Statistics & Decisions 5, 89–104 (1987)

UNBIASED ESTIMATION IN NONPARAMETRIC CLASSES OF DISTRIBUTIONS

L. Rüschendorf

Received: Revised version: November 8, 1985

Abstract. We characterize the set of unbiased estimators of zero and MVUEs in nonparametric classes of distributions determined by generalized moment type conditions. We discuss especially estimation in the class of distributions with given marginals, in invariant models and for generalized translation families. In the case of a dominated family of distributions we determine the general structure of the zero estimators after n independent observations. The convexity of the underlying class of distributions turns out to be an important property.

1. Introduction

Let $(X,\mathcal{B})$ be a measurable space and let F be a set of real, measurable functions on X. Many interesting nonparametric families of distributions can be described as "large" subsets of

(1) $$M_r = M_r(F) = \{Q\in M^1(X,\mathcal{B});\ F\subset L^r(Q), \int f\,dQ = 0,\ \forall f\in F\},$$

AMS 1979 subject classification: 62 G 05

Key words and phrases: MVUE, nonparametric, LB-property, measures with

where $1 \leq r \leq \infty$, $M^1(X,\mathcal{B})$ denotes the set of all distributions on $(X,\mathcal{B})$. Examples for M_r are distributions with given moments, symmetric distributions or, more generally, distributions invariant w.r.t. a group of transformations, distributions with given marginals, distributions of stationary processes or of martingales.

For the case that F is finite, Hoeffding (1977a,b) has considered the problem of determining the symmetric unbiased estimators of zero, while Fisher (1982) in a subsequent paper considered certain nonlinear restrictions.

2. One Observation

At first we consider the case of one observation. For $1 \leq r \leq \infty$ let $D_{o,r}$ denote the set of all unbiased estimators of zero with existing r-th moment, i.e.

(2) $$D_{o,r} = D_{o,r}(M_r) = \{h \in L^r(M_r);\ \int h\, dP = 0,\ \forall\, P \in M_r\}\ .$$

Let $\frac{1}{r} + \frac{1}{s} = 1$ and define for $P \in M_r$

(3) $$F_r^{\perp}(P) = \{h \in L^s(P);\ \int fh\, dP = 0,\ \ \forall f \in F \cup \{1\}\}\ .$$

An important property of F turns out to be the following: F has *property* LB (lowerbounded) w.r.t. P if for $LB(P) = \{h : X \to R^1;\ \text{ess}_P\inf h > -\infty\}$ holds:

(4) $$F_r^{\perp}(P) \cap LB(P) \ \text{ is dense in } \ F_r^{\perp}(P) \text{ w.r.t. } \ L^s(P)\ .$$

This property prevents $F_r^{\perp}$ from being too small in some sense. For $r = 1$, (4) is fulfilled generally by definition for all $P \in M_1$. For $r > 1$ it is typically not satisfied if $F_r(P) = \langle f_1,\dots,f_k\rangle$, where f_i are not bounded from below.

For a set B of real functions let $\langle B\rangle$ denote the linear space generated by B; for $B \subset L^r(P)$ let $L^r(B,P)$ denote the closure of $\langle B\rangle$ in $L^r(P)$.

THEOREM 1.

a) $\bigcap_{P\in M_r} L^r(F,P) \subset D_{o,r}$

b) $D_{o,r} \subset \bigcap_{\substack{P\in M_r \\ F \text{ is LB w.r.t. } P}} L^r(F,P)$

c) $D_{o,1} = \bigcap_{P\in M_1} L^1(F,P)$

Proof.

a) For $h \in L^r(F,P)$ there exists a sequence $(f_n) \subset \langle F\rangle$ such that $f_n \to h$ in $L^r(P)$. Therefore, $0 = \int f_n dP \to \int h\, dP$, i.e. $\bigcap_{P\in M_r} L^r(F,P) \subset D_{o,r}$.

b) If $h \in D_{o,r}$ and if F is LB w.r.t. $P \in M_r$, then choose for $k \in F_r^{\perp}(P) \cap LB($ $c \in R^1$, $c > 0$, such that $c\,k \geq -1$ [P] and define $\tilde{P} = (1 + c\,k)P$. Then $\tilde{P} \in M_r$ and, therefore, $0 = \int h\, d\tilde{P} = \int h\, dP + c\int hk\, dP = c\int hk\, dP$, i.e. h is orthogonal to k. Since $F_r^{\perp}(P) \cap LB(P)$ is dense in $F_r^{\perp}(P)$, h is even orthogonal to $F_r^{\perp}(P)$. From $\int h\, dP = 0$ we infer that

(5) $$\begin{aligned} &h \in (F_r^{\perp}(P) \cup \{1\})^{\perp} \\ &= L^r(F \cup \{1\},P) \cap \{1\}^{\perp} = L^r(F,P), \end{aligned}$$

where (5) follows from the Hahn Banach theorem.

c) follows from a), b). □

While c) gives a complete description of $D_{o,r} = D_{o,1} \cap L^r(M_r)$, a) and b) are concerned with stronger forms of representation of $D_{o,r}$ by means of F. If F is a vectorlattice, then c) implies that $D_{o,r} = \bigcap_{P\in M_r} L^r(F,P)$.

From the covariance method we have for $g: M_r \to R^1$ and $D_{g,r}$ the set of unbiased estimators of g with existing r-th moment w.r.t. M_r:

COROLLARY 2. Let $P,Q \in M_r$, $d \in D_{g,r} \cap L^2(P)$ and let F be LB w.r.t. Q. If $d \in (L^r(F,Q) \cap L^2(P))^{\perp_P}$, then d is MVUE for g in P w.r.t. $D_{g,r}$ ($\perp_P$ denotes the orthogonal complement in $L^2(P)$).

More generally in Corollary 2 it is sufficient to find d in

$$\Big(\bigcap_{Q\in M_r} L^r(F,Q)\cap L^2(P)\Big)^{\perp P}$$

F LB w.r.t. Q.

3. Examples

a) Finite Case

If $F = \langle f_1,\dots,f_k\rangle$, $r\geq 1$, then F is LB for any $P\in M_r$. For the proof let $g\in F_r^{\perp}(P)$ and let g_n^i be bounded elements of $L^s(P)$ with $\langle g_n^i, f_i\rangle = 0$ and $g_n^i \xrightarrow[n\to\infty]{} g$, $1\leq i\leq k$, convergence w.r.t. $L^s(P)$. Define $h_n = \sum_{i=1}^k c_{i,n}\, g_n^i + c_{o,n}$ where the coefficients $c_{i,n}$ are choosen such that $\langle h_n, f_i\rangle = 0$, $\sum_{i=1}^k c_{i,n} = 1$; then $c_{o,n}\to 0$ and $h_n\to g$.

Therefore, $D_{o,r} = F$ for each r. If $d\in D_{g,r}$, then the MVUE for g in $P\in M_r$ w.r.t. $D_{g,r}$ is

$$(6)\qquad d^* = d + (c^*)^T f$$

where $f = (f_1,\dots,f_k)^T$ and c^* is a solution of

$$(7)\qquad E_P df + c^T G = 0;\quad G = G_P = \Big(\int f_i f_j \, dP\Big),$$

denoting the information matrix in P (cf. Hoeffding (1977a), Rüschendorf (1984a)).

b) Invariance

Let $F = \{f - f\circ g;\ g\in G,\ f\in B(X,\mathcal{B})\}$, where G is a group of measurable transformations of $(X,\mathcal{B})$ and $B(X,\mathcal{B})$ is the set of bounded real measurable functions on $(X,\mathcal{B})$. Then M_r is for $r\geq 1$ the set of all G-invariant distributions on $(X,\mathcal{B})$. We assume that $M = M_r \neq \emptyset$. From Theorem 1 we get that $D_{o,1} = \bigcap_{P\in M_1} L^1(F,P)$.

PROPOSITION 3. Let $\tilde g\colon M\to R^1$ be unbiased estimable.

a) If $d\in D_{\tilde g,\infty}$ and d is almost G-invariant w.r.t. M (i.e. $d\circ g = d$ [P], $\forall P\in M$, $g\in G$) then d is UMVUE for $\tilde g$ w.r.t. $D_{\tilde g,1}$.

b) If G is finite, then

b1) $D_{o,1} = \langle F\rangle = \{\sum_{g\in G}(f_g - f_g\circ g);\ f_g\in B(X,\mathcal{B})\}$.

b2) If $d\in D_{o,1}$ and d is G-invariant, then $d = 0$.

b3) If $d\in D_{g,1}$, then $d^* = \frac{1}{|G|}\sum_{g\in G} d\circ g$ is UMVUE for $\tilde g$ w.r.t. $D_{\tilde g,1}$.

b4) If each $g\in G$ is bijective, then the σ-algebra $\mathfrak{U}(G)$ of G-invariant sets is sufficient and complete.

Proof.

a) If $f - f\circ g\in F$, then $E_P\, d(f - f\circ g) = E_P\, d\, f - E_P d\, f\circ g = 0$ for all $P\in M$ since $d = d\circ g$ [P]. Therefore, $E_P\, dh = 0$ for all $h\in\langle F\rangle$. Since d is bounded, this relation extends to $L^1(F,P)$. Using that $D_{o,1}\subset L^1(F,P)$, the covariance method implies that d is UMVUE w.r.t. $D_{\tilde g,1}$.

b1) It is easy to see that
$L^1(F,P) = \langle\{f - f\circ g;\ f\in L^1(P),\ g\in G\}\rangle = \{\sum_{g\in G}(f_g - f_g\circ g);\ f_g\in L^1(P)\}$.
Since $\bigcap_{P\in M} L^1(P) = B(X,\mathcal{B})$, b1) follows.

b2) If $h = \sum_{g\in G}(f_g - f_g\circ g)\in D_{o,1}$ is G-invariant, then for $P\in M$

$$h = \frac{1}{|G|}\sum_{g'\in G} h\circ g' = \frac{1}{|G|}\sum_{g\in G}\sum_{g'\in G}(f_g\circ g' - f_g\circ g\circ g') = 0.$$

b3) Follows from b1).

b4) Sufficiency of $\mathfrak{U}(G)$ is wellknown while completeness follows from b2).□

From general ergodic theory is is known (cf. Theorem 5.1 of Tempel'man (1972)) that $L^1(P)$ has a unique decomposition in $L^1(F,P)\oplus\tilde L(G,P)$, where $\tilde L(G,P)$ denotes the set of all almost G-invariant integrable functions w.r.t. P. Therefore, we can interpret $D_{o,1}$ as the orthogonal complement of the bounded almost invariant functions w.r.t. M.

Under topological assumptions on G we can extend Propositon 3. Let G be a locally compact topological group acting measurably on $(X,\mathcal{B})$ and let $\tilde{\mathfrak{U}}(G)$ denote the σ-algebra generated by the M-almost G-invariant functions. Let μ be the Haar-measure on G. A strong form of an amenability condition is the Emerson-Tempel'man condition (cf. Tempel'man (1972), Th. 6.1, Bondar, Milnes (1981), pg. 120). There exists a summing sequence (G_n) of compact sets such that $\nu(G_n G_n^{-1})/\nu(G_n) \le B < \infty$ for all $n \in \mathbb{N}$, ν being the right invariant Haar-measure on G.

PROPOSITION 4. If G satisfies the Emerson-Tempel'man condition, then $\mathfrak{U}(G) = \tilde{\mathfrak{U}}(G)$ [M] and $\mathfrak{U}(G)$ is sufficient and L_2-complete.

Proof. Let $P \in M$ and $d \in B(X,\mathcal{B})$ and define $d_n = \frac{1}{\mu(G_n)} \int_{G_n} d \circ g \, d\mu(g)$, where (G_n) is a summing sequence. Then (d_n) is bounded and the individual ergodic theorem of Tempel'man (1972), Th. 6.1 (cf. also Bondar, Milnes (1981), pg. 120) implies that $\lim d_n$ exists P a.s. Define $d^* = \overline{\lim}\, d_n$, then d^* is G-invariant and $d^* = E_P(d|\mathfrak{U}(G))$, $\forall P \in M$ (cf. Theorem 6.1, part 3 of Tempel'man (1972)). This implies that $\mathfrak{U}(G)$ is sufficient for M.

Furthermore, by Proposition 3 a) d^* is a bounded UMVUE for $g(P) = E_P d$, $P \in M$. Now Theorem 3 of Padmanabhan (1974) implies that there exists a sufficient, L_2-complete σ-algebra $\mathfrak{U}$, namely, as follows from the proof of that theorem, the σ-algebra generated by the UMVUEs which are bounded, i.e. $\mathfrak{U}(G)$. Since for each $d \in \tilde{\mathfrak{U}}(G)$ we can construct d^* as in the first part of this proof with $d^* = E_P(d|\mathfrak{U}(G))$, both are UMVUE for $g(P) = E_P d$ and, therefore, $d = d^*$ [M], i.e. $\tilde{\mathfrak{U}}(G) = \mathfrak{U}(G)$ [M]. □

Remarks.

1. For $\mathfrak{P} \subset M$, $g \in G$ let $\overline{\mathfrak{U}(g)}$ denote the $\mathfrak{P}$-completion of the g-invariant sets (assuming that all g are one to one) and $\overline{\mathfrak{U}(G)} = \bigcap_{g \in G} \overline{\mathfrak{U}(g)}$.
   Several authors have contributed to the question whether $\overline{\mathfrak{U}(G)} \subset \tilde{\mathfrak{U}}(G)$ is sufficient for $\mathfrak{P}$. The most general results concern different kinds of amenable groups. In contrast to the sufficiency problem there are only few results for the question of completeness. Basu (1970) has shown that if a bounded complete and sufficient σ-algebra exists, then it is a subset of $\overline{\mathfrak{U}(G)}$. Proposition 3 a) and the result of Padmanabhan (1974) imply that sufficiency of $\tilde{\mathfrak{U}}(G)$ for M implies that $\tilde{\mathfrak{U}}(G)$ is L_2-complete w.r.t. M.

2. As consequence of Proposition 3 U-statistics remain UMVUE if one enlarges the model of iid random variables to the larger model of symmetric distributions. From 2 b) the σ-algebra of symmetric sets is symmetrically complete w.r.t. the class of all symmetric distributions.

c) Measures With Given Marginals

Let $(X,\mathcal{B}) = \bigoplus_{i=1}^{k} (X_i,\mathcal{B}_i)$, $P_i \in M^1(X_i,\mathcal{B}_i)$, $1 \le i \le k$,

(10) $$F = \langle \{f_i - \int f_i \, dP_i;\ f_i \in B(X_i,\mathcal{B}_i),\ 1 \le i \le k\} \rangle,$$

where f_i is considered as function on X, $f_i: X \to R^1$, defined as $f_i(x) = f_i(x_i)$. In this case for $r \ge 1$, $M_r = M(P_1,\ldots,P_k)$ is the class of all distributions with given marginals P_i on X_i. For general $P \in M(P_1,\ldots,P_k)$ a closed expression is unknown for $L^r(F,P)$ but the general idea seems to be right, that the spaces are getting bigger if P is tending to come "close" to the extreme points; for extreme points P it is known (at least for r = 1) that $L^1(F,P) = L^1(P)$. In view of Theorem 1 this leads us to concentrate on the opposite extreme, namely the product measure $\overline{P} = P_1 \oplus \ldots \oplus P_k$. The following Proposition answers a question put by Hoeffding (1977b), 2c.

PROPOSITION 5.

a) $L^r(F,\overline{P}) = \{\sum_{i=1}^{k} (f_i - \int f_i \, dP_i);\ f_i \in L^r(X_i,\mathcal{B}_i)\}$, $1 \le r \le \infty$.

b) $F_r^{\perp}(\overline{P}) = \{h \in L^s(\overline{P});\ \int h \, d \bigoplus_{j \ne i} P_j = c = \int h \, d\overline{P}\ [P_i],\ 1 \le i \le k\}$.

c) F is LB w.r.t. $\overline{P}$ for all $r \ge 1$.

d) $D_{o,r} = L^r(F,\overline{P})$ for all $r \ge 1$.

Proof.

a) Consider at first the case $r = 1$. Clearly,

$$\{\sum_{i=1}^{k} (f_i - \int f_i \, dP_i);\ f_i \in L^1(X_i,\mathcal{B}_i)\} \subset L^1(F,\overline{P}).$$

Let on the other hand $\sum_{i=1}^{k} f_{i,n} \to h$ in $L^1(\overline{P})$ for $n \to \infty$ (assuming w.l.g.

$\int f_{i,n} dP_i = 0$); then by Fubini and writing f_i instead of $f_{i,n}$

$\int |\sum_{\ell=1}^{k} f_\ell - h| d\overline{P} = \int [\int | \sum_{\ell=1}^{k} f_\ell - h | d \underset{j\neq i}{\oplus} P_j] dP_i \geq \int |f_i - \int h \, d \underset{j\neq i}{\oplus} P_j | dP_i$,

implying, that $f_i \to \int h \, d \underset{j\neq i}{\oplus} P_j$ in $L^1(P_i)$ and, therefore,

$\sum_{i=1}^{k} f_i \to \sum_{i=1}^{k} \int h \, d \underset{j\neq i}{\oplus} P_j$ w.r.t. $L^1(\overline{P})$, i.e. $h = \sum_{i=1}^{k} \int h \, d \underset{j\neq i}{\oplus} P_j \; [\overline{P}]$.

The case $r \geq 1$ follows from the case $r = 1$.

b) If $h \in L^s(\overline{P})$ and $\int f \, h \, d\overline{P} = 0$ for all $f \in F$, then

$\int f_i h \, d\overline{P} = \int f_i [\int h \, d \underset{j\neq i}{\oplus} P_j] \, dP_i = 0$ for all $f_i \in B(X_i, \mathcal{B}_i)$ with

$\int f_i dP_i = 0$, and, therefore, $\int f_i h_i dP_i = 0$, for all $f_i \in L^r(P_i)$,

$\int f_i dP_i = 0$, where $h_i = \int h \, d \underset{j\neq i}{\oplus} P_j$. Since $\int f_i h_i dP_i = \int f_i (h_i - \int h_i dP_i) dP_i$,

we obtain $\int f_i (h_i - \int h_i dP) \, dP_i = 0$, for all $f_i \in L^r(P_i)$, and, therefore,

$h_i = \int h_i dP_i = \int h \, d\overline{P} = c \; [P_i]$.

c) can be seen easily from b), while d) follows from Theorem 1. □

To consider a concrete example for Proposition 5 let $(X_i, \mathcal{B}_i) = (R^1, \mathcal{B}^1)$, $1 \leq i \leq k$, $g(P) = \int \prod_{i=1}^{k} x_i \, dP(x)$ and assume that $\int |x_i|^{rv2} dP_i < \infty$, $1 \leq i \leq k$. To determine the MVUE for g in $\overline{P}$ define

$$d^*(x) = \prod_{i=1}^{k} x_i - \sum_{i=1}^{k} x_i \left(\int \prod_{j\neq i} y_j \, d \underset{j\neq i}{\oplus} P_j\right) + k \prod_{i=1}^{k} \int x_i \, dP_i .$$

(11) d^* is MVUE for g in $\overline{P}$ (w.r.t. $D_{g,r}$).

<u>Proof.</u> By definition $d^* \in D_{g,r}$ and for $1 \leq i \leq k$ holds

$\int d^* \, d \underset{j\neq i}{\oplus} P_j = \prod_{j=1}^{k} \int x_j \, dP_j = c$, i.e. $d^* \in F_r^{\perp}(\overline{P})$ and, therefore, by Corollary 2 d^* is MVUE in $\overline{P}$. □

The additional terms to $\prod_{j=1}^{k} x_j$ reflect our knowledge of the marginals. d^* is no UMVUE. If $v_j \in L^r(P_j)$, $\int v_j dP = 0$ and $\prod^{k} v_j \geq -1$, then

$\tilde{P} = (1 + \prod_{j=1}^{k} v_j) \, \overline{P} \in M_r = M_r(P_1, \ldots, P_k)$. $\tilde{P}$ is a generalized FGM (Farlie-Gumbel-Morgenstern) distribution. The MVUE for g in $\tilde{P}$ (w.r.t. $D_{g,r}$) is given by

$$\begin{aligned} \tilde{d}(x) = {} & \prod_{j=1}^{k} x_j - \sum_{j=1}^{k} x_j \int \prod_{i\neq j} y_i \, d \underset{i\neq j}{\oplus} P_i \\ & - \sum_{j=1}^{k} x_j v_j(x_j) \int \prod_{i\neq j} y_i v_i(y_i) \, d \underset{i\neq j}{\oplus} P_i \\ & + k \left[\prod_{i=1}^{k} \int y_i \, dP_i + \prod_{i=1}^{k} \int y_i v_i(y_i) \, dP_i \right] . \end{aligned} \tag{12}$$

The correction term is the projection of $\prod_{i=1}^{k} x_i$ on $D_{0,r} = L^r(F, \overline{P})$ w.r.t. $\tilde{P}$. For the determination of the projection on $L^r(F,P)$ for general underlying $P \in M(P_1, \ldots, P_k)$ cf. Rüschendorf (1983).

d) <u>Measure Extensions, Generalized Translation Families</u>

Let $\mathcal{B}_0 \subset \mathcal{B}$ be a sub σ-algebra of $\mathcal{B}$, let $P \in M^1(X, \mathcal{B})$ and $P_0 = P|\mathcal{B}_0$. For $F = \{f \in B(X, \mathcal{B}_0); \int f \, dP_0 = 0\}$ $M = E(P_0) = \{Q \in M^1(X, \mathcal{B}); Q|\mathcal{B}_0 = P_0\}$ is the set of all measure extensions of $P_0|\mathcal{B}_0$ to the larger σ-algebra $\mathcal{B}$. $M \neq \emptyset$ by assumption, since $P \in E(P_0)$.

<u>Example.</u> Let G be a group of one to one transformations of $(X, \mathcal{B})$, $\mathcal{B}_0 = \mathfrak{U}(G)$ the σ-algebra of G-invariant sets and $P \in M^1(X, \mathcal{B})$ with $P_0 = P|\mathcal{B}_0$, then $E(P_0) = \{Q \in M^1(X, \mathcal{B}); Q^T = P^T\}$ where T is a maximal invariant w.r.t. G such that $\mathfrak{U}(T) = \mathcal{B}_0$. Especially, $E(P_0) \supset \mathcal{P} = \{P^g; g \in G\}$. If, especially, $X = R^n$ and G is a permutation group, then we get all distributions Q such that $S(Q) = \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} Q^\pi$ equals a given symmetric distribution P (assuming P to be nonatomic). If $G = R$ is the translation group on R^n, $gx = (x_1 + g, \ldots, x_n + g)$, $g \in G$, then $E(P_0)$ is a generalized translation-model; $E(P_0) \supset \mathcal{P} = \{P^g; g \in G\}$ the usual (parametric) translation-model.

<u>PROPOSITION</u> 6. For $P \in E(P_0)$, $\tilde{g}: E(P_0) \to R^1$, it holds:

a) $L^1(F,P) = \{f \in L^1(\mathcal{B}, P); \int f \, dP = 0, f = E_P(f|\mathcal{B}_0) \; [P]\}$

b) Let $d \in D_{\tilde{g},1} \cap L^2(P)$, then $d^* = d - E_P(d|\mathcal{B}_0) + E_P d \in D_{\tilde{g},1}$ and d^* is MVUE for $\tilde{g}$ in P w.r.t. $D_{\tilde{g},1}$.

c) $d \in L^2(E(P_0))$ is UMVUE $\Leftrightarrow E_P(d|\mathcal{B}_0) - E_P d$ is independent of $P \in E(P_0)$.

Proof.

a) If $f \in L^1(F,P)$, then there exists a sequence $(f_n) \subset F$ such that $f_n \to f$ in $L^1(P)$, implying the existence of a subsequence $(f_{n'}) \subset (f_n)$ which is a.s. convergent. Therefore, $f = \overline{\lim} f_{n'} = f^*[P]$, $f^* = E_P(f|\mathcal{B}_0) = f\,[P]$ and $\int f\, dP = \lim \int f_n dP = 0$. The converse inclusion is obvious.

b) For all $R \in E(P_0)$ it holds $E_R d^* = E_R d - E_R E_P(d|\mathcal{B}_0) + E_P d = E_R d = \tilde{g}(R)$, i.e. we have $d^* \in D_{\tilde{g},1}$. Moreover, for $f \in L^1(F,P) \cap L^2(P)$ we obtain $E_P d^* f = E_P df - E_P E_P(d|\mathcal{B}_0) f = 0$ which by Corollary 2 implies the assertion.

c) follows from b).

Remarks.

1. In part b) of Proposition 6 we construct an improvement of an estimator d. In the case of a generalized translation-model this improvement is equivalent to that discussed by Rao for the translation-model $\mathcal{P} = \{P^g;\ g \in G\} \subset E(P_0)$. If d is an equivariant estimator, i.e. $d(gx) = d(x) + g$ for $g \in G = R^1$ and $E_P d = 0$, then $E_{Pg}(d|\mathcal{B}_0) - E_{Pg} d = E_P(d+g|\mathcal{B}_0) - g = E_P(d|\mathcal{B}_0) - E_P d$. This implies that with $\tilde{g}(Q) = E_Q d$ the Pitman-estimator (in the representation of Rao) $d^* = d - E_P(d|\mathcal{B}_0)$ is MVUE for g for the whole translation class $\mathcal{P}$. Note that $\tilde{g}(P^g) = g$, for all $g \in G$. So in contrast to the situation of unbiased estimation in the location model $\mathcal{P}$ where Bondesson (1975) showed that UMVUE is very rare, we obtain in the generalized translation model that the Pitman estimator is always MVUE for the subclass $\mathcal{P}$.
The following simple argument communicated to the author by H. Luschgy explains this behaviour of d^*. If d_1, d_2 are unbiased for $\tilde{g}$ in $E(P_0)$, then $d_1 - d_2$ is by part a) of Proposition 6 almost surely G-invariant. Therefore, the optimality of d^* w.r.t. unbiased estimators is equivalent to the wellknown optimality of d^* w.r.t. all equivariant estimators.

2. It was shown to the author by D. Plachky that d^* is a UMVUE for $E(P_0)$ if and only if $d\, d_0$ is P a.s. $\mathcal{B}_0$-measurable for all $d_0 \in D_0 \cap L^2(P)$ and for all $P \in E(P_0)$. Therefore, if d is not $\mathcal{B}_0$-measurable, P a.s. and P_0 is not trivial (i.e. $P_0(A) \in (0,1)$ for some $A \in \mathcal{B}_0$) then d^* is not UMVUE.

4. Dominated Case

Consider now the following variant of Theorem 1 concerning dominated subsets of M_2. Let μ be a σ-finite measure on $(X,\mathcal{B})$, let $F \subset L^2(\mu)$ and let H be a subset of

$$(13) \qquad \{h \in L^2(\mu);\ \ h \geq 0,\ \int h\, d\mu = 1,\ \int h f\, d\mu = 0,\ \ \forall f \in F\}.$$

Then $\mathcal{P} = H\mu = \{h\mu;\ h \in H\} \subset M_2 = M_2(F)$ is a subset of M_2 dominated by μ. We denote by $\overline{B}$ the closure of a subset $B \subset L^2(\mu)$ in $L^2(\mu)$. Let $D_0(\mathcal{P}) = \{h \in L^2(\mu);\ \int h\, dP = 0,\ \forall P \in \mathcal{P}\}$; we have the following obvious relations:

LEMMA 7.

a) $D_0(\mathcal{P}) = D_0(\text{con}\ \mathcal{P}) = H^\perp$, ($\perp$ the orthogonal complement in $L^2(\mu)$).

b) $\mathcal{P}$ is complete w.r.t. $L^2(\mu)$ if and only if $\langle H \rangle$ is dense in $L^2(\mu)$. Let, especially, $H_0 \subset L^2(\mu)$ be a subset of all probability densities w.r.t. μ and $H = H_0 \cap F^\perp$, then we have for $g: \mathcal{P} \to \mathbb{R}^1$:

PROPOSITION 8.

a) $\mathcal{P} = H\mu \subset M_2(F)$, $\mathcal{P} \ll \mu$

b) $D_0(\mathcal{P}) = H^\perp = H_0^\perp + L^2(F,\mu)$

c) $D_0(\mathcal{P}) = L^2(F,\mu)$ if and only if $\overline{\langle H_0 \rangle} \supset F^\perp$

d) If $d \in D_{g,2}(\mu,\mathcal{P}) = \{h \in L^2(\mu);\ \int h\, dQ = g(Q)$ for all $Q \in \mathcal{P}\}$ and if $d \in \overline{\langle H \rangle \cap L^2(Q)}^{Q}$ (closure in $L^2(Q)$), then d is MVUE for g in $Q \in \mathcal{P}$. □

5. Several Observations

Consider the situation from section 4, i.e. $\mathcal{P} = H\mu$, $H \subset L^2(\mu)$ and assume that in order to estimate $g: \mathcal{P} \to R^1$ we make n independent observations, i.e. our model is $\mathfrak{P}^n = \{P^n;\ P \in \mathfrak{P}\}$. In contrast to the case of one observation we do not have the important relation

$$(14) \qquad D_o(\mathfrak{P}^n) = D_o((\mathrm{con}\ \mathfrak{P})^n),$$

but generally there is strict inequality. Caused by this fact completeness of $\mathfrak{P}$ w.r.t. $L^2(\mu)$ does not imply symmetric completeness of $\mathfrak{P}^n$ w.r.t. $L^2(\mu^n)$, the opposite direction being obvious.

Example. Let $(X,\mathfrak{B}) = (R^1,\mathfrak{B}^1)$, μ being a σ-finite measure on $(R^1,\mathfrak{B}^1)$, different from a one or two point measure. Let $\mathfrak{P}$ denote the set of all $P \in M^1(R^1,\mathfrak{B}^1)$ with $V(P) = 1$, $\int x^4 dP < \infty$ which have a square integrable density w.r.t. μ. With H being the set of all square integrable μ-densities h satisfying $\int x^2 h\, d\mu - (\int xh\, d\mu)^2 = 1$ and $\int x^4 h\, d\mu < \infty$ we have $\mathfrak{P} = H\mu$ and it holds:

1. $\mathfrak{P}$ is $L^2(\mu)$-complete
2. $u(x,y) = \frac{1}{2}(x-y)^2 - 1 \in D_o(\mathfrak{P}^2)$
3. $\mathfrak{P}^2$ is not symmetrically complete w.r.t. $L^2(\mu)$.
4. $(\mathrm{con}\ \mathfrak{P})^2$ is symmetrically complete w.r.t. $L^2(\mu)$.

Proof.

1. By Lemma 7 we have to prove that $\langle H\rangle$ is dense in $L^2(\mu)$ or, equivalently, that $H^\perp = \{0\}$. Define $H_a = \{h \in H;\ \int x h\, d\mu = a\}$; then $H_a^\perp = \{k \in L^2(\mu);\ \int k h\, d\mu = 0,\ \forall h \in L^2(\mu)$ with $\int (x-a) h\, d\mu = 0,\ \int (x^2 - (1+a^2)) h\, d\mu = 0\}$ $= (\{x^2 - (1+a^2),\ x-a\}^\perp)^\perp = \langle x^2 - (1+a^2),\ x-a\rangle$.
Since μ is not a two point measure, we have that $H_{a_1} \cap H_{a_2} = \{0\}$ for $a_1 \neq a_2$ and, therefore, $H^\perp = \left(\bigcup_{a\in R^1} H_a\right)^\perp = \bigcap_{a\in R^1} H_a^\perp = \{0\}$.

2., 3. are obvious while 4. follows from Theorem 8 of Rüschendorf (1984a).□

The following lemma is probably well known. We include a proof of it since we are not aware of a reference.

LEMMA 9. Let $H_i \subset \overline{H}$, $1 \le i \le k$, be subsets of a Hilbertspace $\overline{H}$ and let $\bigotimes^k H_i$ be the topological tensorproduct of H_i, $1 \le i \le k$, (i.e. the closure

of $\{\sum_{j=1}^k \alpha_j h_{j1} \otimes \ldots \otimes h_{jk};\ h_{ji} \in H_i\}$ in $\bigotimes_{i=1}^k \overline{H}$. Then $\left(\bigotimes_{i=1}^k H_i\right)^\perp =$

$$= \sum_{i=1}^k \overline{H} \otimes \ldots \otimes H_i^\perp \otimes \ldots \otimes \overline{H}\ .$$

Proof. We consider only the case $k = 2$; the general case following from induction.
Let $(f_i)_{i\in I}$ be an ON-basis of $\overline{H}$ such that for $I_1, I_2 \subset I$, $(f_i)_{i\in I_k}$ is an ON-basis of $H_k^\perp$, $k = 1,2$ and $(f_i)_{i \in I_k^c}$ is an ON-basis of $\overline{\langle H_k\rangle}$, $k = 1,2$; the existence following from the projection theorem. Then $(f_i \otimes f_j)_{(i,j)\in I\times I}$ is an ON-basis of $\overline{H}\otimes\overline{H}$ (cf. Neveu (1968), Lemma 6.5) and for $g \in (H_1 \otimes H_2)^\perp$ there exists a countable subset $D \subset I \times I$ such that

$$g = \sum_{(i,j)\in D} c_{ij}\, f_i \otimes f_j$$
$$= \Bigg(\sum_{\substack{(i,j)\in D\\ i\notin I_1, j\notin I_2}} + \sum_{\substack{(i,j)\in D\\ i\in I_1}} + \sum_{\substack{(i,j)\in D\\ i\notin I_1, j\in I_2}} \Bigg) c_{ij}\, f_i \otimes f_j\ .$$

Since $g \in (H_1 \otimes H_2)^\perp$ we get for $i \in I_1^c$, $j \in I_2^c$, $0 = \langle g, f_i \otimes f_j\rangle = c_{ij}$; i.e. $g = g_1 + g_2$ where $g_1 \in H_1^\perp \otimes \overline{H}$, $g_2 \in \overline{H} \otimes H_2^\perp$. □

As consequence we now obtain

THEOREM 10. Let $H_+ = \mathbb{R}_+ H = \{c h;\ c \in \mathbb{R}_+, h \in H\}$ be a convex cone, then

a) $D_o(\mathfrak{P}^n) = \sum_{i=1}^n L^2(\mu) \otimes \ldots \otimes H^\perp \otimes \ldots \otimes L^2(\mu)$,

b) $D_o(\mathfrak{P}^n) = \sum_{i=1}^n L^2(\mu) \otimes \ldots \otimes \langle F\rangle \otimes \ldots \otimes L^2(\mu)$,
$\Leftrightarrow \overline{\langle H\rangle} = F^\perp$.

Proof.
a) $D_o(\mathfrak{P}^n) = \{f \in L^2(\mu^n);\ \int f\, dQ^n = 0,\ \forall Q \in \mathfrak{P}\} = \{f \in L^2(\mu^n);\ \int f \bigotimes_{i=1}^n h\, d\mu^n = 0,$ $\forall h \in H\}$. By Corollary 5 of Rüschendorf (1984a) (compare also the proof of Theorem 8) and the assumption that H_+ is a convex cone we get

$$D_0(\mathfrak{P}^n) = \{f \in L^2(\mu^n);\ \int f \bigotimes_{i=1}^{n} h_i \, d\mu^n = 0 \text{ for all } h_i \in H,\ 1 \le i \le n\} =$$

$$= \left(\bigotimes_{i=1}^{n} H\right)^{\perp} = \sum_{i=1}^{n} L^2(\mu)\otimes \ldots \otimes H^{\perp} \otimes \ldots \otimes L^2(\mu) \text{ by Lemma 9.}$$

b) follows from a) since $H^{\perp} = \overline{<F>}$ is equivalent to $\overline{<H>} = (H^{\perp})^{\perp} = (\overline{<F>})^{\perp} = F^{\perp}$. □

Remarks and Examples.

a) Theorem 10 generalizes Theorem 1 B of Hoeffding (1977a) (note that Theorem 1 A can be deduced from 1 B) in two respects. Firstly we admit more general sets of moment functions F. Therefore, we can answer some questions of his paper (1977b). Secondly, we admit more general classes of hypotheses also in the case of finite F. This is caused by the fact that Hoeffding's proof is based on some involved finite approximation results.

b) Condition b) of Theorem 10 corresponds to the LB-condition of Theorem 1 (for $r = 2$). We see that in the dominated case the LB-condition is necessary for a result as in Theorem 1.

c) Let $\mathfrak{P} = M_{\overline{P}}(P_1,\ldots,P_k)$ be the set of all distributions in $M(P_1,\ldots,P_k)$ with L^2-density w.r.t. $\mu = \overline{P} = \bigotimes_{j=1}^{k} P_j$. From the proof of Proposition 5 we infer that F (cf. Example 1c)) is LB w.r.t. $\overline{P}$ and $\overline{<H>} = F^{\perp}$. Therefore, by Theorem 10 noting that H is convex we get

$$D_0(\mathfrak{P}^n) = \sum_{i=1}^{n} L^2(\overline{P})\otimes \ldots \otimes \overline{F} \otimes \ldots \otimes L^2(\overline{P}), \tag{15}$$

where $\overline{F} = \{\sum_{j=1}^{k} (f_j - \int f_j \, dP_j);\ L^2(P_j)\}$, i.e. typical elements of $D_0(\mathfrak{P}^n)$ are of the form

$$h(x_1,\ldots,x_n) = \sum_{i=1}^{n} \sum_{j=1}^{k} f_{ij}(x_{i,j}) h_i(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n),$$

$x_i = (x_{i,1},\ldots,x_{i,k})$, $h_i \in L^2(\overline{P}^{(n-1)})$, $f_{ij} \in L^2(P_j)$, $\int f_{ij} \, dP_j = 0$ for all i,j.

$$d^*(x_1,\ldots,x_n) = \frac{1}{n} \sum_{i=1}^{n} \{ \prod_{j=1}^{k} x_{i,j} - \sum_{j=1}^{k} x_{i,j} \int \prod_{r\neq j} y_r \, d(\bigotimes_{r\neq j} P_r) + k \prod_{j=1}^{k} \int y_j \, dP_j \} \tag{16}$$

is MVUE for $g(P^n)$ in $\overline{P}^n$.

Proof. If $\int f_{\ell s} dP_s = 0$, then

$E_{\overline{P}^n} d^* f_{\ell s}(x_{\ell,s}) h_{\ell}(x_1,\ldots,x_{\ell-1}, x_{\ell+1},\ldots,x_n) = 0$.

This relation follows if one considers at first indices $i \neq \ell$ and uses that $\int f_{\ell s} dP_s = 0$, while for $i = \ell$ we can argue as in relation (11). □

Similarly, if we want to estimate $g(P^n) = P(x_1 < x_2)$ in the case $k = 2$, the MVUE in $\overline{P}^n$ is

$$d^*((x_1,y_1),\ldots,(x_n,y_n)) = \frac{1}{n} \sum_{j=1}^{n} [1_{(x_j < y_j)} - P_2(x_j,\infty) - P_1(-\infty,y_j) + 2\int P_2(x,\infty)dP_1(x)]. \tag{17}$$

Acknowledgement. The author wishes to thank two referees for their extremely careful reading of the manuscript and for many helpful suggestions. The author is also indebted to H. Luschgy and D. Plachky for some useful remarks.

References

[1] Basu, D. (1970). On sufficiency and invariance. In: R. C. Bose et al. (eds.): Essays in Probability and Statistics. Univ. of North Carolina Press, Chapel Hill.

[2] Bondar, J. V., Milnes, P. (1981). Amenability: A survey for statistical applications of Hunt-Stein and related conditions on groups. Zeitschr. W.theorie 57, 103 - 128.

[3] Bondesson, L. (1975). Uniformly minimum variance estimation in location parameter families. Ann. Statist. 3, 637 - 660.

[4] Fisher, N. I. (1982). Unbiased estimation in some non-parametric families of distributions. Ann. Statist. 10, 603 - 615.

[5] Hoeffding, W. (1977a). Some incomplete and boundedly complete families of distributions. Ann. Statist. 5, 278 - 291.

[6] Hoeffding, W. (1977b). More on incomplete and boundedly complete families of distributions. In: S. Gupta and D. Moore (eds.): Statistical Decision Theory and Related Topics II, 157 - 164, Academic Press.

[7] Johnson, N. L. and Kotz, S. (1975). On some generalized Farlie-Gumbel-Morgenstern distribution. Commun. Statistics 4, 415 - 427.

[8] Lehmann, E. L. (1959). Testing Statistical Hypotheses. Wiley.

[9] Neveu, J. (1968). Processus Aleâtoires Gaussiens. Le Presse de l'Universitê de Montreal.

[10] Padmanabhan, A. R. (1974). Two comments on a paper of Bahadur. Metrika 21, 201 - 203.

[11] Rüschendorf, L. (1985). Projections and iterative procedures. In: P. R. Krishnaiah (ed.): Multivariate Analysis VI, 485 - 493, Elsevier Science Publishers B. V.

[12] Rüschendorf, L. (1984a). Symmetric functions and complete classes of distributions. To be published.

[13] Rüschendorf, L. (1984b). On the theory of unbiased estimation. To be published.

[14] Tempel'man, A. A. (1972). Ergodic theorems for general dynamical systems. Trans. Mosc. Math. Soc. 26, 94 - 132.

L. Rüschendorf
Institut für Mathematische Statistik
Universität Münster
Einsteinstraße 62
D-4400 Münster