Proc. of the 5th Pannonian Symp.
on Math. Stat., Visegrád, Hungary 1985
W. Grossmann, J. Mogyoródi, I. Vincze, W. Wertz, eds

UNBIASED ESTIMATION AND LOCAL STRUCTURE

L. RÜSCHENDORF

Münster

We develop the importance of the local structure of the estimation model to the theory of unbiased estimation, especially, in the case of large nonparametric models. On one side this gives us an useful tool for the construction of MVUE, justifying already in the finite theory intuitively appealing construction methods, which are known from the asymptotic theory. On the other side our results show under general conditions, that UMVUE typically are asymptotically efficient and, furthermore, that asymptotically efficient estimators are obtained from MVUE for a linearized estimation functional combined with an adaption procedure.

AMS subject classification: 62G05

1. INTRODUCTION

In the theory of unbiased estimation many examples are known, where unbiased estimators are very bad and, more worse, where unbiased estimators do not exist (e.g. in many models involving nuisance paramters). Furthermore, for bounded loss functions the principle of MVUE is not effective as was demonstrated by Basu (1955). A beautiful presentation of the relation of unbiasedness with different estimation principles can be found in the recent book of Lehmann (1983).

On the other hand it was shown by Sharma (1973) and Portnoy (1977), that UMVUE in regular exponential families typically are asymptotically equivalent to the MLE and, therefore, are asymptotically efficient. So the bad behaviour in these cases is due to too small sample sizes. We want to show that this behaviour extends to a general class of estimation problems, especially in nonparametric models, and that on the other side MVUE lead in a natural way to a construction method for asymptotically

In the asymptotic theory the importance of the local structure was impressively demonstrated in the LAN-theory of Le Cam and Hajek for parametric models, while for general models one can find a convincing and far reaching presentation in the recent book of Pfanzagl (1982). We want to show, that the local structure is already important in the finite unbiased theory, at least for large nonparametric models.

2. UNBIASED ESTIMATION AND TANGENT CONES

Let $\mathcal{P}$ be a class of probability measures on $(X,\mathcal{B})$ and $g:\mathcal{P}\to\mathbb{R}^1$ a function which we want to estimate assuming a quadratic loss. The leading idea of the following construction method of minimum variance unbiased estimators (MVUE) of g is that the local structure of $\mathcal{P}$ contains essential information on the estimation problem at least in large nonparametric models.

We shall concentrate in this paper on the first order tangent cones and, therefore, to linear estimation functionals g; nonlinear estimation functionals and the corresponding higher order tangent structure will be treated in a forthcoming paper. For $P\in\mathcal{P}$ let $T(P,\mathcal{P})\subset L^2(P)$ denote the set of all *tangent vectors* at P in $\mathcal{P}$, i.e. the set of all $h\in L^2(P)$ such that there exists a path $(P_t)_{t\geq 0}\subset\mathcal{P}$ with

$$\text{a) } P_0 = P$$

$$\text{b) } \left\|2\left[\left(\frac{dP_t}{dP}\right)^{\frac{1}{2}}-1\right]-th\right\| = o(t) \qquad (1)$$

$$\text{c) } P_t\left\{\frac{dP_t}{dP}=\infty\right\} = o(t^2) \text{ for } t\to 0$$

$\|\;\|$ denoting the norm in $L^2(P)$, $\frac{dP_t}{dP}$ the generalized Radon Nikodym derivative.

For the general theory of L^2-differentiation we refer to Pukelsheim (1981) and Witting (1985). The importance of L^2-differentiation and of the concept of tangent cones in asymptotic estimation theory was fully established by Pfanzagl (1982) where one can find many parametric and nonparametric examples for the calculation of $T(P,\mathcal{P})$.

If $(P_t)\subset\mathcal{P}$ and $h, \frac{dP_t}{dP}\in L^2(P)$, $t\geq 0$ such that

$$\left\|\frac{dP_t}{dP}-1-th\right\| = o(t) \text{ and } P_t\left\{\frac{dP_t}{dP}=\infty\right\} = o(t^2) \qquad ($$

then $h\in T(P,\mathcal{P})$. We remark that derivatives in the sense of (2 have been considered by Barankin (1949) and Parzen (1959) in parametric model in connection with unbiased estimation We extend this result to the general nonparametric case. Define

$$T^1(P,\mathcal{P}) = T(P,\mathcal{P}) \cup \{1\}; \qquad ($$

the constant 1 corresponding to the derivative 0 of the constant path $P_t=P$, $t\geq 0$. Furthermore let

$$D_g := \{d\in L^2(\mathcal{P});\ \int d\,dP = g(P),\ \forall P\in\mathcal{P}\} \qquad ($$

be the set of all unbiased estimators of g which are square integrable w.r.t. $\mathcal{P}$ and $D_0 = D_0(\mathcal{P})$ the corresponding class of estimators of zero

<u>Theorem 1</u>.

a) $D_0 \subset \bigcap_{P\in\mathcal{P}} T^1(P,\mathcal{P})^{\perp}$
where $T^1(P,\mathcal{P})^{\perp} = \{h\in L^2(P);\ \int hf\,dP = 0,\ \forall f\in T^1(P,\mathcal{P})\}$.

b) Let $H_P := cl\langle T^1(P,\mathcal{P})\rangle$ - the closure of the linear space generated by $T^1(P,\mathcal{P})$ in $L^2(P)$; if $d^*\in D_g\cap H_P$, then d^* is MVUE for g in P.

<u>Proof</u>. a) Let $h\in D_0$ and $f\in T(P,\mathcal{P})$, then there exists a path $(P_t)\subset\mathcal{P}$ with $P_0=P$ and L^2-derivative f. The function $F(t):=\int h\,dP_t$, $t\geq 0$ is differentiable in 0 with $F'(0)=\int hf\,dP$ (cf. Pukelsheim (1981), Pfanzagl (1982), Prop. 5.1.5 and Witting (1985), Satz 1.179, 1.190). Since $F(t)=0$, we obtain $0=F'(0)$ i.e. h is orthogonal to f. Since by assumption h is orthogonal to 1 we obtain $D_0\subset T^1(P,\mathcal{P})^{\perp}$.

b) Since by a) $D_0 \subset T^1(P,\mathcal{P})^{\perp}$ we get

$$H_P = cl\langle T^1(P,\mathcal{P})\rangle = (T^1(P,\mathcal{P})^{\perp})^{\perp} \subset D_0^{\perp} \quad .$$

Therefore, by the covariance method any $d^* \in D_g \cap H_P$ is MVUE for g in P. □

Corollary 1. If for all $P \in \mathcal{P}$ holds

$$H_P = L^2(P), \text{ then } D_0(\mathcal{P}) = \{0\} \tag{5}$$

i.e. the class $\mathcal{P}$ is L^2-complete. ($\{0\}$ denotes the set of all functions which are equal to $0[\mathcal{P}]$.)

Remark 1. a) A model with the property (5) is called a full model. Typical examples of full models are robust models (e.g. variation neighbourhood models or e.g. absolutely continuous distributions on $(\mathbb{R}^k, \mathbb{B}^k)$ w.r.t. λ^k).

b) By Theorem 1 the following method is proposed: For $d \in D_g$ determine the projection $\tilde{d}$ on H_P. If $\tilde{d} \in D_g$, then $\tilde{d}$ is MVUE for g. If $\tilde{d} \in \bigcap_{P \in \mathcal{P}} H_P$, then $\tilde{d}$ is UMVUE for g. The chance to find such an element is especially good, if $T^1(P,\mathcal{P})$ is independent of $P \in \mathcal{P}$. This projection method turns out to be especially effective when projectioning an optimal estimator from a larger model down to the actual model, in this way obtaining a finite version of the projection method known from asymptotic statistics.

3. EXAMPLES

a) Linear restrictions

Let $f_i : (X,\mathcal{B}) \to (\mathbb{R}^1, \mathbb{B}^1)$, $1 \le i \le k$,

$$\mathcal{P}_i = \{P \in M^1(X,\mathcal{B});\ f_i \in L^2(P),\ \int f_i dP = 0\},\ 1 \le i \le k, \text{ and } \mathcal{P} = \bigcap_{i=1}^{k} \mathcal{P}_i .$$

Then for $P \in \mathcal{P}$

$$T(P,\mathcal{P}_i) = \langle f_i \rangle^{\perp} \tag{6}$$

Proof. For $h \in T(P,\mathcal{P}_i)$ let $(P_t) \subset \mathcal{P}_i$ be a path with $P_0 = P$ and tangent vector h, then with $F(t) = \int f_i \, dP_t$ we obtain as in the proof of Theorem 1, $0 = F'(0) = \int f_i h \, dP$ i.e. $h \in \langle f_i \rangle^{\perp}$. If, conversely, $h \in \langle f_i \rangle^{\perp}$, let $h_t \in \langle f_i \rangle^{\perp}$ satify, $t\|h_t - h\| \to 0$ and $th_t \ge -1$ and define $P_t = (1+th_t)P$. Then (P_t) is a path in $\mathcal{P}_i$ with tangent vector h by (2), i.e. $h \in T(P,\mathcal{P}_i)$. □

As in Pfanzagl (1982), pg. 118, one now obtains

$$T(P,\mathcal{P}) = \bigcap_{i=1}^{k} T(P,\mathcal{P}_i) = \bigcap_{i=1}^{k} \langle f_i \rangle^{\perp} \quad . \tag{7}$$

Define for $c \in \mathbb{R}^k$ and $d \in D_g$

$$d_c = d + c^T f,\ f := (f_1, \dots, f_k)^T$$

and $G_P = (\int f_i f_j \, dP)_{1 \le i,j \le k}$ – the 'information' matrix of $\mathcal{P}$ in P.

Proposition 2.

a) $D_0 = \langle f_1, \dots, f_k \rangle$.

b) If $c^* \in \mathbb{R}^k$ solves the equation

$$(c^*)^T G_P + E_P df = 0, \tag{8}$$

then d_{c^*} is MVUE for g in P.

c) $d^*(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^{n} d_{c^*}(x_i)$ is MVUE for g(Q) in P^n w.r.t. $\mathcal{P}^n = \{P^n;\ P \in \mathcal{P}\}$.

Proof. a) From Theorem 1 and (7)

$D_0 \subset \bigcap_{P \in \mathcal{P}} (T^1(P,\mathcal{P}))^{\perp} = \langle f_1, \dots, f_k \rangle$ the converse inclusion is obvious.

b) By a) we obtain

$D_g = \{d_c;\ c \in \mathbb{R}^k\}$. So by the covariance method d_{c^*} is MVUE in P if and only if

$c^T E_P df + (c^*)^T G_P c = 0$ for all $c \in \mathbb{R}^k$, or, equivalently $E_P df + (c^*)^T G_P = 0$.

c) The tangent space of P^n in $\mathcal{P}^n$ is

$$T(P^n,\mathcal{P}^n) = \{ \sum_{i=1}^{n} h(x_i);\ h \in T(P,\mathcal{P})\} .$$

Since $d^*(x) = \frac{1}{n} \sum_{i=1}^{n} d_{c^*}(x_i) \in D_g \cap H_{P^n}$

Theorem 1 implies its optimality in P^n.

□

We remark that $D_0(\mathcal{P}^n)$ has been characterized by Hoeffding (1977) as the set of all functions of the form $\{ \sum_{i=1}^{k} \sum_{j=1}^{n} f_i(x_j) h_{ij}(x_{(j)})$ where $x_{(j)} := (x_1,\dots,x_{j-1}, x_{j+1},\dots,x_n)$. The formulation in terms of tangent cones allows many modifications. To give a concrete example: Let $X = \mathbb{R}^1$, $\mathcal{P}_0 := \{P \in M^1(\mathbb{R}^1, \mathbb{B}^1)\}$; P symmetric around 0, $f_i \quad L^2(P)$, $\int f_i dP = 0$, $1 \leq i \leq k\}$ then for $P^n \in \mathcal{P}_o^n$ we get

$$T(P^n,\mathcal{P}_o^n) = \{ \sum_{i=1}^{n} h(x_i) \in T(P^n,\mathcal{P}^n);\ h \text{ symmetric around } 0\}.$$

Therefore, defining

$$\tilde{d}(x_1,\dots,x_n) := \frac{1}{2n} \sum_{i=1}^{n} (d_{c^*}(x_i) + d_{c^*}(-x_i)) \qquad (9)$$

we obtain $\tilde{d} \in D_{g^n} \cap T^1(P^n,\mathcal{P}_o^n)$ and, therefore, by Theorem 1 $\tilde{d}$ is MVUE for g in P^n w.r.t. $\mathcal{P}_o^n$.

b) <u>G invariant distributions</u>

Let G be a finite group of measurable transformations of $(X,\mathcal{B})$ and let $\mathcal{P}$ be a large subset of all distributions which are invariant w.r.t. G, large meaning, that $\mathcal{P}$ has the same tangent cone as the full model has, i.e. for $P \in \mathcal{P}$

$$T(P,\mathcal{P}) = \{h \in L^2(P);\ h \circ g = h[P] \text{ for } g \in G,\ \int h\, dP = 0\} \qquad (10)$$

Let for $f \in L^2(\mathcal{P})$, $g(P) = \int f dP$. Without any 'essential' restrictions on $\mathcal{P}$ the UMVUE of g(P) after n independent observations would be

$$d_n(x) = \int f dP_{n,x} = \frac{1}{n} \sum_{i=1}^{n} f(x_i), \text{ where} \qquad (11)$$

$P_{n,x}(B) = \frac{1}{n} \sum_{i=1}^{n} 1_B\,(x_i)$ is the empirical measure. According to our general idea we consider

$$\tilde{P}_{n,x}(B) = \frac{1}{n|G|} \sum_{i=1}^{n} \sum_{g\ G} 1_{g^{-1}(B)}(x_i) \quad , \qquad (12)$$

the projection of $P_{n,x}$ on $\mathcal{P}$.

<u>Proposition 3</u>. For $f \in L^2(\mathcal{P})$ is

$$\tilde{d}(x) = \int f d\tilde{P}_{n,x} = \frac{1}{n|G|} \sum_{i=1}^{n} \sum_{g \in G} f \circ g(x_i) \text{ the UMVUE for} \qquad (13)$$

$g(P^n) = \int f dP$.

<u>Proof</u>. Since $T^1(P^n,\mathcal{P}^n) = \{ \sum_{i=1}^{n} h(x_i);\ h \in L^2(P),\ h \circ g = h[P],\ g \in G\}$ $\tilde{d} \in D_g \cap T^1(P^n,\mathcal{P}^n)$. Therefore, by Theorem 1, $\tilde{d}$ is UMVUE.

□

<u>Special case</u>. 1) If $(X,\mathcal{B}) = (\mathbb{R}^1, \mathbb{B}^1)$, $G = \{id,s\}$, where $s(x) = -x$, $x \in \mathbb{R}^1$, then $\mathcal{P}$ is a large subclass of the distributions which are symmetric around 0. If $f = 1_{(-\infty,x_0]}$, then the UMVUE for $g(P^n) = F_P(x_0)$ is

$$\tilde{F}_{n,x}(x_0) = \frac{1}{2}[F_{n,x}(x_0) + 1 - F_{n,x}((-x_0)-)] =$$

$$= \frac{1}{2n} \sum_{i=1}^{n} [1_{(-\infty,x_0]}(x_i) + 1_{[-x_0,\infty)}(x_i)] \qquad (14)$$

the symmetrized df.

2) If $(X,\mathcal{B}) = (\mathbb{R}^k, \mathbb{B}^k)$, $G = \mathcal{S}_k$ denotes the group of coordinate changes, then $\mathcal{P}$ is a large subclass of the set of symmetric

measures on $(\mathbb{R}^k, \mathcal{B}^k)$. If $f = 1_{(-\infty,x_0]}\times 1_{\mathbb{R}^{k-1}}$, then $g(P^n) = P_1(-\infty,x_0] = F_{P_1}(x_0)$ is the df of the marginal of P. The UMVUE is

$$F_{n,x}(x_0) = \frac{1}{n\cdot k}\sum_{i=1}^{n}\sum_{j=1}^{k} 1_{(-\infty,x_0]}(x_{ij}) \tag{15}$$

where $x = (x_1,\dots,x_n)$, $x_i = (x_{i1},\dots,x_{ik})$. We remark, that for the case that $\mathcal{P}$ denotes the set of all continuous symmetric probability measures one could also apply a completeness result of Smith (1969), pg. 35.

4. UNBIASED ESTIMATORS AND ASYMPTOTICALLY EFFICIENT ESTIMATORS

Consider an asymptotic estimation problem $(\mathcal{P}^n)_{n\in\mathbb{N}}$ with iid observations and let $g\colon \mathcal{P}\to\mathbb{R}^1$ be a differentiable functional which is to be estimated; differentiability of g means that for all $P\in\mathcal{P}$ there exists $g_P\in L^2(P)$ such that $\int g_P dP = 0$ and for all $\{P_t\}_{t\geq 0}\subset\mathcal{P}$ with $P_0=P$ and tangent vector h holds:

$$\lim_{t\to 0}\frac{g(P_t)-g(P)}{t} = \int g_P h dP\ . \tag{16}$$

$g_P = g(\cdot,P)$ is called *gradient of g* in P.

As is clear from the definition (16) a gradient is not uniquely determined. Call $\dot{g}_P(\cdot,P) = \dot{g}_P$ a *canonical gradient* if $\dot{g}_P$ is a gradient and if $\dot{g}_P\in \mathrm{cl}\langle T(P,\mathcal{P})\rangle$; i.e. $\dot{g}_P$ is the (unique) projection of any g_P on $\mathrm{cl}\langle T(P,\mathcal{P})\rangle$.

An estimator sequence is called asymptotically efficient (of first order) if it is as.median-unbiased and as. $N(0,\sigma^2(P))$ distributed, with $\sigma^2(P) = E_P(\dot{g}(\cdot,P))^2$. The typical as. eff. estimator sequences have a stochastic expansion of the form

$$\sqrt{n}(d_n-g(P)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{g}(x_i,P) + o_{P^n}(1). \tag{17}$$

If (d_n^*) is a UMVUE sequence for g constructed according to Theorem 1, then (d_n^*) is as. efficient.

Proposition 4. Let $d_n^*\in D_g(\mathcal{P}^n)$ satisfy

$$d_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} h(x_i),\ h\in\bigcap_{P\in\mathcal{P}} H_P,\ n\in\mathbb{N}\ ,$$

then (d_n^*) is asymptotically efficient.

Proof. Since $d_n^*\in D_g(\mathcal{P}^n)$, $n\geq 1$ we have for $P\in\mathcal{P}$, $d_1^*(x)=g(P)+$ $+f_P(x) = h(x)$, i.e. $f_P\in \mathrm{cl}\langle T(P,\mathcal{P})\rangle = H_P$. Therefore, $g(Q) =$ $=\int d_1^* dQ = g(P) + \int f_P dQ$ for all $Q\in\mathcal{P}$, implying, that $f_P = \dot{g}_P$ is the canonical gradient of g. By the CLT

$$\sqrt{n}(d_n^* - g(P))\xrightarrow{D} N(0,\int(\dot{g}_P)^2dP)$$

i.e. (d_n^*) is as. efficient.

□

Proposition 4 shows that the phenomenon detected by Sharma (1977) holds true in very general situations (at least for linear estimable functionals g). Loosely speaking one can say that for large nonparametric models UMVUE are typically as. efficient.

Let $\delta(Q,P) = \left|\left|\left(\frac{dQ}{dP}\right)^{1/2} - 1\right|\right|$ denote the Hellinger distance ($||\ ||$ in $L^2(P)$). $g_P\in L^2(P)$ is called a strong gradient of g in P w.r.t. δ if $\int g_P dP = 0$ and

$$\begin{aligned} &g(Q) - g(P) = \int g_P dQ + o(\delta(Q,P)) \text{ for } Q \text{ with}\\ &Q(\tfrac{dQ}{dP} = \infty) = o(\delta(Q,P)^2)\end{aligned} \tag{18}$$

As in Pfanzagl (1982), pg. 66, if g_P is a strong gradient, then g_P is a gradient.

Since for nonlinear functionals there are no unbiased estimators (in our context) we consider for strongly differentiable g a linearization

$$\tilde{g}(Q) = g(P) + \int\dot{g}_P dQ \tag{19}$$

of g which approximates g(Q) for P near to Q w.r.t. δ. For linear functionals $g(Q) = \int f dQ$ as considered in section 2 we have $\tilde{g}(Q) = g(Q)$. By Theorem 1 the MVUE of $\tilde{g}$ in P^n is

$$d_n^*(x) = d_n^*(x,P) = g(P) + \frac{1}{n} \sum_{i=1}^{n} \dot{g}_P(x_i) \ . \tag{20}$$

We can now try to make the MVUE d_n^* to a global estimator by an *adaption-step* namely by replacing the unknown P by a 'good' estimator $P_n = P_n(\cdot,x)$, good means according to (18), that P_n should approach P w.r.t Hellinger distance at a certain order. Define

$$\begin{aligned}\tilde{d}_n(x) = d_n^*(x,P_n(\cdot,x)) = g(P_n(\cdot,x)) + \\ + \frac{1}{n} \sum_{i=1}^{n} \dot{g}(x_i,P_n(\cdot,x));\end{aligned} \tag{21}$$

$\tilde{d}_n$ should be a good estimator for g(P) since by (18) $g(P) \approx g(P_n) + \int \dot{g}(\cdot,P_n)dP \approx \tilde{d}_n(x)$ (21) is identical with the *'improved estimator sequence'* generalizing the improvement procedure based on the Newton-Raphson approximation to the solution of the maximum likelihood equation. This procedure was introduced by Pfanzagl (1982), 11.4.2 and shown in many examples to lead to asymptotically efficient estimators. In contrast to the usual motivation for (21) namely to improve P_n, we come to equation (21) by trying to make a MVUE in P^n (for $\tilde{g}$) to a global estimator. Extension of this idea to the construction for higher order approximations of g and to higher order improvements will be considered a forthcoming paper.

Example. Let $\mathcal{P}_0 = \{P \in M^1(\mathbb{R}^1, \mathcal{B}^1),\ \int x^2 dP = m_2,\ \int x^4 dP < \infty\}$ and let $g(P) = \int x dP =: m_1(P)$. By Proposition 2 the MVUE for $g(Q^n) = m_1(Q)$ in P^n is

$$d_n^*(x_1,\ldots,x_n) = \frac{1}{n} \sum_{i=1}^{n} (x_i - c^*(x_i^2 - m_2)) \tag{22}$$

with $c^* = \frac{m_3(P)-m_2m_1(P)}{\int (x^2-m_2)^2 dP}$, $m_i(P)$ denoting the i-th moment of

The adaption step would now yield

$$\tilde{d}_n(x_1,\ldots,x_n) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \frac{\hat{m}_3-m_2\hat{m}_1}{\hat{\sigma}_2^2} (x_i^2 - m_2))$$

where $\hat{m}_i = \frac{1}{n} \sum_{k=1}^{n} x_k^i$, $\hat{\sigma}_2^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i^2 - m_2)^2$

are the corresponding sample moment estimators of the momen of the distributions.

REFERENCES

Barankin, E.W. (1949). Locally best unbiased estimates. *Ann. Math. Statist.* 10, 477-501.

Basu, D. (1955). A note on the theory of unbiased estimatio *Ann. Math. Statist.* 26, 144-145.

Hoeffding, W. (1977). Some incomplete and boundedly complet families of distributions. *Ann. Statist.* 5, 278-291.

Lehmann, E.L. (1983). *Theory of Point Estimation.* Wiley, Ne York.

Pfanzagl, J. (1982). *Contributions to a General Asymptotic Statistical Theory.* Lecture Notes in Statistics 13, Springer.

Parzen, E. (1959). Statistical inference on time series by Hilbert space methods I. In: *Time Series Analysis Pape ed. by E. Parzen, Holden Day.* 1967, pg. 251-382.

Portnoy, S. (1977). Asymptotic efficiency of minimum varian unbiased estimators. *Ann. Statist.* 5, 522-529.

Pukelsheim, F. (1981). On L_p-differentiable distributions, exponential families and the Cramer-Rao Inequality. Un published manuscript.

Sharma, D. (1973). Asymptotic equivalence for two estimator for an exponential family. *Ann. Statist.* 1, 973-980.

Smith, P.J. (1969). Structure of nonparametric tests of som multivariate hypotheses. Ph. D. thesis, Case Western Reserve University.

Stein, C. (1950). Unbiased estimates of minimum variance. *Ann. Math. Statist.* 21, 406-415.

Witting, H. (1985). *Mathematische Statistik I.* Teubner-Verlag.

Ludger Rüschendorf
Institut für Mathematische
Statistik der Universität Münster
Einsteinstraße 62
D-4400 Münster, BRD