \sum_{n=0}^{N-1} e(hx_n) \right|$$

$$= \frac{1}{p} + \frac{1}{Np} \max_{\substack{a,b \in \mathbb{Z}_p \\ a \le b}} \sum_{h \in \mathbb{Z}_p^*} \left| \sum_{j=a}^{b} \chi(hj) \right| |S_{N,g}(h\gamma)| \ .$$

Now, Lemma 3 can be used in order to obtain

$$D_N < \frac{1}{p} + \frac{s-1}{Np^{1/2}} \left(\frac{4}{\pi^2} \log p + 0.38 + \frac{0.608}{p} + \frac{0.116}{p^2} \right) \max_{\substack{a,b \in \mathbb{Z}_p \\ a \le b}} \sum_{h \in \mathbb{Z}_p^*} \left| \sum_{j=a}^{b} \chi(hj) \right| \ .$$

Finally, it follows as in the proof of Lemma 3 by applying Lemma 2 that

$$D_N < \frac{1}{p} + \frac{s-1}{Np^{1/2}} \left(\frac{4}{\pi^2} \log p + 0.38 + \frac{0.608}{p} + \frac{0.116}{p^2} \right)$$

$$\times \max_{\substack{a,b \in \mathbb{Z}_p \\ a \le b}} \sum_{h \in \mathbb{Z}_p^*} \left| \frac{\sin(\pi h(b-a+1)/p)}{\sin(\pi h/p)} \right|$$

$$< \frac{1}{p} + (s-1) \frac{p^{1/2}}{N} \left(\frac{4}{\pi^2} \log p + 0.38 + \frac{0.608}{p} + \frac{0.116}{p^2} \right)^2 \ . \qquad \square$$

Proof of Theorem 2: First, it follows from Lemma 1 in Niederreiter (1990) that

$$D_N \ge \frac{1}{2N} \left| \sum_{n=0}^{N-1} e(x_n) \right| = \frac{1}{2N} |S_{N,g}(\gamma)| \ .$$

Hence, it suffices to show that for any given monic permutation polynomial g there are more than $A_N(t)(p-1)$ values of $\gamma \in \mathbb{Z}_p^*$ with $|S_{N,g}(\gamma)| \ge 2tp^{1/2}$. This property can be proved by contradiction, since otherwise Lemma 3 implies

that

$$\sum_{\gamma \in \mathbb{Z}_p^*} |S_{N,g}(\gamma)|^2 < (1 - A_N(t))(p-1)4t^2p + A_N(t)(p-1)(s-1)^2p \times \left(\frac{4}{\pi^2}\log p + 0.38 + \frac{0.608}{p} + \frac{0.116}{p^2}\right)^2 = N(p-N)$$

which contradicts Lemma 4. □

Acknowledgement: The author would like to thank the referee for valuable comments.

References

Chung KL (1949) An estimate concerning the Kolmogoroff limit distribution, Trans. Amer. Math. Soc. 67:36–50

Cochrane T (1987) On a trigonometric inequality of Vinogradov, J. Number Th. 27:9–16

Eichenauer J, Grothe H, Lehn J (1988) Marsaglia's lattice test and non-linear congruential pseudo random number generators, Metrika 35:241–250

Eichenauer-Herrmann J (1992) Inversive congruential pseudorandom numbers: a tutorial, Int. Statist. Rev. 60:167–176

Eichenauer-Herrmann J, Niederreiter H (1992) On the statistical independence of nonlinear congruential pseudorandom numbers (submitted for publication)

Lidl R, Niederreiter H (1983) Finite fields, Addison-Wesley, Reading, Mass.

Niederreiter H (1988a) Remarks on nonlinear congruential pseudorandom numbers, Metrika 35: 321–328

Niederreiter H (1988b) Statistical independence of nonlinear congruential pseudorandom numbers, Monatsh. Math. 106:149–159

Niederreiter H (1990) Lower bounds for the discrepancy of inversive congruential pseudorandom numbers, Math. Comp. 55:277–287

Niederreiter H (1991) Recent trends in random number and random vector generation, Ann. Operations Res. 31:323–345

Niederreiter H (1992a) Nonlinear methods for pseudorandom number and vector generation. In: Pflug, G. and Dieter, U. (eds.) Simulation and Optimization, Lecture Notes in Economics and Math. Systems 374:145–153, Springer, Berlin

Niederreiter H (1992b) Random number generation and quasi-Monte Carlo methods, SIAM, Philadelphia

Niederreiter H (1992c) Pseudorandom numbers and quasirandom points, Z. Angew. Math. Mech. (to appear)

Weil A (1948) On some exponential sums, Proc. Nat. Acad. Sci. U.S.A. 34:204–207

Received 09.11.92
Revised version 10.02.93

Metrika (1993) 40:339–348

Metrika

A Semiparametric Technique to Estimate Survival Probabilities

NADER EBRAHIMI

Division of Statistics, Department of Mathematical Sciences, Northern Illinois University, DeKalb, Illinois 60115-2854, USA

Abstract: The role of the so-called surplus processes in the assessment of probability of survival of a company is well-known in risk theory and applications thereof. However, the insurance models used in this regard ignore the fact that, in many situations, no relevant information is available for the assessment of survival after the company goes out of business. In this paper, we revisit the classical risk model in order to remedy this situation. Having stopped the deficit process, which is negative of the surplus process, at the time of ruin, under two different sampling schemes, we obtain inference procedure for ruin probabilities. As by products of our methodology, we also obtain procedures to assess the reliability of systems whose survival depends on a cumulative damage process, which is equivalent to the aggregate claim size process of the classical risk model.

Key Words and Phrases: Counting process, Compound Poisson process, Consistent estimator, Damage process.

1 Introduction

In the classical risk model (Seal (1978) and Grandell (1991)) of a company that deals with non-life insurance, the surplus process, $Y(t)$, is given by

$$Y(t) = (ct + a) - Z(t) \ , \tag{1.1}$$

where

$$Z(t) = \sum_{i=1}^{N(t)} X_i \ , \tag{1.2}$$

c denoting the units of premium money that the company receives per unit time, a denoting the initial risk reserve of the company, $N(t)$ being the number of claims on the company during the time period $(0, t]$ and X_i's being the successive positive stochastic amounts of money paid out by the company. Although c is positive in the insurance applications of this model, we shall let c be non-negative in order to subsume a special case, which we describe in the next

0026–1335/93/5/339–348 $2.50

paragraph. For convenience, we shall use the deficit process, $W(t)$, which is given by

$$W(t) = -Y(t) \ . \tag{1.3}$$

Clearly, in this model, the company is ruined at the random time

$$T = T(a, c) = \inf\{t: W(t) > 0\} \tag{1.4}$$

The finite-horizon and the infinite-horizon survival probabilities, $\Psi(a, t)$, $\Psi(a)$, can now be defined respectively by

$$\Psi(a, t) = P(T > t) \tag{1.5}$$

$$\Psi(a) = P(T = \infty) \tag{1.6}$$

The classical risk theory deals with the study of $\Psi(a, t)$ and $\Psi(a)$, which is well documented in Grandell (1991) and Seal (1978).

A special case of (1.3) and (1.4) in which $c = 0$ provides a useful model of cumulative damages which can be used to study the reliability of various systems. For example, if $N(t)$ denotes the number of shocks to a system during $(0, t]$, X_i's are the successive shocks inflicted upon the system, and if the system fails when the sum of the shocks exceeds a known threshold a, then the time to system failure, S_1, can be written as

$$S_1 = \inf\{t: Z(t) > a\} \tag{1.7}$$

In survival analysis, one is interested in the probability, $\bar{F}_1(t)$, that the system survives beyond time t, namely

$$\bar{F}_1(t) = P(S_1 > t) \tag{1.8}$$

For more details regarding (1.8), see Barlow and Proschan (1981).

Remark 1: There are situations in which there is no practical way to inspect a system to determine its threshold a. In this case the failure time of a system S_2 say, is

$$S_2 = \inf\{t: Z(t) > V\} \ , \tag{1.9}$$

where V is a positive random threshold, with a known probability density function $f_V(x)$, that is independent of the process $Z(t)$. For more details see Esary, Marhsall and Proschan (1973). For systems of this type, the survival, $\bar{F}_2(t)$, is given by

$$\bar{F}_2(t) = P(S_2 > t) \tag{1.10}$$

Often, we do not know the structure of the damage process $Z(t)$ in (1.2) and of the deficit process $W(t)$ in (1.3), but would like to assess, in advance of ruin(failure), $\Psi(a, t)(\bar{F}_1(t))$, given the current deficit(damage) $W(s)(Z(s))$, up to a time $s < t$, and\or the number of claims up to a time $s < t$, i.e. $N(s)$. The methodologies that currently exist in the literature on risk theory and survival analysis do not address these assessment issues adequately for situations in which the following crucial assumptions hold:

(i) No relevant information regarding survival is available after the failure of the company(system), which stops honoring the payments on the claims-(operating) at the time of ruin(failure), $T(S_1$ or $S_2)$.
(ii) Claims on the company(shocks to the system) are likely to arrive beyond the time $T(S_1$ or $S_2)$.

Consequently, our main goal in this paper will be to develop procedures to assess $\Psi(a, t)(\bar{F}_1(t))$, using sample paths of the process $N(t)$ and the following stopped deficit(damage) process:

$$U(t) = \begin{cases} W(t) , & \text{if } t < T \\ 0 , & \text{if } t \geq T \end{cases}, \tag{1.11}$$

$$U_1(t) = \begin{cases} Z(t) , & \text{if } t < S_1 \\ a , & \text{if } t \geq S_1 \end{cases}, \tag{1.12}$$

Procedures to assess the reliabilities of systems using $\bar{F}_2(t)$ in (1.10) will also be outlined.

In order to keep our discussions non-trivial, unless stated otherwise, we shall always assume that $P(N(0) = 0) = 1$ and $P(Z(0) = 0) = 1$ and that a version of $Z(t)$ process, whose sample paths are right-continuous with finite left limits at all times, is available. Finally, we consider in this paper:

(i) The process $\{N(t), t > 0\}$ is homogeneous Poisson with intensity $\lambda > 0$.
(ii) The sequence $\{X_i, i \geq 1\}$ consists of *i.i.d.* random variables with common pdf $g(x)$, common CDF $G(x)$ and common survival function $\bar{G} = 1 - G$.
(iii) The processes $\{N(t), t \geq 0\}$ and $\{X_i, i \geq 1\}$ are independent.

Now, we describe two schemes of collecting data.

Scheme A: For estimating $\bar{F}_1(t)$, provided by (1.8), we observe k independent copies of stopped damage process and number of shocks process. The i-th replication is observed at pre-determined epochs. More specifically, we let subscript $i = 1, \ldots, k$ refer to the label of the system and subscript $j = 1, \ldots, M_i$, denote the label of proposed observations over time for the i-th system. For the j-th observation on the i-th system, let t_{ij} be the time from the initial observation date, $t_{i0} = 0$. It is assumed that all systems are initially working. It is important to note that if $U_1(t_{ij}) < a$, the i-th system is working at time t_{ij}. However, if for a given j, $U_1(t_{i,j-1}) < a$ and $U_1(t_{ij}) = a$, then we stop gathering data from the failed system and record the actual time to failure of the i-th system S_{1i}, where $t_{i,j-1} < S_{1i} \leq t_{ij}$. Furthermore, we relabel j as m_i and the corresponding proposed observation time, t_{im_i} as S_{1i}. It should be noted that $U_1(t_{il}) = a$, $l = m_i, \ldots, M_i$. Thus, for $i = 1, \ldots, k$, we observe $U_1(t_{ij})$, $j = 1, \ldots, M_i$, $L_i = \min(S_{1i}, t_{iM_i})$, and $N(t_{ij})$, $j = 1, \ldots, M_i$.

For estimating $\psi(a, t)$, provided by (1.5), we observe $U(t_{ij})$, $L_i = \min(T_{1i}, t_{iM_i})$ and $N(t_{ij})$, $i = 1, \ldots, k$, $j = 1, \ldots, M_i$. Here T_{1i} is the actual ruin time of the i-th company.

Scheme B: For estimating $\bar{F}_1(t)$ we only observe $U_1(t_{ij})$ and $L_i = \min(S_{1i}, t_{iM_i})$, $i = 1, \ldots, k$, $j = 1, \ldots, M_i$. For estimating $\Psi(a, t)$ we only observe $U(t_{ij})$ and $L_i = \min(T_{1i}, t_{iM_i})$, $i = 1, \ldots, k$, $j = 1, \ldots, M_i$.

This paper is organized as follows. In Section 2, we derive a semiparametric estimations of $\bar{F}_1(t)$ as well as $\Psi(a, t)$ under the sampling Scheme A. A semiparametric estimations of $\bar{F}_1(t)$ and $\Psi(a, t)$ under sampling Scheme B were obtained in Section 3.

2 Estimation of Survival Probabilities: Scheme A

The log-likelihood function of the data $N(t_{ij}) = n_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, M_i$, $l_A(\lambda)$ can be obtained using the structure of the Poisson process $\{N(t), t \geq 0\}$. That is,

$$l_A(\lambda) = \sum_{i=1}^{k} [n_{i,M_i} \log \lambda - \lambda t_{i,M_i}] + d \ , \tag{2.1}$$

where the term d is free from λ. In order to maximize the log-likelihood in (2.1), we consider

$$\frac{dl_A(\lambda)}{d\lambda} = \sum_{i=1}^{k} \left(\frac{n_{i,M_i}}{\lambda} - t_{i,M_i} \right) = 0 \ . \tag{2.2}$$

Solving Equation (2.2), the maximum likelihood estimate of λ is

$$\hat{\lambda}_A = \frac{\sum_{i=1}^{k} n_{i,M_i}}{\sum_{i=1}^{k} t_{i,M_i}} \tag{2.3}$$

Now, we start with describing our inference procedure for $\bar{F}_1(t_0)$, the reliability of a system over a specified time period $[0, t_0)$.

It is known that

$$E(Z(t)) = \mu\lambda t \ , \tag{2.4}$$

and

$$K(s, t) = \mathrm{Cov}(Z(s), Z(t)) = (\sigma^2 + \mu^2)\lambda \min(t, s) \ , \tag{2.5}$$

where $\mu = \int_0^\infty \bar{G}(x)\,dx$, $\sigma^2 = -\mu^2 + 2\int_0^\infty x\bar{G}(x)\,dx$. Given that for the ith system we observe $U_1(t_{i1}), \ldots, U_1(t_{i,M_i})$, $i = 1, \ldots, k$, our first goal is to predict $Z_i(t_0)$, the total damages to the i-th system by time t_0, $i = 1, \ldots, k$. Clearly if $U_1(t_{iM_i}) = a$, then $Z(t_{iM_i}) > a$ and if $U_1(t_{iM_i}) < a$, then $U_1(t_{ij}) = Z(t_{ij}) = z_{ij}$, $j = 1, \ldots, M_i$. For the latter case, a commonly used class of predictors are linear combinations of the form

$$\hat{Z}_i(t_0) = \sum_{j=1}^{M_i} a_j Z(t_{ij}) \ . \tag{2.6}$$

The Kriging predictor, $\hat{Z}_i(t_0)$, is the unbiased linear predictor that minimizes the variance of the prediction error. Kriging is a method of prediction for random field popular in mining and hydrology. (See Rendu (1978), Journel and Huijbergts (1978) and Kitanidis (1983).) The weights $a(i) = (a_1, \ldots, a_{M_i})'$ defining $\hat{Z}_i(t_0)$ are given by

$$a(i) = C^{-1}(i)c(i) + C^{-1}(i)D'(i)(D(i)C^{-1}(i)D(i))^{-1}[t_0 - D(i)C^{-1}(i)c(i)] \ , \tag{2.7}$$

where $D(i) = [t_{i1}, \ldots, t_{i,M_i}]$, $c(i) = [K(t_0, t_{i1}), \ldots, K(t_0, t_{i,M_i})]'$ and $C(i)$ is $M_i \times M_i$ matrix with (j, l)-th element $K(t_{ij}, t_{il})$, $j, l = 1, \ldots, M_i$, $c(i)$ has full rank and $(a_1, \ldots, a_{M_i})'$ is the transpose of the vector $(a_1, \ldots, a_{M_i})$. (See Goldberger, (1962)). The prediction error is denoted by $e_i(t_0) = \hat{Z}_i(t_0) - Z_i(t_0)$ and its variance is $\sigma^2(i)$ where

$$\sigma^2(i) = K(t_0, t_0) - c'(i)C^{-1}(i)c(i)[t_0 - D(i)C^{-1}(i)c(i)]$$

$$\times (D(i)C^{-1}(i)D'(i))^{-1}(t_0 - D(i)C^{-1}(i)c(i))$$

The covariance function $K(.,.)$ is not specified and must be estimated from data. Using (2.5) we propose to estimate $K(.,.)$ by

$$\hat{K}(t_{il}, t_{ij}) = (\hat{\mu}^2 + \hat{\sigma}^2)\hat{\lambda}_A t_{il} , \qquad j \geq l ,$$

where unbiased estimators of μ and σ^2 are

$$\hat{\mu} = \frac{1}{N_1} \sum_{i=1}^{k} \sum_{j=1}^{d_i} w_{ij} , \tag{2.8}$$

$$\hat{\sigma}^2 = \frac{1}{N_1 - 1} \sum_{i=1}^{k} \sum_{j=1}^{d_i} (w_{ij} - \hat{\mu})^2 , \tag{2.9}$$

respectively. Here

$$d_i = \sum_{j=1}^{M_i} I(n_{ij} > n_{i,j-1}) , \qquad w_{ij} = \frac{1}{n_{ij} - n_{i,j-1}} (S_{n_{ij}} - S_{n_{i,j-1}}) ,$$

$N_1 = \sum_{i=1}^{k} d_i$, $S_{n_{ij}}$ is the total damages received by the ith system at time t_{ij}, and I is the indicator function.

Now, using (2.6) we estimate $\bar{F}_1(t_0) = P(Z(t_0) \geq a)$ by

$$\hat{\bar{F}}_1(t_0) = \frac{1}{k} \sum_{i=1}^{k} \{I(\hat{Z}_i(t_0) \geq a) I(Z_i(t_{i_{M_i}}) < a) + I(Z_i(t_{i_{M_i}}) \geq a))(I(S_{1i} > t_0))\} . \tag{2.10}$$

Turning now to the estimation of $\bar{F}_2$ in (1.10),

$$\bar{F}_2(t) = P(S_2 \geq t) = P(Z(t) \leq V) = \int_0^\infty P(Z(t) < v) f_V(v)\, dv . \tag{2.11}$$

Now, for any fixed but arbitrary $v \leq 0$, one can estimate $P(Z(t) \leq v)$ using techniques described earlier in this section (threshold known case) and consequently $\bar{F}_2(t)$ can be estimated through equation (2.11).

We close this part with a discussion of the asymptotic properties of $\hat{\bar{F}}_1(t_0)$. For this part, we shall assume that $M_1 = M_2 = \cdots = M_k = M$ and the data are equally spaced, that is $\Delta_{ij} = t_{ij}$, $-t_{i,j-1} = \Delta$, $i = 1, \ldots, k$, $j = 1, \ldots, M$.

Theorem 1: As $k \to \infty$, then

(a) $\hat{\bar{F}}_1(t_0)$ is a consistent estimator of $\bar{F}_1(t_0)$

(b) $\sqrt{k}(\hat{\bar{F}}_1(t_0) - \bar{F}_1(t_0))$ is asymptotically normal with mean zero and variance $\bar{F}_1(t_0)(1 - \bar{F}_1(t_0))$.

Proof: It is clear that, in view of the assumptions of equally-spaced data and all M_i being equal, our proposed estimator is based on k i.i.d. copies of the random vector $(U_1(\Delta), \ldots, U_1(M\Delta), N(\Delta), N(2\Delta), \ldots, N(M\Delta), L)$, where $L = \min(S_1, M\Delta)$, and therefore k i.i.d. copies of $V = I(\hat{Z}(t_0) \geq a)I(Z(M\Delta) < a) + I(S_1 > a)I(Z(M\Delta) \geq a)$. Using central limit theorem, we get the result.

Using part (b) of Theorem 1, $(1 - \alpha)\%$ confidence interval for $\bar{F}_1(t_0)$ is

$$\hat{\bar{F}}_1(t_0) \pm z_{1-(\alpha/2)} \sqrt{\frac{\hat{\bar{F}}_1(t_0)(1 - \hat{\bar{F}}_1(t_0))}{k}} , \tag{2.12}$$

where z_α is α-th percentile of standard normal distribution.

Now, we turn to estimation of $\Psi(a, t)$. In general it should be noted that if $U(t_{i_{M_i}}) < 0$, then $U(t_{ij}) = W(t_{ij})$ and therefore $Z(t_{ij}) = W_{ij} + (ct_{ij} + a)$, $j = 1, \ldots, M_i$. If $U(t_{i_{M_i}}) = 0$, then $Z(t_{i_{M_i}}) > ct_{i_{M_i}} + a$. Using this fact we propose $\hat{\Psi}(a, t_0)$ as an estimator of $\Psi(a, t_0)$, where

$$\hat{\Psi}(a, t_0) = \frac{1}{k} \sum_{i=1}^{k} I(\hat{Z}_i(t_0) > ct_0 + a)I(Z_i(t_{i_{M_i}}) < ct_{i_{M_i}} + a)$$

$$+ I(U(t_{i_{M_i}}) = 0)I(T_i > t_0)] , \tag{2.13}$$

where $\hat{Z}_i(t_0)$ is given by the equation (2.6). It can be shown that as $k \to \infty$, $\sqrt{k}(\hat{\Psi}(a, t_0) - \Psi(a, t_0))$ is asymptotically normal with mean zero and variance $\Psi(a, t_0) \times (1 - \Psi(a, t_0))$. Also $(1 - \alpha)\%$ confidence interval for $\Psi(a, t_0)$ is

$$\hat{\Psi}(a, t_0) \pm z_{1-(\alpha/2)} \sqrt{\frac{\hat{\Psi}(a, t_0)(1 - \hat{\Psi}(a, t_0))}{k}}$$

Now, to see how our proposed estimators perform, we did Monte Carlo studies of biases and mean squared errors. In our limited studies we assume $G(x) = 1 - \exp(-x)$, and that the shocks are arriving at the rate $\lambda = 5$ per year. Moreover, we assume that $M_1 = \cdots = M_k = M = 20$, $t_{ij} - t_{i,j-1} = \Delta = \frac{1}{2}$, $i = 1, \ldots, k$, $j = 1, \ldots, 20$, $a = 45$. For a given k we generated 2000 vectors $(U_1(t), \ldots, U_k(t))$ and $(N_1(t), \ldots, N_k(t))$ and using (2.3) and (2.10) we estimated λ and $\bar{F}$ (7) respectively. Simulated biases and mean squared errors (MSE) for

Table 1. Estimated bias and MSE

k	3	5	10	20	50
$\hat{Bias}(\hat{\lambda})$	$(21) \times 10^{-3}$	$(13) \times 10^{-3}$	2×10^{-3}	$(6.1) \times 10^{-4}$	$(17.3) \times 10^{-5}$
$\hat{Bias}(\hat{\bar{F}}_1(7))$	3×10^{-3}	$(1.8) \times 10^{-3}$	$(7.6) \times 10^{-4}$	$(1.3) \times 10^{-4}$	$(1.2) \times 10^{-5}$
$\hat{MSE}(\hat{\lambda})$	$(1.8) \times 10^{-4}$	$(1.1) \times 10^{-4}$	$(7.8) \times 10^{-5}$	$(1.3) \times 10^{-5}$	$(6.1) \times 10^{-6}$
$\hat{MSE}(\hat{\bar{F}}_1(7))$	$(1.1) \times 10^{-4}$	$(7.2) \times 10^{-5}$	$(3.1) \times 10^{-5}$	$(1.2) \times 10^{-5}$	$(.6) \times 10^{-5}$

different values of k are given in table 1. It is apparent from looking at these numbers that both bias and MSE decreases as k increases. Moreover, our estimator performs well even for small k. We should mention also that $\hat{\lambda}$ is always an unbiased estimator of λ.

3 Reliability Estimation: Scheme B

Recall that our goal is to estimate $\bar{F}_1(t)$ as well as $\Psi(a, t_0)$. We start with $\bar{F}_1(t)$. From (2.5), it is clear that

$$E\left[\frac{Z(t)}{t}\right] = \lambda\mu \ , \tag{3.1}$$

and

$$\text{Var}\left[\frac{Z(t)}{t}\right] = (\sigma^2 + \mu^2)\lambda \ . \tag{3.2}$$

Given that for the i-th system we observe only $U_1(t_{i1}), \ldots, U_1(t_{iM_i})$, $i = 1, 2, \ldots, k$, it is clear that if $U_1(t_{iM_i}) = a$, then $Z(t_{i_{M_i}}) > a$ and if $U_1(t_{iM_i}) < a$, then $U_1(t_{ij}) = Z(t_{ij}) = z_{ij}$, $j = 1, \ldots, M_i$. For the latter case we use

$$\hat{Z}_i(t_0) = \sum_{j=1}^{M_i} a_j Z_i(t_{ij}) \ , \tag{3.3}$$

where a_j is given by Equation (2.7). Similar to Section 2, the covariance function $K(.,.)$ is not specified and must be estimated from the data. Using the equation

(3.2) one can estimate $b = (\sigma^2 + \mu^2)\lambda$ by

$$\hat{b} = \frac{1}{N_2 - 1} \sum_{i=1}^{k} \sum_{j=2}^{M_i} \left[\left[\frac{U_1(t_{ij}) - U_1(t_{i,j-1})}{(t_{ij} - t_{i,j-1})} \right] - \frac{1}{N_2} \sum_{j=2}^{M_i} \frac{U_1(t_{ij}) - U_1(t_{i,j-1})}{t_{ij} - t_{i,j-1}} \right]^2 \tag{3.4}$$

and $K(t, s)$ by

$$\hat{K}(t, s) = \hat{b} \operatorname{Min}(t, s) \ , \tag{3.5}$$

where $N_2 = \sum_{i=1}^{k} M_i$.

Now, using (3.3)

$$\hat{\bar{F}}_{1B}(t) = \frac{1}{k} \sum_{i=1}^{k} \{ I(\hat{Z}_i(t_0) \geq a) I(Z_i(t_{i_{M_i}}) < a) + I(Z(t_{i_{M_i}}) \geq a) I(S_{1i} > t_0) \} \ . \tag{3.6}$$

Assuming that $M_1 = \cdots M_k = M$ and $t_{i,j} - t_{i,j-1} = \Delta$, $i = 1, \ldots, k$, $j = 1, \ldots, M$, we have the following theorem.

Theorem 2: If $k \to \infty$, then

a) $\hat{\bar{F}}_{1B}(t_0)$ is a consistent estimator of $\bar{F}_1(t_0)$

b) $\sqrt{k}(\hat{\bar{F}}_{1B}(t_0) - \bar{F}_1(t_0))$ is asymptotically normal with mean zero and variance $(\bar{F}_1(t_0))(1 - \bar{F}_1(t_0))$.

Proof: The proof of theorem is similar to the proof of Theorem 1 and it is omitted.

Using Theorem 2, $(1 - \alpha)\%$ confidence interval for $\bar{F}_{1B}(t_0)$ is

$$\hat{\bar{F}}_{1B}(t_0) \pm z_{1-(\alpha/2)} \sqrt{\frac{\hat{\bar{F}}_{1B}(t_0)(1 - \hat{\bar{F}}_{1B}(t_0))}{k}} \ . \tag{3.7}$$

One can apply similar argument used in the previous section to estimate $\bar{F}_2(t_0)$.

To estimate $\Psi(a, t_0)$, one can follow the same argument used in the previous section with replacing $\hat{Z}_i(t_0)$ in Equation (2.13) by $\hat{Z}_i(t_0)$ in Equation (3.3).

Acknowledgement: This research was partially supported by the Air Force Office of Scientific Research Grant Number 89-0402.

References

1. Barlow RE, Proschan F (1981) Statistical theory of Reliability and life testing. To begin with, *Silver Spring, MD*
2. Esary JD, Marshal AW, Proschan F (1973) Shock models and wear processes. *Ann. Probability*, 1:627–649
3. Goldberger A (1962) Best linear unbiased prediction in the generalized linear regression model. *J. Amer. Stat. Assoc.* 57:369–375
4. Grandell J (1991) Aspects of risk theory. Springer Verlag, New York
5. Journel AG, Huijbregts CJ (1978) *Mining Geostatistics.* Academic Press, London
6. Kitanidis PK (1983) Statistical estimation of polynomial generalized covariance function. *Hydrological Applications: Water Resour. Res.* 19:909–921
7. Rendu BD (1978) *An Introduction to Geostatistical Methods of Mineral Evaluation.* S. Afr. Inst. Mining Metall
8. Seal H (1978) *Survival Probabilities, The goal of risk theory.* John Wiley and Sons Inc., New York

Received 30.06.1992
Revised version 26.11.1992

Metrika (1993) 40:349–359

Optimal Designs for Models with Block-Block Resp. Treatment-Treatment Correlations

KONRAD ENGEL and SYLKE GIERER

Universität Rostock, Fachbereich Mathematik, 18051 Rostock, Germany

Summary: In this paper we study optimal designs assuming two special covariance structures of the observations, namely that the covariance between the observations depends only on the blocks resp. the treatments. We show that the weighted least squares estimator equals the ordinary least squares estimator. Then we prove that block-block correlations resp. treatment-treatment correlations do not have any influence on the A- and MV-optimality resp. A-optimality. For the study of the MV-optimality in the case of treatment-treatment correlations we use the idea of invariance to find optimal C-matrices.

Key Words: Optimal designs, correlated observations, C-matrix, invariant problems, linear inference.

1 Introduction

In the theory of optimal designs it is very often assumed that the observations are uncorrelated and have all the same variance (briefly we call this the *uncorrelated case*). With this article we will emphasize that also for models in which the covariances between the observations depend only on the blocks resp. the treatments there exists a satisfactory theory at least for the construction of optimal C-matrices and in several cases also for the construction of optimal designs.

Fundamental work for designs with correlated observations was done by Kiefer and Wynn (1981), where one can find in addition some historical background. As it is also mentioned in the book of Shah and Sinha (1989), Kiefer and Wynn proposed a two stage procedure: At stage one we look for a class of designs which are optimal in the uncorrelated case. At the second stage we consider the ordinary least squares (OLS) estimates and look for designs within the class obtained at stage one which maximize the precision of these OLS estimates with respect to the true covariance structure. In particular, Kiefer and Wynn studied the "nearest neighbour" model assuming observations in different blocks to be uncorrelated whereas observations in the same block are assumed to be correlated only if they are neighbours.

Of course, if the covariance structure of the observations is precisely known, one has to use weighted least squares (WLS) estimates to find the really optimal

0026–1335/93/5/349–359 $2.50

designs. But we shall see that in our models already the OLS estimates have the maximum precision. Considering A- and MV-optimality, we shall prove that the case where the covariances between the observations depend only on the blocks can be reduced to the uncorrelated case. Further we shall see that the same holds for treatment-treatment correlations and A-optimality. The investigation of MV-optimality in the case where the covariances between the observations depend only on the treatments enables the application of the concept of invariance (see Sinha (1982)) and the method of convex functionals (Kiefer (1975)) to find optimal designs.

Let us mention that related results for models with correlated observations were obtained by Budde (1984), Bischoff (1992), Krafft and Schaefer (1992) where in particular *D*-optimality was studied.

2 Two Models for Correlated Observations

We consider the situation where v treatments are to be compared using b blocks each of k different treatments, that is, $k \leq v$. A *block design* d is a $k \times b$ array of treatment symbols $1, \ldots, v$ where the columns represent the blocks.

The usual linear additive model specifies the expectation of the observation y_{ij} on treatment i in block j as

$$E(y_{ij}) = \alpha_i + \beta_j$$

where α_i is the (unknown) effect of treatment i and β_j is the (unknown) effect of the jth block, $i = 1, \ldots, v$, $j = 1, \ldots, b$. In matrix notation we have:

$$E(\boldsymbol{y}) = (X_1 | X_2) \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix}$$

where $\boldsymbol{y}$ is the bk-dimensional vector of the observations, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the vectors of the treatment resp. block effects, and X_1 and X_2 are the *design matrices* for the treatment and block effects, respectively.

Let

$$V = \operatorname{cov}(\boldsymbol{y})$$

be the covariance matrix of $\boldsymbol{y}$. We assume the following two special structures for V:

Model 1:

$$\operatorname{cov}(y_{ij}, y_{i'j'}) = \begin{cases} \sigma^2 & (1 + w_{jj'}) \quad \text{if } (i, j) = (i', j') \ , \\ \sigma^2 & w_{jj'} \quad \text{otherwise} \end{cases}$$

(the covariances between and the variances of the observations depend only on the blocks).

Model 2:

$$\operatorname{cov}(y_{ij}, y_{i'j'}) = \begin{cases} \sigma^2 & (1 + w_{ii'}) \quad \text{if } (i, j) = (i', j') \ , \\ \sigma^2 & w_{ii'} \quad \text{otherwise} \end{cases}$$

(the covariances between and the variances of the observations depend only on the treatments).

In both models the entries $w_{jj'}$ resp. $w_{ii'}$ are known or unknown real numbers (not depending on the design d) subject to the condition that V is symmetric and positive definite. To avoid too much parameters we will suppose throughout

$$\sigma^2 = 1 \ .$$

Let I be the identity matrix. The following observation is fundamental for this paper.

Proposition 1: a) *In Model 1 we have*

$$\operatorname{cov}(y) = X_2 W X_2' + I$$

where $W = (w_{jj'})$, $j, j' = 1, \ldots, b$.

b) *In Model 2 we have*

$$\operatorname{cov}(y) = X_1 W X_1' + I$$

where $W = (w_{ii'})$, $i, i' = 1, \ldots, v$. □

As mentioned in the introduction we will speak of the (usual) *uncorrelated case* if $W = 0$.

It is well-known that for an estimable linear combination of the treatment effects $l'\alpha = (l'|0')\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ (i.e. $(l'|0')$ is in the row span of $(X_1|X_2)$) the WLS estimator being best linear unbiased is given by

$$\widehat{l'\alpha} = (l'|0')(X'V^{-1}X)^{-}X'V^{-1}y \tag{1}$$

where A^{-} denotes any g-inverse of the matrix A, $X = (X_1|X_2)$, and $0' = (0, \ldots, 0)$. For the variance of the WLS estimator we have

$$\operatorname{var}(\widehat{l'\alpha}) = l'C_{V,d}^{-}l \tag{2}$$

where $C_{V,d} = X_1'V^{-1}X_1 - X_1'V^{-1}X_2(X_2'V^{-1}X_2)^{-1}(X_2'V^{-1}X_1)$.
(Note that $C_{V,d}$ depends on d through the design matrices X_1 and X_2.)

If $V = I$, then $C_{V,d}$ is the usual *C-matrix* of the design d, briefly denoted by C_d, and one can calculate C_d in the following way:

$$C_d = r^{\delta} - \frac{1}{k}N_dN_d'$$

where $r^{\delta} = \operatorname{diag}(r_1, \ldots, r_v)$, r_i is the number of replications of treatment i in d and $N_d = (n_{d_{ij}})$ is the incidence matrix of d, $n_{d_{ij}} = 1$ if treatment i is contained in the jth block and $n_{d_{ij}} = 0$ otherwise.

Throughout we will suppose that d is connected, i.e. $\operatorname{rg} C_d = v - 1$ which also implies $\operatorname{rg} C_{V,d} = v - 1$.

It is well-known that then $l'\alpha$ is estimable iff $l'\mathbf{1} = 0$ where $\mathbf{1} = (1, \ldots, 1)'$, i.e. iff $l'\alpha$ is a *contrast*.

One can easily see that in both models the columns of the matrix $V(X_1|X_2)$ are contained in the column space of the matrix $(X_1|X_2)$. By a result of Zyskind (1967) (cf. Kiefer and Wynn (1981, p. 739)) the OLS estimator equals the WLS estimator, thus we obtain (using the Moore-Penrose g-inverse A^{+} of the matrix A)

Proposition 2: In Model 1 and Model 2 the WLS estimator is given by

$$\widehat{l'\alpha} = (l'|0')(X'X)^{-}X'y = (l'|0')X^{+}y \ .$$

□

Theorem 1: a) *In Model 1 we have*

$$\operatorname{var}(\widehat{l'\alpha}) = l'C_d^{-}l \ .$$

b) *In Model 2 we have*

$$\operatorname{var}(\widehat{l'\alpha}) = l'(C_d^- + W)l \ .$$

Proof: Since $(l'|\mathbf{0}')$ is in the row span of $X = (X_1|X_2)$, there exists some x with

$$X_1'x = l \qquad \text{and} \qquad X_2'x = \mathbf{0} \ .$$

Because XX^+ is an orthoprojector onto the column space of X we obtain

$$\begin{aligned}
\operatorname{var}(\widehat{l'\alpha}) &= (l'|\mathbf{0}')(X^+VX^{+\prime})\begin{pmatrix} l \\ \mathbf{0} \end{pmatrix} \\
&= x'XX^+VX^{+\prime}X'x \\
&= x'X_iWX_i'x + l'C_d^-l \\
&= \begin{cases} l'C_d^-l & \text{if } i = 2 \ , \\ l'(C_d^- + W)l & \text{if } i = 1 \ . \end{cases}
\end{aligned}$$

□

3 Optimality Criteria

Let L be a given set of *contrast vectors*, i.e. of vectors l with $l'\mathbf{1} = 0$. The design d is called *A- resp. MV-optimal with respect to L (w.r.t. L)* if

$$\sum_{l \in L} \operatorname{var}(\widehat{l'\alpha}) \qquad \text{resp.} \qquad \max_{l \in L} \operatorname{var}(\widehat{l'\alpha}) \tag{3}$$

is minimum in the class of all connected designs with parameters v, b, and k. We consider here only such sets L for which the sum resp. the maximum exists.

Let e_i be the unit vector with 1 at coordinate i.
If $L = \{e_i - e_j\colon 1 \le i, j \le v, i \ne j\}$ the A- and MV-optimality w.r.t. L coincide with the usual A- and MV-optimality, respectively.
If $L = \{e_1 - e_j\colon 2 \le j \le v\}$ we have the case of comparing test treatments $2, \ldots, v$ with a standard treatment 1 (Bechhofer and Tamhane (1981), cf. Hedayat, Jacroux, and Majumdar (1988) for an overview).
If $L = \{l\colon l'\mathbf{1} = 0, \|l\| = 1\}$ the MV-optimality w.r.t. L is the usual E-optimality.

The matrix W is called *completely symmetric* (c.s.) if it is of the form $W = \xi I + \eta J$, where J is the matrix all of whose entries equal 1.

Theorem 2: Let L be a given set of contrast vectors.

a) *The design d is A- resp. MV-optimal w.r.t. L in Model 1 iff it is A- resp. MV-optimal w.r.t. L in the uncorrelated case.*
b) *The design d is A-optimal w.r.t. L in Model 2 iff it is A-optimal w.r.t. L in the uncorrelated case. If W is completely symmetric and all contrast vectors in L have some fixed norm δ, then d is MV-optimal w.r.t. L in Model 2 iff it is MV-optimal w.r.t. L in the uncorrelated case.*

Proof: a) The statement follows directly from Theorem 1 since W does not appear in $\operatorname{var}(\widehat{l'\alpha})$.

b) We have

$$\operatorname{var}(\widehat{l'\alpha}) = l'C_d^- l + l'Wl$$

and, for $W = \xi I + \eta J$ and $\|l\|^2 = \delta^2$,

$$\operatorname{var}(\widehat{l'\alpha}) = l'C_d^- l + l'(\xi I + \eta J)l = l'C_d^- l + \xi\delta^2 \ .$$

Hence the objective functions (3) differ from the correlated to the uncorrelated case only by the constant $\sum_{l \in L} l'Wl$ resp. $\xi\delta^2$ (which does not depend on d). □

4 MV-Optimal C-Matrices for Model 2

In the general case it seems to be hopeless to find MV-optimal designs for model 2. But using the idea of invariance, cf. Sinha (1982), we can restrict ourselves to the study of much smaller classes of C-matrices.

First of all we have to prove the convexity of the objective function. Let $\mathscr{C}$ be the class of all non-negative definite (n.n.d.) $v \times v$ matrices C with $\operatorname{rg}(C) = v - 1$ and let $\mathscr{C}_0 = \{C \in \mathscr{C} : C1 = 0\}$. Let $\Phi: \mathscr{C}_0 \to \mathbb{R}$ be defined by

$$\Phi(C) = \max_{l \in L} l'(C^- + W)l \ .$$

Lemma 1: The functional Φ is convex, i.e.

$$\Phi(\lambda C + (1 - \lambda)D) \leq \lambda\Phi(C) + (1 - \lambda)\Phi(D) \qquad \forall C, D \in \mathscr{C}_0 \quad \forall \lambda \in [0, 1] \ .$$

Proof: If we define $\Psi: \mathscr{C} \to \mathbb{R}$ by

$$\Psi(E) = \max_{l \in L} l'(E + W)l \ ,$$

then Ψ is evidently convex and *monotone*, i.e. $E \leq F$ (in the sense of $F - E$ being n.n.d.) implies $\Psi(E) \leq \Psi(F)$.

For $C \in \mathscr{C}_0$ we use the g-inverse

$$C^- = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & C_{11}^{-1} & \\ 0 & & \end{bmatrix}$$

where C_{11} is the submatrix of C deleting the first row and first column. By a result of Kiefer (1959), for $\lambda \in [0, 1]$

$$(\lambda C_{11} + (1 - \lambda)D_{11})^{-1} \leq \lambda C_{11}^{-1} + (1 - \lambda)D_{11}^{-1}$$

where D_{11} is defined analogously as C_{11}. This implies by the monotonicity and convexity of Ψ

$$\begin{aligned} &\Psi \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & (\lambda C_{11} + (1 - \lambda)D_{11})^{-1} & \\ 0 & & \end{bmatrix} \\ &\leq \Psi \left[\lambda \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & C_{11}^{-1} & \\ 0 & & \end{bmatrix} + (1 - \lambda) \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & D_{11}^{-1} & \\ 0 & & \end{bmatrix} \right] \\ &\leq \lambda \Psi \left[\begin{bmatrix} 0 & \cdots & 0 \\ \vdots & C_{11}^{-1} & \\ 0 & & \end{bmatrix} \right] + (1 - \lambda)\Psi \left[\begin{bmatrix} 0 & \cdots & 0 \\ \vdots & D_{11}^{-1} & \\ 0 & & \end{bmatrix} \right] , \end{aligned}$$

i.e.

$$\begin{aligned} \Phi(\lambda C + (1 - \lambda)D) &= \Psi((\lambda C + (1 - \lambda)D)^-) \\ &\leq \lambda \Psi(C^-) + (1 - \lambda)\Psi(D^-) = \lambda \Phi(C) + (1 - \lambda)\Phi(D) \ . \end{aligned}$$

□

Remark 1: Though the above proof is contained implicitely in the work of Kiefer we could not find it in the fundamental article on universal optimality by Kiefer (1975) resp. other papers nor in the book of Shah and Sinha (1989).

Let S_v be the symmetric group of permutations of $\{1, \ldots, v\}$. For $\pi \in S_v$ let P_π be the $v \times v$ matrix representation of π. Let G be a subgroup of S_v. A $v \times v$ matrix A is said to be *G-invariant* if

$$P_\pi' A P_\pi = A \qquad \forall \pi \in G \ .$$

(Note that $\{\pi \in S_v: P_\pi' A P_\pi = A\}$ is a subgroup of S_v.) Furthermore, the set L of contrast vectors is called *G-invariant* if

$$P_\pi \boldsymbol{l} \in L \qquad \forall \boldsymbol{l} \in L, \forall \pi \in G \ .$$

(Note that then by $\boldsymbol{l} \to P_\pi \boldsymbol{l}$ there is given a bijective mapping $L \to L$.) Of course, G-invariance implies G'-invariance if G' is a subgroup of G.

Finally, we mention that if d is a connected design corresponding to parameters v, b, and k, then C_d is a matrix with $C_d \in \mathscr{C}_0$ and $\operatorname{tr} C_d = (k-1)b$; note our assumption that the blocks contain k *different* treatments.

Theorem 3: Let the matrix W and the set L of contrast vectors be G-invariant. Then to any matrix $C \in \mathscr{C}_0$ there exists a G-invariant matrix $\tilde{C} \in \mathscr{C}_0$ with the same trace (the averaged version of C) such that

$$\Phi(\tilde{C}) \leq \Phi(C) \ .$$

Proof: First we show that Φ is *symmetric*, i.e.

$$\Phi(C) = \Phi(P_\pi' C P_\pi) \qquad \forall \pi \in G \ .$$

Note that $(P_\pi' C P_\pi)^+ = P_\pi' C^+ P_\pi$ and $P_\pi' W P_\pi = W$ for all $\pi \in G$. Now

$$\begin{aligned} \Phi(P_\pi' C P_\pi) &= \max_{\boldsymbol{l} \in L} \boldsymbol{l}'((P_\pi' C P_\pi)^+ + W)\boldsymbol{l} \\ &= \max_{\boldsymbol{l} \in L} ((P_\pi \boldsymbol{l})' C^+ (P_\pi \boldsymbol{l}) + (P_\pi \boldsymbol{l})' W (P_\pi \boldsymbol{l})) \\ &= \max_{\boldsymbol{l}_1 \in L} \boldsymbol{l}_1'(C^+ + W)\boldsymbol{l}_1 = \Phi(C) \ . \end{aligned}$$

Now let

$$\tilde{C} = \frac{1}{|G|} \sum_{\pi \in G} P_\pi' C P_\pi \ .$$

Then $\tilde{C}$ is obviously G-invariant and tr $\tilde{C}$ = tr C. By the convexity of Φ we have

$$\Phi(\tilde{C}) \leq \frac{1}{|G|} \sum_{\pi \in G} \Phi(P_\pi' C P_\pi) = \frac{1}{|G|} \sum_{\pi \in G} \Phi(C) = \Phi(C) \ .$$

□

Remark 2: We do not need the supposition that L contains $v - 1$ linear independent vectors as it is assumed in the paper of Sinha (1982).

Remark 3: In the case of G-invariance of W and the set of contrast vectors L, by Theorem 3, the investigation can be restricted to the study of G-invariant C-matrices which enables the determination of the MV-optimal C-matrix if G is "large", e.g. if $G = S_v(1)$, the group of permutations with fixpoint 1.

Of course, there need not exist designs yielding this MV-optimal C-matrix. But we propose to take designs whose C-matrix is "near" to the MV-optimal C-matrix. Since it cannot be assumed that W is precisely known this approach should be sufficient though an explicit calculus of errors remains as an open problem.

5 Generalizations

In order to introduce our Models 1 and 2 we restricted ourselves to binary designs, i.e. no treatment is contained more than once in some block. This restriction is not essential, otherwise one has to distinguish in Model 1 and 2 only the cases where the observations are the same or not (instead of $(i, j) = (i', j')$ or not). Using the covariance structures from Proposition 1 as starting point there is really no restriction to binary designs since we never needed that $(X_1 | X_2)$ contains only pairwise different rows. The only difference will be that in the case of non-binary designs the trace of the C-matrix is not fixed, given the parameters v, b, and k. In an application of Theorem 3 one has to find such G-invariant C-matrices which have *maximum trace.*

Another generalization concerns row-column designs. A *row-column design* is an arrangement of v treatments in a $k \times b$ array where the rows as well as the

columns represent the blocks of two corresponding block designs. Assume for the bk observations the linear additive model and for V

$$V = X_i W_i X_i' + I \ , \qquad i = 1, 2, 3 \ ,$$

where the X_i's are the design matrices for the treatment, row and column effects, respectively, and $W_i = (w_{ll'})_{l,l'=1,\ldots,m_i}$, $m_1 = v$, $m_2 = k$, $m_3 = b$. Using the same arguments as in the case of block designs we have:

$$\operatorname{var}(\widehat{l'\alpha}) = \begin{cases} l' C_d^- l & \text{if } i = 2, 3, \\ l'(C_d^- + W) l & \text{if } i = 1 \ . \end{cases}$$

where $C_d = r^\delta - \frac{1}{b} M_d M_d' - \frac{1}{k} N_d N_d' + \frac{1}{bk} rr'$ (the C-matrix of the row-column design in the uncorrelated case; for the derivation see Shrikhande (1951) or Cheng (1978) for the more general context of multiway heterogeneity setting) and M_d and N_d are the treatment-row and treatment-column incidence matrices, respectively. Therefore, A- and MV-optimal designs in the uncorrelated case remain optimal if row-row or column-column correlations are assumed. Concerning treatment-treatment correlations we have the same results as in the case of block designs. Finally, in a natural way results can be obtained for establishing the optimality of block or row-column designs in the more general model

$$V = I + \sum_{i=1}^{n} X_i W_i X_i' \ , \qquad n = 2, 3 \ .$$

Acknowledgement: We would like to thank the unknown referee for many valuable suggestions which essentially improved the presentation of the article.

References

Bechhofer RE, Tamhane AC (1981) Incomplete block designs for comparing treatments with a control: General theory. Technometrics 23:45–57

Bischoff W (1992) On exact D-optimal designs for regression models with correlated observations. Ann. Inst. Statist. Math. 44:229–238

Budde M (1984) Optimale Zweifachblockpläne bei seriell korrelierten Fehlern. Metrika 31:203–213

Cheng CS (1978) Optimal designs for the elimination of multi-way heterogeneity. Ann. Statist. 6:1262–1272

Hedayat AS, Jacroux M, Majumdar D (1988) Optimal designs for comparing test treatments with controls. Statistical Science 3:462–491

Kiefer J (1959) Optimum experimental designs. J. R. Statist. Soc. B 21:272–319

Kiefer J (1975) Construction and optimality of generalized Youden designs. In: A Survey of Statistical Designs and Linear Models (J.N. Srivastava, ed.). North-Holland, Amsterdam, pp 333–353

Kiefer J, Wynn HP (1981) Optimum balanced block and latin square designs for correlated observations. Ann. Statist. 9:737–757

Krafft O, Schaefer M (1992) D-optimal designs for a multivariate regression model. J. Multivariate Anal. 42: 130–140.

Shah KR, Sinha BK (1989) Theory of optimal designs. Lecture Notes in Statistics No. 54. Springer-Verlag, Berlin, Heidelberg

Shrikhande SS (1951) Designs for two-way elimination of heterogeneity. Ann. Math. Statist. 22:235–247

Sinha BK (1982) On complete classes of experiments for certain invariant problems of linear inference. J. Statist. Plann. Infer. 7:171–180

Zyskind G (1967) On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. Ann. Math. Statist. 38: 1092–1109

Received 23.06.1992
Revised version 11.01.1993

Book Review

D. Feldman, M. Fox: Probability. The Mathematics of Uncertainty. Marcel Dekker, New York, Basel, Hong Kong, 1991, XV/404 pp., $135.00

The authors consider their book as preparation for a follow-up course in mathematical statistics; the students should have taken a course in advanced calculus previously. Indeed this book gives all the definitions and theorems which are necessary for an introduction to mathematical statistics. Sometimes more complicated proofs are omitted, for example: for the characterization theorem (p. 153) of distribution functions F it is not shown that the measure generated by F is countably additive. The central limit theorem (p. 305) is only formulated for identically distributed random variables and, for the proof of this, the relationship between the limit of distribution functions and the limit of accompanying moment generating functions is only cited (p. 380). Special attention must be called to the wealth of very good examples and of useful and carefully constructed problems.

What are the specific qualities of this book in comparison with other introductions to probability theory? The authors start from events and their relationships and not from elementary events (points of a sample space). This approach – attractive at first glance – has often been used in the past decenniums. But the difficulties and fussinesses are important, see, for example, the laborious introduction of a random variable. Add to this that the students should learn to read the literature in the usual terminology. Perhaps understanding would be facilitated if the theorem of Stone (there can be associated with every algebra of events an algebra of sets isomorphic to it) were cited and explained.

Würzburg W. Uhlmann

Metrika (1993) 40:361–365

Metrika

An Identity for Expectations of Functions of Order Statistics

Udo Kamps

Institut für Statistik und Wirtschaftsmathematik, RWTH Aachen, Wüllnerstr. 3, 52062 Aachen, Germany

Lutz Mattner

Institut für Mathematische Stochastik, Universität Hamburg, Bundesstr. 55, 20146 Hamburg 13, Germany

Summary: We consider an identity for expectations of general functions of order statistics valid in a parametric class of probability distributions. Corresponding characterization results are indicated.

Explicit expressions for moments of order statistics are often very complicated and thus of minor use. Hence identities connecting them are frequently examined. For an excellent review of relations valid for specific distributions we refer to Balakrishnan et al. (1988). Some isolated results from the literature are subsumed and generalized in Kamps (1991) by considering the class of distribution functions F satisfying

$$(F^{-1})'(t) = \frac{1}{d} t^p (1-t)^{q-p-1} \qquad \text{on } (0, 1) \tag{1}$$

for some $d > 0$ and integers p and q.

Special choices of p and q lead, e.g., to exponential, Pareto, power function, and logistic distributions.

In this note we consider an identity for expectations of general functions of order statistics. It yields as particular cases the identities for moments mentioned above but also, e.g., identities for characteristic functions.

Let throughout this paper $X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}$, $n \in \mathbb{N}$, denote the order statistics corresponding to independent, identically distributed random variables $X_1, X_2, \ldots, X_n$ with distribution function F.

Theorem: Let F satisfy (1), and let k be an integer with $2 \le k \le n$ and $1 \le k + p \le n + q$. Then

$$E[g(X_{k,n}) - g(X_{k-1,n})] = c \cdot Eg'(X_{k+p,n+q}) \,, \tag{2}$$

0026–1335/93/5/361–365 $2.50

where

$$c = c(k, n, p, q, d) = \frac{1}{d} \frac{\binom{n}{k-1}}{(k+p)\binom{n+q}{k+p}},$$

holds for every absolutely continuous function g provided that the right hand side in (2) exists.

Examples:

i) The choice $g(x) = x^m$, $m \in \mathbb{N}$, yields the equalities for ordinary moments of orders m and $m-1$ obtained in Kamps (1991). Considering $m = 1$ we observe that $c = E[X_{k,n} - X_{k-1,n}]$.

ii) Putting more generally $g(x) = x^{\alpha/\beta+1}$ with $\alpha \in \mathbb{R}$, $\beta > 0$, $\alpha + \beta \neq 0$, and rewriting the result in terms of $Y = (\beta X)^{1/\beta}$ leads to equalities for moments of orders $\alpha + \beta$ and α valid in a transformed, and by this enlarged, class of distributions as shown in Kamps (1992). (The quantities g and Y to be well defined, the support of F has to be chosen appropriately.)
Putting similarly $g(x) = e^{\alpha x}$ and $Y = e^X$ yields identities involving moments of order α only.

iii) Choosing $g(x) = e^{itx}$, $t \in \mathbb{R}$, we obtain

$$\varphi_{k,n}(t) - \varphi_{k-1,n}(t) = \mathrm{cit}\varphi_{k+p,n+q}(t) \ , \tag{3}$$

where $\varphi_{j,n}$ denotes the characteristic function of $X_{j,n}$. Such identities for characteristic functions of order statistics may be useful since simple explicit expressions are not available in general. The logistic distribution is an exceptional case with

$$p = q = -1 \qquad \text{and} \qquad \varphi_{k,n}(t) = \frac{\Gamma(k+it)\Gamma(n-k+1-it)}{\Gamma(k)\Gamma(n-k+1)}$$

(see Balakrishnan, Cohen 1991, p. 38).

To prepare for the proof of the theorem, and for the characterization results indicated below, we state the following lemma, which is probably well known.

Lemma: Let φ be an increasing and continuous function on some open interval (a, b), and let g be an absolutely continuous function on the range of φ. Then,

for every Borel function h on (a, b), we have

$$\int_{(a,b)} h \, dg \circ \varphi = \int_{(a,b)} h \cdot g' \circ \varphi \, d\varphi \ ,$$

provided that one of the above Lebesgue-Stieltjes integrals exists.

Proof: It suffices to consider $h = \mathbf{1}_{(c,d]}$ for $a < c < d < b$.

Applying the change-of-variables theorem (e.g. Royden 1968, p. 264), we have

$$\int_c^d dg \circ \varphi = g(\varphi(d)) - g(\varphi(c)) = \int_{\varphi(c)}^{\varphi(d)} g'(x) \, dx = \int_c^d g' \circ \varphi \, d\varphi \ ,$$

and thus the assertion.

Proof of the Theorem: For later use of what follows let first F denote an arbitrary distribution function. We may write, using the quantile transformation, and the density of an order statistic from a uniform distribution,

$$c(k, n, p, q, d) E g'(X_{k+p,n+q}) = \frac{1}{d} \binom{n}{k-1} \int_0^1 g'(F^{-1}(t)) t^{k+p-1} (1-t)^{n+q-k-p} \, dt \ ,$$

subject to the existence of $E g'(X_{k+p,n+q})$.

Now assume that F is strictly increasing on $(F^{-1}(0+), F^{-1}(1-))$, which is the case iff F^{-1} is continuous. Using the joint density of two successive order statistics (e.g. David 1981, p. 10) and the lemma above we get

$$\begin{aligned} E[g(X_{k,n}) - g(X_{k-1,n})] &= \frac{n!}{(k-2)!(n-k)!} \int_0^1 \int_0^1 (g \circ F^{-1}(v) - g \circ F^{-1}(u)) \\ &\quad \times 1_{\{u \le v\}}(u, v) u^{k-2} (1-v)^{n-k} \, du \, dv \\ &= \frac{n!}{(k-2)!(n-k)!} \int_0^1 \int_0^1 \int_0^1 1_{\{u \le t\}}(u, t) 1_{\{t \le v\}}(t, v) \\ &\quad \times u^{k-2} (1-v)^{n-k} d(g \circ F^{-1}(t)) \, du \, dv \\ &= \binom{n}{k-1} \int_0^1 g'(F^{-1}(t)) t^{k-1} (1-t)^{n-k+1} \, dF^{-1}(t) \end{aligned}$$

(see Khan, Yaqub, Parvez 1983, Lin 1988 in the case of ordinary moments).

If (1) holds, F^{-1} is necessarily continuous, and thus we find

$$E[g(X_{k,n}) - g(X_{k-1,n})] - cEg'(X_{k+p,n+q})$$
$$= \binom{n}{k-1} \int_0^1 g'(F^{-1}(t)) t^{k-1}(1-t)^{n-k+1} \{dF^{-1}(t) - \frac{1}{d} t^p (1-t)^{q-p-1}\, dt\} = 0 \ . \tag{4}$$

Inspecting the above proof, it is obvious that, under suitable conditions, characterization results may be obtained based on (2). We close this note by indicating some of them.

Assume throughout that F is strictly increasing on $(F^{-1}(0+), F^{-1}(1-))$, and that p, q and d are fixed.

If (2) holds for some fixed function g, with $g'(x) \neq 0$ except for isolated points, and all k and n, then (1) easily follows from considering (4). In fact, not all k and n are needed, applying certain complete sequences of functions. For details and surveys we refer to Lin (1988), Kamps (1991), and Huang (1989), Lin (1989) with respect to $g(x) = x^m$.

If instead (2) holds for a fixed pair (k, n), and for a sufficiently rich class of functions g, then (1) follows again. E.g., we may take all functions $x \mapsto e^{itx}$, $t \in \mathbb{R}$, which amounts to assuming (3) for all $t \in \mathbb{R}$.

Finally, an inequality for moments of order statistics, and by this a characterization result is evidently obtained involving only one equation in contrast to infinite many used above. E.g., we assume that g' is positive and $\{\ldots\}$ in (4) is nonnegative.

If F is absolutely continuous with density f, the latter condition reads

$$\frac{1}{d} F^p(t)(1 - F(t))^{q-p-1} f(t) \leq 1 \ ,$$

which for $p = q = 0$ (corresponding to exponential distributions) is a condition on the hazard rate of F.

When considering other inequalities for moments, interesting characterizations result under mild restrictions. We refer to the detailed paper of Gajek, Gather (1991) on inequalities of Hölder type.

Acknowledgement: The authors would like to thank the referee for helpful remarks.

References

Balakrishnan N, Cohen AC (1991) Order Statistics and Inference. Academic Press, Boston

Balakrishnan N, Malik HJ, Ahmed SE (1988) Recurrence relations and identities for moments of order statistics, II: Specific continuous distributions. Commun Statist-Theory Meth 17:2657–2694

David HA (1981) Order Statistics. 2 nd ed Wiley, New York

Gajek L, Gather U (1991) Moment inequalities for order statistics with applications to characterizations of distributions. Metrika 38:357–367

Huang JS (1989) Moment problem of order statistics: A review. International Statistical Review 57:59–66

Kamps U (1991) A general recurrence relation for moments of order statistics in a class of probability distributions and characterizations. Metrika 38:215–225

Kamps U (1992) Recurrence relations for moments of order statistics and record values. In: Gritzmann P et al. (eds) Operations Research '91:338–340. Physica, Heidelberg

Khan AH, Yaqub M, Parvez S (1983) Recurrence relations between moments of order statistics. Naval Research Logistics Quarterly 30:419–441

Lin GD (1988) Characterizations of distributions via relationships between two moments of order statistics. Journal of Statistical Planning and Inference 19:73–80

Lin GD (1989) Characterizations of distributions via moments of order statistics: A survey and comparison of methods. In: Dodge Y (ed) Statistical Data Analysis and Inference: 297–307. North-Holland, Amsterdam

Royden HL (1968) Real Analysis. 2 nd ed Macmillan, New York

Received 24.08.1992
Revised version 26.01.1993

Book Review

L. Ljung/G. Pflug/W. Walk: Stochastic approximation and optimization of random systems, Birkhäuser-Verlag, 1992, 113 pp., sFr. 38.00/DM 44.00

Stochastic approximation is a fine family of algorithms for the optimization of random systems. Such optimization, like minimization of risk for some observations, is a basic problem of mathematical statistics. Unfortunately, stochastic approximation did not become the real part of mathematical statistics, since mathematical statistics is not interested in the complexity of the algorithm applied; in a fairly one-sided way it is interested only in the best possible inference which can be made for the given observations. So the research of stochastic approximation was mainly the interest of probability theorists and engineers. Engineers were (and still are) interested in it typically for fast signal processing applications, where a microcomputer with the program of best possible inference would be either expensive or too slow with respect to the actual data speed.

Stochastic approximation is already a well-developed theory, and there are excellent survey articles and monographs. It is, therefore, not obvious why anyone should write and publish an additional book like the one under consideration. This book, however, gives several models and points, proving that there are a lot of new and interesting directions in this area.

This book is not a homogeneous monograph. It consists of three individual parts written by three leading experts in the field. Part I (*Foundation of stochastic approximation* by Harro Walk) summarizes the basic properties. In the usual works, the properties of stochastic approximation can be found if the observations are driven by a sequence of independent and identically distributed random variables. Here the behaviour of the risk along the iteration is investigated if the observations are driven by a time series. Part II (*Application aspects of stochastic approximation* by Georg Pflug) is mainly concerned with the so-called constant gain stochastic approximation, which is less complex than conventional stochastic approximation, but it is not consistent. For small gain, it is a homogeneous Markov chain with unique limit distribution. Recently, it was shown that the average of this procedure is consistent and has an optimum rate of convergence. This part contains sections on limit distribution of the conventional stochastic approximation, too, and on some nice applications of the theory. Part III (*Application to adaptation* by Lennart Ljung) deals with several models and problems concerning the importance of stochastic approximation for adaptation. The main task of adaptation is to track the changes of a stochastic system. Maybe this is the most promising direction towards the fruitful extensions and applications of stochastic approximation, when much more general models of changes are taken into account.

Budapest L. Györfi

Metrika (1993) 40:367–379

Metrika

Testing a Linear Regression Model Against Nonparametric Alternatives

GERHARD WEIHRATHER

Institut für Mathematik, Universität Augsburg, 86135 Augsburg, Germany

Summary: As a test statistic for testing goodness-of-fit of a linear regression model, we propose a ratio of quadratic forms measuring the distance between parametric and nonparametric fits, relative to the estimated error variance. The test statistic is a modification of the statistic suggested by Härdle and Mammen (1988). The asymptotic distribution under the hypothesis is established. The finite sample behaviour of the test is investigated in a Monte Carlo study, and is illustrated for two applications.

Key Words: Asymptotic normality; Linear regression; Nonparametric estimation; Simulation; Test.

1 Introduction

In many fields of application linear regression models play an important role. Let us assume that n observations $Y_{i,n}$ $(1 \leq i \leq n; n \in \mathbb{N})$ follow a linear regression model with fixed design

$$Y_{i,n} = x(t_{i,n})'\theta + \varepsilon_{i,n} \qquad (1 \leq i \leq n) \ . \tag{1}$$

The second index n indicates the dependence from the number of observations, which is needed for asymptotic considerations in section 2. The observations are obtained at the known design points $a_1 < t_{1,n} < \cdots < t_{n,n} < a_2$, where $A = [a_1; a_2] \subset \mathbb{R}$ is a fixed compact interval. The regression function $x: A \to \mathbb{R}^p$ $(p \in \mathbb{N})$ is known by the experimenter, whereas the unknown parameter vector $\theta \in \mathbb{R}^p$ is estimated from the data. Let us further assume that for fixed n the error variables $(\varepsilon_{i,n})_{1 \leq i \leq n}$ are independent and identically distributed with mean 0, variance $\sigma^2 \in (0; \infty)$, and finite fourth moment. After estimating the unknown parameters θ and σ^2 the question arises whether the linear model fits the data well.

Härdle and Mammen (1988) propose a method for testing the goodness-of-fit of a parametric regression model. They use as test statistic a weighted squared L_2-distance between the parametric and a nonparametric fit. The nonparametric

0026–1335/93/5/367–379 $2.50

regression model is a generalisation of the linear model. The two models differ in that the parametric regression function $x(t)'\theta$ is replaced by an unknown 'smooth' function $g(t)$, whence the nonparametric regression model can be written as

$$Y_{i,n} = g(t_{i,n}) + \varepsilon_{i,n} \qquad (1 \leq i \leq n) \ , \tag{2}$$

for some $g \in C^2[A]$. Whereas the mean function in model (1) lies in a p-dimensional space, the function g of model (2) is an element of an infinitely dimensional function space. The Härdle and Mammen test depends on the unknown variance σ^2 which is estimated by a time-consuming bootstrap procedure, called 'wild bootstrap'.

In the present paper a new method for testing the goodness-of-fit of a linear regression model is proposed. The test statistic is a discretisized version of that of Härdle and Mammen, divided by an estimate of the error variance. The critical value no longer depends on any unknown parameters, and no bootstrapping is necessary.

In section 2 the test statistic is introduced, and the asymptotic distribution under the null hypothesis is derived. The finite sample properties of the test are investigated in a simulation study in section 3, with a special view towards the goodness of the asymptotic approximation of the null distribution, and the power of the test under special alternatives. In the sections 4 and 5 the test is illustrated with the Forbes data of Atkinson (1985), and a data set on milk yield.

Alternative methods for testing the goodness-of-fit of parametric regression models are discussed by, e.g., Kozek (1990), and Eubank and Spiegelman (1990). Those authors consider stochastic design points (Kozek), and use slightly different distance measures than ours.

2 The Test Statistic

In section 1 the parametric regression model (1) and its nonparametric counterpart (2) are introduced. Using the $n \times p$ matrix $X_n = (x(t_{i,n})')_{1 \leq i \leq n}$ and the $n \times 1$ vectors $Y_n = (Y_{1,n}, \ldots, Y_{n,n})'$ and $\varepsilon_n = (\varepsilon_{1,n}, \ldots, \varepsilon_{n,n})'$, the linear model (1) can also be written as

$$Y_n = X_n\theta + \varepsilon_n \ .$$

We assume that the matrix X_n has full column rank p, and that the components of the regression function x are in $C^2[A]$. Our goal is to test the null hypothesis

of a linear model, against the alternative of a more general nonparametric model. More precisely, we wish to test the null hypothesis

$$H_0: g \in \mathcal{M}_{\mathcal{X}} = \{f \in C^2[A]: \exists \theta \in \mathbb{R}^p \forall t \in A: f(t) = \theta' x(t)\} \tag{3}$$

against the alternative

$$H_1: g \notin \mathcal{M}_{\mathcal{X}} .$$

The test statistic is based on a distance measure between the fit of the linear model and the fit of the nonparametric model. The null hypothesis H_0 is rejected for large values of the test statistic.

Let $\hat{\theta}_n = (X_n' X_n)^{-1} X_n' Y_n$ denote the least squares estimator of the unknown parameter vector θ. Using the 'hat matrix' $H_n = X_n (X_n' X_n)^{-1} X_n'$, the least squares fits and residuals can be written as $\hat{Y}_n = X_n \hat{\theta}_n = H_n Y_n$ and $R_n = (I_n - H_n) Y_n$, where I_n denotes the $n \times n$ identity matrix. In the nonparametric model (2), we estimate the regression function g using the kernel estimator of Gasser and Müller (1979). With $s_{0,n} = a_1$, $s_{n,n} = a_2$, $s_{i,n} = \frac{1}{2}(t_{i,n} + t_{i+1,n})(i = 1, \ldots, n-1)$ we define

$$W_{i,n}(t) = \frac{1}{b_n} \int_{s_{i-1,n}}^{s_{i,n}} K\left(\frac{t-x}{b_n}\right) dx \qquad (i = 1, \ldots, n) ,$$

where the bandwidth $b_n \in \left(0; \frac{a_2 - a_1}{2}\right)$ serves as a smoothing parameter. The prescribed kernel $K: \mathbb{R} \to \mathbb{R}$ is assumed to have compact support $[-1; 1]$, to be continuously differentiable on $\mathbb{R}$, and to satisfy the moment condition $\int_{\mathbb{R}} K(x)\, dx = 1$. Then the kernel estimate for $g \in C^2[A]$ is

$$\hat{g}_n(t) = \sum_{i=1}^{n} W_{i,n}(t)\, Y_{i,n} .$$

Following the lead of Härdle and Mammen (1988), we use the test statistic

$$\begin{aligned} T_n^{HM} &= n b_n^{1/2} \int_A \left[\hat{g}_n(t) - \sum_{j=1}^{n} W_{j,n}(t)\, \hat{Y}_{j,n}\right]^2 dt \\ &= n b_n^{1/2} \int_A \left[\sum_{j=1}^{n} W_{j,n}(t) R_{j,n}\right]^2 dt \\ &= n b_n^{1/2} R_n' A_n^{HM} R_n , \end{aligned}$$

where the $n \times n$ matrix $A_n^{HM} = (a_{ij,n}^{HM})$ is defined by $a_{ij,n}^{HM} = \int_A W_{i,n}(t) W_{j,n}(t)\, dt$. Härdle and Mammen actually use the Nadaraya-Watson kernel estimator instead of the Gasser-Müller estimator, and they consider a random design. For large values of T_n^{HM} the null hypothesis H_0 is rejected.

With a polynomial kernel function K we can calculate the test statistic T_n^{HM} exactly, but for each element $a_{ij,n}^{HM}$ we have to evaluate a high-degree polynomial. Hence exact evaluation of T_n^{HM} is rather expensive. Instead the test statistic is computed by numerical integration. Just as well we may then consider a discretisized version of T_n^{HM} right from the beginning, using the Riemann sum approximation

$$T_n = nb_n^{1/2} \sum_{i=1}^{n} \left[\sum_{j=1}^{n} W_{j,n}(t_{i,n}) R_{j,n} \right]^2 (s_{i,n} - s_{i-1,n}) \ .$$

With the $n \times n$ diagonal matrix $\Delta_n = diag((s_{i,n} - s_{i-1,n}))$ and the $n \times n$ matrix $G_n = (W_{j,n}(t_{i,n}))$, we rewrite T_n as a weighted sum of squares of the smoothed least squares residuals,

$$T_n = nb_n^{1/2} (G_n R_n)' \Delta_n (G_n R_n) \ .$$

Under the null hypothesis (3), T_n becomes

$$T_n = nb_n^{1/2} \varepsilon_n' (I_n - H_n) G_n' \Delta_n G_n (I_n - H_n) \varepsilon_n \ .$$

The null distribution of T_n and of T_n^{HM} depends on the unknown error variance σ^2. Härdle and Mammen use their 'wild bootstrap' method to estimate the critical values. The major goal of this paper is to avoid this time-consuming bootstrapping procedure.

The null distribution of $\frac{1}{\sigma^2} T_n$ does not depend on σ^2. Hence the statistic T_n is divided by an estimator $\hat{\sigma}_n^2$ of the error variance. Ideally the variance estimator has two properties: The null distribution of $T_n / \hat{\sigma}_n^2$ is independent of the unknown parameters θ and σ^2, and $\hat{\sigma}_n^2$ is a consistent estimator of σ^2 in the nonparametric model (2).

Gasser, Sroka and Jennen-Steinmetz (1986) propose the error estimate

$$\hat{\sigma}_n^2 = \frac{1}{n-2} \sum_{i=2}^{n-1} (a_{i,n}^2 + b_{i,n}^2 + 1)^{-1} (a_{i,n} Y_{i-1,n} - Y_{i,n} + b_{i,n} Y_{i+1,n})^2 \ ,$$

with $a_{i,n} = \dfrac{t_{i+1,n} - t_{i,n}}{t_{i+1,n} - t_{i-1,n}}$ and $b_{i,n} = \dfrac{t_{i,n} - t_{i-1,n}}{t_{i+1,n} - t_{i-1,n}}$. With a symmetric $n \times n$ band-

matrix N_n we rewrite $\hat{\sigma}_n^2$ as a quadratic form, $\hat{\sigma}_n^2 = \frac{1}{n-2} Y_n' N_n Y_n$. However, the null distribution of $T_n/\hat{\sigma}_n^2$ depends on the unknown parameter vector θ. We can circumvent this dependence by a slight modification of $\hat{\sigma}_n^2$.

We consider the variance estimator

$$S_n = \frac{1}{n-2} Y_n'(I_n - H_n)N_n(I_n - H_n)Y_n$$

$$= \frac{1}{n-2} R_n' N_n R_n$$

$$= \frac{1}{n-2} \sum_{i=2}^{n-1} (a_{i,n}^2 + b_{i,n}^2 + 1)^{-1}[a_{i,n}R_{i-1,n} - R_{i,n} + b_{i,n}R_{i+1,n}]^2$$

and the test statistic

$$Q_n = \frac{T_n}{S_n} .$$

In the parametric model (1), S_n is a consistent estimator of σ^2. However, the problem remains open whether it is consistent also in the wider nonparametric model (2).

We now analyse the null distribution of the test statistic Q_n. In the case of normally distributed errors, the exact null distribution for sample size n is a ratio of stochastic dependent mixtures of χ^2-distributions, with unknown mixing coefficients. Therefore we turn to an asymptotic approximation.

Härdle and Mammen (1988) show under mild regularity assumptions that T_n^{HM} is asymptotically normal, where the asymptotic null distribution depends on σ^2. Under similar conditions we show that Q_n is asymptotically normal under H_0, and does not depend on θ and σ^2. To be more precise, the following result can be derived.

Theorem: Assume the following asymptotic orders, as $n \to \infty$:

(i) $\max_{1 \le i \le n+1} |t_{i,n} - t_{i-1,n}| = O\left(\frac{1}{n}\right)$, *where* $t_{0,n} = a_1$ *and* $t_{n+1,n} = a_2$.

(ii) $\max_{1 \le i \le n} \left| s_{i,n} - s_{i-1,n} - \frac{1}{n}(a_2 - a_1) \right| = O\left(\frac{1}{n^{1+\delta}}\right)$ *for some* $\delta > 0$.

(iii) $b_n \to 0$, $n^{\delta} b_n^{1/2} \to \infty$ *with* δ *of* (ii), *and* $n b_n^{3/2} \to \infty$.

(iv) $\sup_{t \in A} |J_n(t)| = O(nb_n)$,
where $J_n(t) = \{i \in \{1, \ldots, n\} : (s_{i-1,n}; s_{i,n}) \cap (t - b_n; t + b_n) \neq \varnothing\}$.

(v) $\max_{1 \leq i \leq n} h_{ii,n} = O\left(\frac{1}{n}\right)$.

Then Q_n *is asymptotically normal with mean*

$$\bar{\mu}_n = b_n^{-1/2}(a_2 - a_1) \int_{-1}^{1} K(x)^2 \, dx$$

and with variance

$$\bar{\sigma}^2 = 4(a_2 - a_1)^2 \int_0^2 \left[\int_{-1}^{1-u} K(x)K(x+u) \, dx \right]^2 du \ ,$$

that is, $(Q_n - \bar{\mu}_n)/\bar{\sigma}$ *converge in distribution to a standard normal distribution.*

Assumptions (i), (ii), (iv) are fulfilled for equidistant design points $t_{i,n}$. Assumption (v) is satisfied for a polynomial regression model with equidistant design points. When working with a symmetric kernel K, Härdle and Mammen derive the same asymptotic constants in the case of a random design model. The proof of the theorem is based on a central limit theorem for quadratic forms of random variables by de Jong (1987) and on using the Slutsky-lemma.

We can use the result for approximating the critical value $z_n(\alpha)$ of Q_n, for a given level of significance $\alpha \in (0; 1)$. For large n and appropriate bandwidth b_n we get the approximation

$$z_n(\alpha) \approx \bar{\mu}_n + \bar{\sigma}\Phi^{-1}(1 - \alpha) \ , \tag{4}$$

where $\Phi^{-1}(1 - \alpha)$ denotes the α-fractile of the standard normal distribution.

With the quartic kernel $K(x) = \frac{15}{16}(1 - 2x^2 + x^4)1_{[-1;1]}(x)$ the values of the integrals in the asymptotic constants $\bar{\mu}_n$ and $\bar{\sigma}^2$ are found to be $\int_{-1}^{1} K(x)^2 \, dx = \frac{5}{7} \approx 0.7143$ and $\int_0^2 [\int_{-1}^{1-u} K(x)K(x+u) \, dx]^2 \, du = \frac{5}{49}(-198 + \frac{5}{11}882 - \frac{5}{13}504 - \frac{5}{17}24 + \frac{10}{19}) \approx 0.2582$. Thus, for a design region $A = [0; 1]$, we obtain $\bar{\mu}_n = \frac{5}{7}b_n^{-1/2}$ and $\bar{\sigma}^2 \approx 1.0328$.

3 Simulation Results

In this section the usefulness of the asymptotic approximation of the null distribution is investigated in a simulation study. We investigate the goodness of the

asymptotic approximation (4) of the critical value $z_n(\alpha)$, and examine the power of the test of how it compares to the wild-bootstrap method.

We consider as a null hypothesis the straight-line regression model

$$g(t) = \theta_0 + \theta_1 t \qquad (\theta_0 = 0; \theta_1 = 1) ,$$

and the quadratic regression model

$$g(t) = \theta_0 + \theta_1 t + \theta_2 t^2 \qquad (\theta_0 = 1; \theta_1 = 0; \theta_2 = -1) .$$

We choose equidistant design points $t_{i,n} = (i - 0.5)/n$, and $N(0; \sigma^2)$-distributed errors ($\sigma = 0.1$). Random samples of size $n \in \{25; 50; 75; 100\}$, with bandwidths $b_n \in \{k/20: k = 1, \ldots, 10\}$, were drawn by the TURBO PASCAL 5.0 generator. This was done for 1000 independent samples in each of the experimental settings.

The empirical quantiles of the null distribution of Q_n are given in table 1. Table 2 show the percentages of the simulated values of Q_n which were below the asymptotic approximation (4). The asymptotic approximation works well for some small bandwidths, but is inadequate for larger bandwidths. The bandwidth-intervals of good approximation are small and tend to zero for an increasing number of observations, which is in line with the assumption '$b_n \to 0$' of the theorem. They also depend on the regression model. The approximation in the straight-line model is better than in the quadratic model.

Because of the dependence of the regression model there is no general rule which bandwidth-interval the approximation yields good results. Thus it is advisable to use empirical quantiles for performing the test. These critical values have to be generated for each *type* of a linear regression model, whereas the bootstrapping procedure must be carried out for each *instance* of testing problem.

In the remainder of this section we compare the power of the new test with that of the wild-bootstrap method. We consider nearly the same experimental setting as did Härdle and Mammen (1988). We want to test the null hypothesis of a quadratic regression model ($x(t) = (1; t; t^2)$),

$$H_0: g \in \mathcal{M}_{\mathcal{X}} = \{f \in C^2[A]: \exists \theta = (\theta_0, \theta_1, \theta_2)' \in \mathbb{R}^3 \forall t \in A: f(t) = \theta_0 + \theta_1 t + \theta_2 t^2\} .$$

We take as alternative regression functions the cubic polynomials

$$g_c(t) = 2t - t^2 + c(t - 1/4)(t - 1/2)(t - 3/4) \qquad c \in \{0; 0.5; 1; 2\} .$$

Note that $g_0 \in \mathcal{M}_{\mathcal{X}}$. We choose level of significance $\alpha = 0.05$. In the simulation we use equidistant design points on $A = [0; 1]$ and generate $N(0; \sigma^2)$-distributed

Table 1. Empirical 0.95-quantiles of the null distribution of Q_n. The asymptotic approximation (4) of $z_n(0.05)$ is given in the last column

(a) Straight-line regression model

Bandwidth b_n	Number of observations n				Asymptotic approximation
	25	50	75	100	
0.05	5.47	5.42	5.07	4.84	4.87
0.10	4.09	3.84	3.64	3.36	3.93
0.15	3.46	2.97	2.68	2.60	3.52
0.20	2.35	2.26	2.09	1.94	3.27
0.25	1.94	1.85	1.55	1.60	3.10
0.30	1.65	1.34	1.25	1.26	2.98
0.35	1.26	1.06	1.03	0.93	2.88
0.40	0.90	0.81	0.86	0.77	2.80
0.45	0.75	0.62	0.61	0.56	2.74
0.50	0.53	0.45	0.45	0.42	2.68

(b) Quadratic regression model

Bandwidth b_n	Number of observations n				Asymptotic approximation
	25	50	75	100	
0.05	4.61	4.67	4.81	4.61	4.87
0.10	3.51	3.30	3.14	3.02	3.93
0.15	2.83	2.29	2.17	2.12	3.52
0.20	1.85	1.64	1.47	1.67	3.27
0.25	1.22	1.14	1.14	1.07	3.10
0.30	0.83	0.78	0.76	0.73	2.98
0.35	0.60	0.57	0.53	0.52	2.88
0.40	0.39	0.37	0.34	0.35	2.80
0.45	0.25	0.24	0.24	0.24	2.74
0.50	0.16	0.14	0.15	0.14	2.68

errors ($\sigma = 0.1$). Härdle and Mammen work with the Nadaraya-Watson kernel estimator, and use the quartic kernel. They consider a random-design model with uniformly distributed design points, and for each sample they perform 100 times the bootstrap resampling procedure.

For $c \in \{0; 0.5; 1; 2\}$ and bandwidths $b_n \in \{0.1; 0.2; 0.25; 0.3\}$ we computed 1000 random samples of size $n = 100$. Table 3 shows the empirical power of the new test method based on Q_n and that of the Härdle-Mammen wild-bootstrap method. Both methods work quite well and produce similar results. The larger the factor c, i.e. the greater the distance between g_c and the null hypothesis, the higher is the number of rejections of the null hypothesis. For small bandwidths the method of Härdle and Mammen produces slightly better results, whereas the Q_n-method is preferable for large bandwidths. The Q_n-test offers the interesting feature that the empirical power is increasing in the bandwidth. The variance estimator S_n performed quite well for each value of c. Neither test method

Table 2. Percentages of the simulated values of Q_n lying below the asymptotic approximation (4) of z_n (0.05)

(a) Straight-line regression model

Bandwidth b_n	Number of observations n			
	25	50	75	100
0.05	92.4	91.8	93.6	95.1
0.10	93.6	95.3	96.1	97.8
0.15	95.1	97.8	98.2	99.1
0.20	97.9	98.6	99.2	99.2
0.25	98.5	99.8	99.7	99.7
0.30	98.8	99.7	99.4	99.9
0.35	99.7	100.0	100.0	100.0
0.40	99.9	100.0	100.0	100.0
0.45	99.9	100.0	100.0	100.0
0.50	100.0	100.0	100.0	100.0

(b) Quadratic regression model

Bandwidth b_n	Number of observations n			
	25	50	75	100
0.05	95.5	95.9	95.4	96.7
0.10	96.6	98.0	98.8	99.0
0.15	97.6	99.1	99.3	99.2
0.20	99.4	99.5	99.7	99.6
0.25	99.7	99.8	99.9	100.0
0.30	100.0	100.0	100.0	100.0
0.35	99.8	100.0	100.0	100.0
0.40	100.0	100.0	100.0	100.0
0.45	100.0	100.0	100.0	100.0
0.50	100.0	100.0	100.0	100.0

produces uniformly better results. In the final two sections we apply the Q_n-test to two data sets.

4 Forbes Data

The Forbes data set consists of $n = 17$ observations on the boiling point T of water at different pressures, recorded at different elevations in the Alps (Atkinson 1985; p. 4–7). The response $Y = 100 \log(\textit{pressure})$ is modelled with a straight-line regression

$$Y = \theta_0 + \theta_1 T + \varepsilon \ .$$

Table 3. Empirical power of the Q_n-test and of the wild-bootstrap method. The latter ones are the values obtained by Härdle and Mammen

Bandwidth b_n	$c = 0$		$c = 0.5$		$c = 1$		$c = 2$	
	Q_n	HM	Q^n	HM	Q_n	HM	Q_n	HM
0.10	4.4	10.5	7.0	15.7	19.4	32.5	74.2	78.4
0.20	4.6	5.4	9.0	12.0	26.6	25.2	84.7	79.5
0.25	4.6	5.3	11.0	9.9	34.1	26.3	89.5	76.5
0.30	4.6	3.9	11.2	7.8	40.9	22.5	91.7	71.4

The analysis of Atkinson reveals that there is a strong linear relationship between T and Y. But a residual plot shows small, mostly negative residuals, and one large positive residual. Further analysis demonstrates that there is a strange observation which has a great influence on the fitted regression model. If this outlying observation is omitted from the analysis, the straight-line regression model describes the data very well.

With the Q_n-method for testing the null hypothesis of the straight-line regression model we get the following results. To perform the test, we transform the predictor variable T on the interval [0; 1] and use the bandwidth $b_n = 0.20$. We simulated, under the same conditions as in section 3, an empirical critical value of 2.81, while the test statistic Q_n takes the value 3.53. Therefore the null hypothesis is rejected.

If we drop the outlying observation then the value of Q_n decreases to 2.42. For $n = 16$ and $b_n = 0.20$ we determine the empirical critical value 2.94. Thus there is no significant evidence for rejecting the null hypothesis. These results conform with the analyses of Atkinson (1985). The example also shows that outlying observations can greatly influence the result of the test.

5 Milk Yield Data

Statistical analyse of milk yields plays an important role in agricultural science. Milk yields are recorded for individual cows, at regular time intervals throughout lactation. A common model to describe lactation curves is *Wood's curve*

$$f(t) = \beta_0 t^{\beta_1} \exp(\beta_2 t) \ ,$$

where $f(t)$ denotes the milk yield at time t. One possibility for estimating the unknown parameters β_0, β_1 and β_2 is to transform Wood's equation into a linear regression model. After taking the logarithm on both sides and introducing an error variable ε_t we get the linear model

Table 4. 41 Average daily milk yields per week (in litre) for cow No. 352, between 31st week in 1989 and 26th week in 1990. Data by curtosy of Dr. H. V. Henderson, Ruakura Agricultural Centre, Hamilton, New Zealand

Week	1	2	3	4	5	6	7
Milk yield	20.89	23.21	25.01	23.44	25.66	23.76	24.45
Week	8	9	10	11	12	13	14
Milk yield	24.17	25.87	26.77	25.18	22.30	23.09	24.48
Week	15	16	17	18	19	20	21
Milk yield	23.39	22.57	20.30	20.38	21.03	19.41	20.48
Week	22	23	24	25	26	27	28
Milk yield	19.46	19.65	19.47	20.42	19.11	16.69	14.41
Week	29	30	31	32	33	34	35
Milk yield	12.97	11.75	12.91	9.32	7.43	6.77	7.69
Week	36	37	38	39	40	41	
Milk yield	7.79	7.11	6.89	6.35	4.59	5.05	

$$Y_t = \theta_0 + \theta_1 \log t + \theta_2 t + \varepsilon_t \ ,$$

where $\theta_0 = \log \beta_0$, $\theta_1 = \beta_1$ and $\theta_2 = \beta_2$. The variable Y_t denotes the observed milk yield at time t, measured on a log-scale. Elston, Glasbey und Neilson (1989) argue that Wood's curve is not sufficiently flexible to describe the data. Especially at the time of peak yield Wood's curve does not give a good fit to data.

The milk yield data set consists of the average daily milk yield for an individual cow, recorded at $n = 41$ weeks in the years 1989 and 1990. It is shown in table 4.

In order to test the null hypothesis of the Wood model we transformed the predictor variable 'time' on the interval [0; 1], and computed empirical critical values for different bandwidths. For $b_n = 0.05$ the asymptotic approximation (4) of $z_n(\alpha)$ turned out to be quite good. For this bandwidth we got an empirical critical value of 5.0, while the test statistic Q_n takes the value 52.3. Therefore the null hypothesis is clearly be rejected.

Figure 1 shows the observed log milk yields, together with the least squares fit of the Wood curve (dashed line), and the fitted curve using the Gasser-Müller kernel estimator with $b_n = 0.05$ (solid line). The two fitted curves differ considerable. The nonparametric estimate indicates an extended high level plateau around the time of peak yield ($t = 0.1$ to $t = 0.6$, that is, week $n = 3$ to $n = 26$),

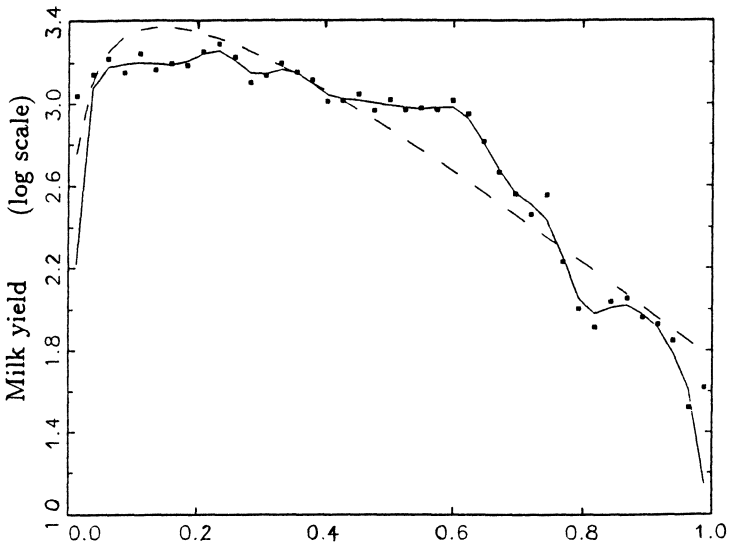

Fig. 1. Dots: 41 observed log milk yields of an individual cow.
Dashed line: Fitted Wood curve $\log f(t) = 4.54 + 0.398 \log t - 2,78t$.
Solid line: Fitted curve of the Gasser-Müller estimator with bandwidth $b_n = 0.05$.

followed by a short low level plateau towards the end of the lactation period ($t = 0.8$ to $t = 0.9$, that is, week $n = 33$ to $n = 38$). The parametric Wood model forces the data into a unimodal curve.

Although the Q_n-test performs well in the two examples discussed here, more practical experience is needed before general recommendations can be made.

Acknowledgement: The milk yield data set was kindly made available by Dr. H. V. Henderson from the Ruakura Agricultural Centre in Hamilton, New Zealand.

References

Atkinson AC (1985) Plots, Transformations, and Regression. Clarendon Press, Oxford

de Jong P (1987) A Central Limit Theorem for Generalized Quadratic Forms. *Probability Theory and Related Fields* 75:261–277

Elston DA, Glasbey CA, Neilson DR (1989) Non-Parametric Lactation Curves. *Animal Production* 48:331–339

Eubank RL, Spiegelman CH (1990) Testing the Goodness of Fit of a Linear Model Via Nonparametric Regression Techniques. *Journal of the American Statistical Association* 85:387–392

Gasser Th, Müller HG (1979) Kernel Estimation of Regression Functions.

Gasser Th, Rosenblatt M (Ed) Smoothing Techniques for Curve Estimation, 23–68, *Lecture Notes in Mathematics* 757, Springer-Verlag, Berlin

Gasser Th, Sroka L, Jennen-Steinmetz C (1986) Residual Variance and Residual Pattern in Nonlinear Regression. *Biometrika* 73:625–633

Härdle W, Mammen E (1988) Comparing Nonparametric versus Parametric Regression Fits. Preprint

(For a discussion of that paper see:

Härdle W (1990) Applied Nonparametric Regression. Cambridge University Press, Cambridge; Section 9.3, pp. 244–254

Kozek AS (1990) A Nonparametric Test of Fit of a Linear Model. *Communications in Statistics, Series A*, 19:169–179

Received 18.01.1993

Book Review

A. Benveniste/M. Metivier/P. Priouret: Adaptive algorithms and stochastic approximations, Springer, Berlin Heidelberg New York, 1990, XI/365 pp., DM 114,– This is the English translation of the book *Algorithms adaptifs et approximations stochastiques* published by Masson, Paris, 1987

Compared with its competitors, namely *Theorie and practice of recursive identification* by Ljung and Soderström (MIT Press, Cambridge, 1983) and *Adaptive filtering, prediction and control* by Goodwin and Sin (Englewood Cliffs, New Jersey), this is the most recent and the most mathematical book among them.

The book is divided into two parts. Part I (by Benveniste) is entitled *Applications*, but it also contains theory, theorems and proofs. This part is not easy to read, since the reader needs some background in engineering as well as in analysis and probability theory. The guiding example is that of adaptive equalization in data transmission. The recursive algorithm

$$\vartheta_n = \vartheta_{n-1} + \gamma_n H(\vartheta_{n-1}, X_n) + \gamma_n^2 \varepsilon_n(\vartheta_{n-1}, X_n) \tag{1}$$

where X_n stems from a Markov process controlled by ϑ_n is the general theme of the book. The method of comparison with ordinary differential equation (ODE method) is the main tool for analyzing it. A *guide to the analysis of adaptive algorithms* describes the different steps needed for analyzing the convergence properties of a given algorithm. Similarly, tracking algorithms are considered, and the practical steps described in another guide. These guides are nice summaries of the major stages in the applicaton. Change point detection is a further important topic, and it is also treated from an engineering point of view. The extensive statistical literature on change point detection is not considered. The first part closes with a very brief overview on systems theory in general such as Kalman filter. Many exercises are contained in the first part, but I doubt that any reader will work on them in detail, since they extend the material more than deepen it. For instance, the whole presentation of neural networks and the Gibbs sampler is done in exercises.

Part II, entitled "*Theory*" (by Métivier and Priouret) contains the main theoretical results about local and global convergence of algorithms of the form of Eq. 1. The style is rigorously mathematical and, to my taste, not very user-friendly, since numerous assumptions are distributed and cross-referenced over several pages, which makes the reading difficult. But it is, on the other hand, the first complete treatment of stochastic approximation algorithms with Markovian error structure.

To summarize, the book is certainly a major step in a high-level presentation of the mathematics of adaptive algorithms. It is not easy to read because it

contains engineering examples which are not understandable for mathematicians on the one hand and modern techniques of stochastic processes, like martingales, mixingales, diffusion processes, weak convergence etc. on the other hand. It is of high quality and I hope, but I am not quite sure, that this book will be greeted by interested reader.

Vienna G. Pflug

P. Hackl (Ed.): Statistical analysis and forecasting of economic structural change, Springer, Berlin Heidelberg New York, 1989, XIX/488 pp., DM 178,–

At the time it was completed (1989), this volume represented the latest in the statistical analysis of structural change. The editor, Peter Hackl, deserves much credit for organizing the volume and for being the inspiration behind the informal International Institute for Applied Systems Analysis (IIASA) Working Group to deal with problems on the statistical analysis of economic data for structural change. This volume represents contributions from conferences in Lodz, Poland (May 1985), Berlin (June 1986), and Sulejov, Poland (September 1986), and contains a variety of approaches to the study of structural change. All aspects of the field are well-represented, from the Bayesian to the conventional approaches, and from the theoretical to the more applied data-analytic studies.

Since Quandt's 1958 paper on switching regression, the statistical community has devoted much effort to this problem. Of course, the switching regression problem occurs in all scientific areas; however, the economists (see Poirer 1976) appear to have made an above-average effort to structural change problems, which is evident upon reading the review by Hackl and Westlund (1985).

The volume is divided into four sections and contains 25 chapters that were written by a total of 32 authors. In the introductory chapter, "*What Can Statistics Contribute to the Analysis of Economic Structural Change*", by Anderson and Mizon, the role of statistics in the detection of structural change and the role of respective methods in the evaluation of econometric models is discussed. Also, this article puts the present volume into a broader perspective by bridging the gap between the contributions by the statisticians and economists.

Identification of structural change is the topic of the second section of the volume, and is mostly concerned with statistical tests to detect a change. For example, Harvey and Phillips develop exact tests to detect changes in the parameters of simultaneous equations models. An interesting study by Kramer measures the robustness of the Chow test to autocorrelation in the observations. He shows that the Chow test is very nonrobust to changes in the autocorrelation, thus adjustments are necessary if autocorrelation is present. Hackl and Katzenbeisser compare several tests for detecting change and declare that some simple ones are almost as efficient as the well-known CUSUM procedure.

Tsurumi uses a Bayesian approach to detect structural change with an interval of a Bayesian predictive density and compares the Bayesian predictive approach with maximum likelihood and with Bayesian posterior density techniques. Wasilewski develops graphical techniques using residuals, which gives important information about the data structure and its effect on the estimation process. The plots of the smoothed residuals show the general tendency in the residual configuration, while with partial residual plots one can recognize the influential subsets of data. Such techniques are quite valuable in the investigation of structural change and represent a unique contribution to the subject.

"*Model Building in the Presence of Structural Change*" is the topic of the third section of the volume and gives the reader some new developments in the field. For example, Tong introduces nonlinear time series models to the analysis of structural change by using the threshold autoregressive class. In such a process, the present state of the process changes between two or more regimes, depending on the previous state of the process, thus the change is state- as well as time-dependent. Inferences are made via maximum likelihood estimation and illustrated with some artificial data. Nonlinear dynamics also play an important role in the article "*Statistical Identification of Nonlinear Dynamics in Macroeconomics Using Nonlinear Time Series Models*" by Ozaki and Ozaki. A mathematical model is introduced to explain the dynamics of the Hicksian IS-LM paradigm, in which the difference between Keynesian and monetarist interpretations of economic activity can be represented by the difference in the parameters of the model. Models based on linear and nonlinear time series processes are frequently used to model structural change, and in addition to the Tong and Ozaki articles, Broemeling works with ARMA processes that have constant mean, but changing autocorrelation and with models that have a constant covariance structure but a changing mean value function. Some of the articles have a computing flavor, as does the Kleffe paper "*Updating Parameters of Linear Change Point Models.*" He presents a more efficient way of computing the estimates of the regression coefficients and change point in the one-change regression model. The method shows considerably higher speed and better numerical stability than using the standard routines for linear regression.

There are many more articles than can be reviewed here, thus readers are encouraged to read this book and see for themselves a most recent account of structural change in economics. It should be mentioned that in 1991 the IIASA produced a follow-up to the volume reviewed here. It is entitled *Economic structural change, analysis and forecasting* and is published by Springer-Verlag with Hackl and Westlund as editors.

Galveston L. D. Broemeling

H. Kunita: Stochastic flows and stochastic differential equations. Cambridge studies in advanced mathematics 24, Cambridge University Press, 1990, XIV/346 pp., $69.50/£40.00

About 50 years ago, K. Ito founded the theory of stochastic differential equations (SDE's). Since then this theory, called stochastic analysis or stochastic calculus, has been developed extensively and it now provides us with an indispensable method in probability theory to analyse sample functions of stochastic processes.

A typical Ito SDE is given usually in the following form:

$$dX_t^i = \sum_{\alpha=1}^{r} \sigma_\alpha^i(t, X_t)dw^\alpha(t) + b^i(t, X_t)dt \ , \qquad i = 1, \ldots, d$$

$$X_0 = x \tag{1}$$

where $X_t = (X_t^1, \ldots, X_t^d)$ is a d-dimensional continuous process and $w(t) = (w^1(t), \ldots, w^r(t))$ is an r-dimensional Wiener process. This equation may be denoted in the following form:

$$dX_t = A(dt, X_t) \ , \qquad X_0 = x \ , \tag{2}$$

where $A(t, x) = (A^i(t, x))$ is defined by

$$A^i(t, x) = \sum_{\alpha=1}^{r} \int_0^t \sigma^i(s, x)dw^\alpha(s) + \int_0^t b^i(s, x)ds \ . \tag{3}$$

$A(t, x)$ is a Gaussian random field which may be called a Wiener vector field or a Wiener process taking values in vector fields. Rewriting Eq. 1 in the form of Eq. 2 provides us with a way of extending the notion of SDE's: We may take more general random fields $A(t, x)$ than given in the form of Eq. 3.

The uniqueness of the present book by Kunita is exactly in this point, namely, a characteristic feature which distinguishes it from existing standard text books on SDE theory is that it treats mainly SDE of the form of Eq. 2, where $A(t, x)$ is allowed to be a more general random field which may be called a semimartingale with values in vector fields. Such a generalization of SDE is not a mere generalization but has a substantial meaning in the study of stochastic flows as will be explained below.

Ito's motivation in introducing SDE was to construct diffusion processes associated to the Kolmogorov differential equations. If we denote the solution X_t of Eq. 1 by $X(t, x, w)$, then for each x, the map $w \to [t \to X(t, x, w)]$ defines a random variable with values in d-dimensional continuous paths so that this family of random paths defines a diffusion process on R^d. If we consider the map

$x \to X(t, x, w)$, however, it defines a homeomorphism on R^d (or diffeomorphism on R^d if the coefficients σ and b are smooth enough), so that we have a random one-parameter family (or a flow) of homeomorphisms, or diffeomorphisms when we regard the solution as giving the following map: $w \to [t \to [x \to X(t, x, w)]]$. This is called a stochastic flow of homeomorphisms or diffeomorphisms and it is to the study of this notion and its applications that this book is primarily devoted.

It is well-known that solutions of an ordinary differential equation (or a dynamical system) define a one-parameter group of diffeomorphisms, and hence it is quite natural and important to study the diffeomorphic properties of solutions for SDE. It should be noted that the study of SDE in connection with stochastic flows was already given in the Soviet Union in early works on the dependence of solutions on the initial values by, e.g., Gihman, Skorohod, Blagovescčnskii, and Freidlin. In recent years, it has been studied more extensively and, among others, a precise diffeomorphic property of solutions of Ito SDE has been finally established by Kunita and Bismut. In this book, this is further developed to solutions of general SDE in the form of Eq. 2 by a masterful method of the author.

The notion of stochastic flows can be defined independently of the notion of SDE. Roughly, it is a (right) invariant Brownian motion on the infinite dimensional Lie group formed of diffeomorphisms. An important and interesting class of stochastic flows has been introduced and studied by T. E. Harris. However, this class of stochastic flows cannot be produced as solutions of Ito's SDE, and it is here that the SDE in the form of Eq. 2 with a more general Wiener vector field plays a crucial role. Indeed, it was found by LeJan and Kunita that the most general class of stochastic flows can be constructed by this class of generalized SDE's. In the present book, the one-to-one correspondence is established between stochastic flow and the SDE of the form of Eq. 2 generating the flow. Furthermore, various important properties of this correspondence are studied in detail; among others, SDE's satisfied by the inverse flow and the composite of flows are determined precisely.

Many important applications of stochastic flows and the SDE's generating them are also discussed. Among other things, we can find applications to stochastic partial differential equations, Zakai and Kushner-Stratonovich equations in nonlinear filtering theory, in particular, and limit theorems in which the following three types of limit theorems are given in a unified way: (a) approximation theorems for SDE and stochastic flows studied by many people as refinement and generalization of the famous Wong-Zakai theorem, (b) limit theorems for driven processes due to Papanicolaou-Stroock-Varadhan, (c) limit theorems for stochastic ordinary differential equatons due to Khasminskii, Papanicolau-Kohler and others.

In conclusion, this is the first book – and thus unique – to treat the theory of SDE's and stochastic flows, to establish their relationship in full generality and in the best possible up-to-date and precise form, and to also give many important applications. It can be highly recommended as a text for graduate students as well as a research monograph for specialists.

Kyoto S. Watanabe

M. Schäl: Markoffsche Entscheidungsprozesse, Teubner-Verlag Stuttgart, 1990, XV/182 pp., DM 29,–

Many treatises on Markovian decision processes exist (also called stochastic dynamic programming, DP, or stochastic optimal control with discrete time parameter). They range from books for practising operations researchers to monographs on sophisticated foundational questions. The present textbook, written by a well-known researcher in the field, introduces the reader to both an up-to-date firm theoretical basis and the essentials of important applications. These include inventory models, the controlled queueing system M/G/1, replacement problems, and optimal selection. A completely novel feature is the unified treatment of stochastic DP, gambling, and optimal stopping, three theories which have for a long time developed quite independently. It is not surprising that this ambitious program can be carried out within less than 200 pages only by restricting oneself to the case of a countable state space (and by omitting some topics such as Bayesian models). This avoids most measure-theoretic and topological subtleties inherent in the case of a general Borel state space. Only rarely are some standard facts from measure theory needed. Thus, despite its high level, an introductory course on probability will suffice as a prerequisite for nearly all parts.

The book consists of two chapters. In the first one (about one-third of the book), basic, and some novel, facts from Markov chains with discrete time parameter are presented in a succinct and refreshing manner, skilfully interwoven with the modelling of the applications mentioned above and with a first excursion into DP, e.g. the performance of special strategies is studied. The bulk of the book is the second chapter where many theoretical results of stochastic dynamic programming are systematically developed and applied to the problems introduced previously. The emphasis is on the infinite-stage model with its several approaches.

A more detailed description of the contents is given in the following: The introductory paragraph whets the appetite with challenging gambling problems up to the optimality of the bold strategy in a subfair game. The examples already cited include the M/G/1 system (under three different assumptions) which is analysed by means of an embedded branching process. After deriving the basic facts about Markov chains, the main limit theorem is proven by a coupling method. For finite Markov chains, the continuity of the limit matrix as function of the transition matrix is shown, which is useful for the treatment of the average reward criterion in Chap. II. Chapter I closes with a treatment of optimal stopping problems.

Chapter II starts with the introduction of the finite-stage model with arbitrary action space and (deterministic) non-Markovian strategies. Throughout the

existence of an (upper) bounding function is assumed. This ensures the existence of the n-stage value function $i \to v^n(i) :=$ maximal expected n-stage reward for initial state i. The first basic result is the validity of the optimality equation (OG) and the existence of ε-optimal Markovian strategies. Under compactness and (semi-)continuity assumptions the existence of an optimal strategy is established. As an interlude, the classical result about the existence of an optimal (s, S)-ordering strategy in an inventory problem with K-convex value functions is derived in a succinct manner.

As is well-known, in infinite-stage models there are several possibilities for defining the value function v, e.g. as in DP by $v_1 := \sup_\delta \overline{\lim}_n E_\delta u_n$, where u_n is the random reward of the n-stage model, or as in gambling problems by $v_2 := \sup_\delta E_\delta \overline{\lim}_n E_\delta u_n$. Some of the main problems treated are to find conditions such that (i) v_1 and /or v_2 satisfy the OG, (ii) v_1 and/or v_2 are the (pointwise) limit of v^n for $n \to \infty$, (iii) the infinite-stage problem has a stationary (ε-) optimal strategy. The now classical cases (the negative, the discounted and the positive case) are treated in detail under weakened conditions, which are useful also for applications. Interesting results from the author's research form a large portion of the following parts of the book: (a) a section on positive models including so-called leavable models, where in each state there is an action which freezes the process for one period, (b) a section on undiscounted models with a new total reward criterion (as limit of β-discounted rewards for $\beta \uparrow 1$), (c) a section on so-called lim-sup models which contain most of the preceding models and (d) a section on generalized stopping problems where also one-stage rewards are allowed. With these last sections, the author completes his program of unifying the theories of DP, gambling and optimal stopping. It turns out that besides the notions of a "conserving" and "equalizing" strategy, according to Dubins, Savage and Sudderth, an additional uniform integrability condition is important for the characterisation of optimal policies. Two sections are devoted to the average reward criterion for the case of a finite state space and a compact action space.

I noticed a few misprints, but all of a harmless nature. In a new edition the inclusion of exercises would enhance the great value of the text.

Despite its high standard, the exposition is very elegant, clear and readable. It includes useful motivations, examples and counterexamples, giving much insight into the subtle interplay between assumptions and results and between the stochastic development and the sequential character of the optimization problem. The book can be highly recommended for course work as well as for seminars. Everybody who has mastered it should be very well prepared to study literature, either of theoretical or of applied character. I think it would be worthwhile to make the book accessible to a wider audience by a translation into English.

Karlsruhe K. Hinderer

J. Pfanzagl: Elementare Wahrscheinlichkeitsrechnung, 2nd edn., de Gruyter Lehrbuch, Walter de Gruyter, Berlin New York, 1991, XII/347 pp. DM 92,– (hardcover) DM 52,– (soft cover)

Despite the statements made at the beginning of its Preface, this text book (reviewed here in its second edition) was conceived in the spirit of measure and integration theory. The author pursues mathematical rigor and at the same time presents a variety of applications. No wonder that a second edition became necessary.

The difference between this edition and the first one is not large, but significant. The author has added Chapter 12 on the distribution of extremes, in which he discusses the asymptotic behavior of maxima and minima of random variables, identifies the corresponding limit distributions, and applies the results to band width and sample mean estimators. Since an introduction to basic problems of mathematical statistics is given only in the following chapter (13), their applications, although of a very special nature, can serve as an *hors d'oeuvre* to the more general but modest treatment of estimation and testing.

In the extended Preface, the author mentions that the idea of adding a further chapter on stochastic processes occurred to him as problematic, since a satisfactory exposition would have necessitated more sophisticated mathematical tools. Although every academic teacher faces this problem, some of them are successful in presenting at least part of the stochastic dynamics (e.g., Markov chains and random walks) on a mathematically acceptable level.

As in all later editions of publications, in this one, too, previous insufficiencies have been eliminated, naturally with the kind help of committed readers.

Tübingen H. Heyer