On the attainment of the Cramer-Rao bound in Lr-differentiable families of distributions

U. Müller-Funk, Friedrich Pukelsheim, Hermann Witting

Angaben zur Veröffentlichung / Publication details:

Müller-Funk, U., Friedrich Pukelsheim, and Hermann Witting. 1989. "On the attainment of the Cramer-Rao bound in Lr-differentiable families of distributions." *The Annals of Statistics* 17 (4): 1742–48. https://doi.org/10.1214/aos/1176347392.

The Annals of Statistics
1989, Vol. 17, No. 4, 1742–1748

ON THE ATTAINMENT OF THE CRAMÉR–RAO BOUND IN $\mathbb{L}_r$-DIFFERENTIABLE FAMILIES OF DISTRIBUTIONS[1]

BY ULRICH MÜLLER-FUNK, FRIEDRICH PUKELSHEIM[2] AND HERMANN WITTING

Universität Münster, Universität Augsburg and Universität Freiburg im Breisgau

A rigorous proof is presented that global attainment of the Cramér–Rao bound is possible only if the underlying family of distributions is exponential. The proof is placed in the context of $\mathbb{L}_r(P_\vartheta)$-differentiability, requiring strong differentiability in $\mathbb{L}_r(P_\vartheta)$ of the rth root of the likelihood ratio relative to P_ϑ.

1. Introduction. It is part of the folklore of parametric statistics that the Cramér–Rao lower bound is globally attained only if the underlying family is exponential. A rigorous proof of this result depends on which concept of differentiability is adopted. Wijsman (1973) employs the logarithmic derivatives of the density functions and solves the associated differential equation, including a detailed discussion of the ensuing measurability problems. Fabian and Hannan (1977) assume weak $\mathbb{L}_2$-differentiability of the likelihood ratio; see also Barankin (1949), Section 6. Čencov (1982), Theorem 15.4 uses the notion of weak differentiability of charges.

We here place our derivation in the context of $\mathbb{L}_r$-differentiable families of distributions, that is, strong $\mathbb{L}_r$-differentiability of the rth root of the likelihood ratio. A detailed exposition of this type of differentiability is given in the textbook by Witting (1985). Ibragimov and Has'minskii (1981) work with regular experiments which essentially coincide with our continuous $\mathbb{L}_2$-differentiability. The notion of $\mathbb{L}_2$-differentiability is due to Hájek (1962), page 1124 and Le Cam (1966), Section 4.

With increasing parameter $r \geq 1$ there evolves a hierarchy of differential smoothness that is statistically meaningful: $\mathbb{L}_1$-differentiability is appropriate for deriving locally optimal tests; see Witting (1985), Sections 2.2.4 and 2.4.3. $\mathbb{L}_2$-differentiability applies to estimation problems; see Witting (1985), Section 2.7.2 or Ibragimov and Has'minskii (1981), Section I.7.2. For local asymptotic normality see Ibragimov and Has'minskii (1981), Chapter II. $\mathbb{L}_r$-differentiability, for all $r \geq 1$, holds in exponential families, reflecting their well appreciated smoothness properties.

$\mathbb{L}_r$-differentiability relates to the α-connections that play a central role in differential–geometrical methods in statistics, see Amari (1985). The relation is

Received November 1987; revised February 1989.

[1]Paper BU-533 in the Biometrics Unit, Cornell University.

[2]Research supported by the Stiftung Volkswagenwerk, Hannover, and by the Mathematical Sciences Research Institute, Cornell University.

AMS 1980 *subject classification.* 62F10.

Key words and phrases. Parametric families, regular experiments.

given by the formula $r = 2/(1-\alpha)$, with $r = 1, 2, \infty$ corresponding to $\alpha = -1, 0, 1$. However, the derivation in the present paper does not require any prerequisites from differential geometry.

In Section 2 we recall the Cramér–Rao inequality in $\mathbb{L}_2$-differentiable families (Theorem 1). Global attainment of the Cramér–Rao bound leads to a differential equation with continuous coefficients, as discussed in Section 3. The solution of this differential equation determines an exponential family (Theorem 2).

2. The Cramér–Rao inequality. First the notion of $\mathbb{L}_r$-differentiability is reviewed, since it is central to what follows. Let $\mathscr{P} = \{P_\vartheta \colon \vartheta \in \Theta\}$ be a family of probability distributions with parameter $\vartheta \in \Theta \subseteq \mathbb{R}^k$, on some fixed sample space $\mathscr{X}$ with sigma-algebra $\mathscr{B}$. The likelihood ratio of a member P_ϑ relative to another member P_{ϑ_0} is denoted by $L_{\vartheta/\vartheta_0}$,

$$L_{\vartheta/\vartheta_0}(x) = p_\vartheta(x)/p_{\vartheta_0}(x) \in [0, \infty]$$

for $(P_\vartheta + P_{\vartheta_0})$-almost all x, where p_ϑ and p_{ϑ_0} are the respective densities of P_ϑ and P_{ϑ_0} relative to some common dominating measure μ, say.

For $r \geq 1$ the $\mathbb{L}_r(P_{\vartheta_0})$-norm of a random variable T is $\|T\|_r = (\int |T|^r \, dP_{\vartheta_0})^{1/r}$, while $|\vartheta| = (\sum_{i=1}^k \vartheta_i^2)^{1/2}$ designates the Euclidean norm of the $k \times 1$ vector ϑ.

Let ϑ_0 be an interior point of Θ and let $r \geq 1$. The family $\mathscr{P}$ is called $\mathbb{L}_r(P_{\vartheta_0})$*-differentiable* when there exists a $k \times 1$ random vector $\dot{L}_{\vartheta_0}$ with components in $\mathbb{L}_r(P_{\vartheta_0})$ such that, for $\vartheta \to \vartheta_0$,

$$\left\| r\left(L_{\vartheta/\vartheta_0}^{1/r} - 1\right) - (\vartheta - \vartheta_0)^{\mathsf{T}} \dot{L}_{\vartheta_0} \right\|_r = o(|\vartheta - \vartheta_0|), \tag{1a}$$

$$P_\vartheta\left(\{L_{\vartheta/\vartheta_0} = \infty\}\right) = o(|\vartheta - \vartheta_0|^r). \tag{1b}$$

In the case $r = 1$, condition (1b) is redundant, since it follows from (1a) by Hilfssatz 1.178 in Witting (1985). If (1a) and (1b) are satisfied, then the statistic $\dot{L}_{\vartheta_0}$ is P_{ϑ_0}-almost surely unique and is called *the* $\mathbb{L}_r$*-derivative of* $\mathscr{P}$ *at* ϑ_0 or, for short, *the* $\mathbb{L}_r(P_{\vartheta_0})$*-derivative*. When $r \geq 2$ the covariance matrix

$$\mathscr{I}(\vartheta_0) = \operatorname{Cov}_{\vartheta_0}\left[\dot{L}_{\vartheta_0}\right] \tag{2}$$

is called the *information matrix of* $\mathscr{P}$ *at* ϑ_0.

The following theorem presents the Cramér–Rao bound as a property of $\mathbb{L}_2$-differentiable families of distributions.

THEOREM 1. *Suppose the family* $\mathscr{P} = \{P_\vartheta \colon \vartheta \in \Theta\}$ *is* $\mathbb{L}_2(P_{\vartheta_0})$*-differentiable at an interior point* ϑ_0 *of* $\Theta \subseteq \mathbb{R}^k$. *For some dimension* $l \geq 1$ *let* T *be an* $l \times 1$ *statistic whose components have second moments in a neighborhood around* ϑ_0,

$$\limsup_{\vartheta \to \vartheta_0} \operatorname{Var}_\vartheta\left[T_j\right] < \infty \quad \textit{for all} \quad j \leq l.$$

Then the mean-value function $\vartheta \mapsto E_\vartheta[T]$ *is differentiable at* ϑ_0 *with Jacobian matrix* $\mathscr{G}(\vartheta_0)$, *say, and the covariance matrix obeys the Cramér–Rao inequality*

$$\operatorname{Cov}_{\vartheta_0}[T] \geq \mathscr{G}(\vartheta_0) \mathscr{I}(\vartheta_0)^{-} \mathscr{G}(\vartheta_0)^{\mathsf{T}}. \tag{3}$$

Moreover, equality holds in (3) *if and only if*

$$T(x) - E_{\vartheta_0}[T] = \mathscr{G}(\vartheta_0)\mathscr{I}(\vartheta_0)^{-}\dot{L}_{\vartheta_0}(x) \tag{4}$$

for P_{ϑ_0}-almost all x. The expressions in (3) *and* (4) *do not depend on the choice of the generalized inverse for $\mathscr{I}(\vartheta_0)$.*

PROOF. See Witting (1985), Satz 2.133, or Ibragimov and Has'minskii (1981), Theorem I.7.3. □

The merits of this theorem lie in that it establishes the Cramér–Rao inequality as a property of the underlying family $\mathscr{P}$, covering all statistics except those which show an aberrant variance behavior around ϑ_0 anyway, compare Pitman (1979), page 39. In contrast, Joshi (1976) presents the Cramér–Rao inequality as a joint property of both, the underlying family of distributions and the particular estimator under investigation.

Next we proceed to show that global attainment of the Cramér–Rao bound entails exponentiality. The essential step of the proof entirely relies on the weakest type of $\mathbb{L}_r$-differentiability, namely, on $\mathbb{L}_1$-differentiability.

3. Attainment of the Cramér–Rao bound. In this section we study the global attainment of the Cramér–Rao bound (4), with statistic T and parameter ϑ having the same dimensionality, $l = k$. Let every parameter vector ϑ be an interior point of Θ, that is, the parameter domain Θ is open. Let us consider the equality

$$\mathrm{Cov}_\vartheta[T] = \mathscr{G}(\vartheta)\mathscr{I}(\vartheta)^{-}\mathscr{G}(\vartheta)^{\mathsf{T}} \quad \text{for all } \vartheta \in \Theta.$$

When the covariance matrix of T is nonsingular, the matrices $\mathscr{G}(\vartheta)$ and $\mathscr{I}(\vartheta)$ are nonsingular as well. Thus (4) leads to the differential equation

$$\dot{L}_\vartheta = \mathscr{A}(\vartheta)^{\mathsf{T}}T - b(\vartheta),$$

with $\mathscr{A}(\vartheta)^{\mathsf{T}} = \mathscr{I}(\vartheta)\mathscr{G}(\vartheta)^{-1}$ and $b(\vartheta) = \mathscr{I}(\vartheta)\mathscr{G}(\vartheta)^{-1}E_\vartheta[T]$. In other words, the derivative $\dot{L}_\vartheta$ is an affine transformation of a statistic T not depending on ϑ, with coefficients $\mathscr{A}(\vartheta)$ and $b(\vartheta)$ depending on ϑ. To solve this differential equation it is helpful to have the coefficients depend on ϑ continuously.

To this end we introduce continuous $\mathbb{L}_r$-differentiability. Let ϑ vary over a neighborhood of ϑ_0. The derivative $\dot{L}_\vartheta$ is a member of the space $\mathbb{L}_r^k(P_\vartheta)$ with norm $\|S\|_{r,\vartheta} = \sum_{i=1}^k(\int|S_i|^r\,dP_\vartheta)^{1/r}$, that is, it is a $k \times 1$ random vector with components in $\mathbb{L}_r(P_\vartheta)$. Multiplication with $L_{\vartheta/\vartheta_0}^{1/r}$ yields a member of the space $\mathbb{L}_r^k(P_{\vartheta_0})$. Thus the family $\mathscr{P}$ is called *continuously* $\mathbb{L}_r(P_{\vartheta_0})$*-differentiable* when $\mathscr{P}$ is $\mathbb{L}_r(P_\vartheta)$-differentiable for all ϑ in a neighborhood of ϑ_0 and when, for $\vartheta \to \vartheta_0$,

$$\|\dot{L}_\vartheta L_{\vartheta/\vartheta_0}^{1/r} - \dot{L}_{\vartheta_0}\|_{r,\vartheta_0} = o(1), \tag{5a}$$

$$\|\dot{L}_\vartheta 1_{\{L_{\vartheta/\vartheta_0}=\infty\}}\|_{r,\vartheta} = o(1). \tag{5b}$$

It is straightforward to show that continuous $\mathbb{L}_2$-differentiability of $\mathscr{P}$ on Θ implies that the information matrix $\mathscr{I}(\vartheta)$ in (2) and the Jacobian matrix $\mathscr{G}(\vartheta)$ in

(3) depend continuously on ϑ. Again, smoothness increases with r; in particular, continuous $\mathbb{L}_2(P_{\vartheta_0})$-differentiability implies continuous $\mathbb{L}_1(P_{\vartheta_0})$-differentiability.

The only additional assumption not mentioned so far is that the parameter domain Θ be connected so that any two points ϑ_0 and ϑ can be joined by a continuous path ϑ_s, with $s \in [0,1]$. Let $\mathrm{GL}(k)$ denote the general linear group of nonsingular $k \times k$ matrices.

THEOREM 2. *Suppose the family* $\mathscr{P} = \{P_\vartheta\colon \vartheta \in \Theta\}$ *has a parameter domain* $\Theta \subseteq \mathbb{R}^k$ *that is open and connected. Let* T *be a* $k \times 1$ *random vector; when the Jacobian matrices of the mean-value function* $\vartheta \mapsto E_\vartheta[T]$ *exist they are denoted by* $\mathscr{G}(\vartheta)$. *Then the following three statements are equivalent*:

(a) $\mathscr{P}$ *is an exponential family of order* k *in* T *and* $\alpha(\vartheta)$, *for some continuously differentiable mapping* $\alpha\colon \Theta \to \mathbb{R}^k$ *whose Jacobian matrices* $\mathscr{A}(\vartheta)$ *are nonsingular.*

(b) $\mathscr{P}$ *is continuously* $\mathbb{L}_2$*-differentiable on* Θ, *and the covariance matrices* $\mathrm{Cov}_\vartheta[T]$ *are nonsingular and attain the Cramér–Rao bound* $\mathrm{Cov}_\vartheta[T] = \mathscr{G}(\vartheta)\mathscr{I}(\vartheta)^{-1}\mathscr{G}(\vartheta)^{\mathsf{T}}$ *for all* $\vartheta \in \Theta$.

(c) $\mathscr{P}$ *is continuously* $\mathbb{L}_1$*-differentiable on* Θ *and the derivatives* $\dot{L}_\vartheta$ *admit a representation* $\dot{L}_\vartheta = \mathscr{A}(\vartheta)^{\mathsf{T}}T - b(\vartheta)$ *for all* $\vartheta \in \Theta$, *where the mappings* $\mathscr{A}\colon \Theta \to \mathrm{GL}(k)$ *and* $b\colon \Theta \to \mathbb{R}^k$ *are continuous, and where the distributions of* T *under* $\mathscr{P}$ *do not concentrate on a proper affine subspace of* $\mathbb{R}^k$.

PROOF. That (a) implies (b) is easy to verify. That (b) implies (c) follows from the smoothness hierarchy mentioned below (5b). It remains to show that (c) implies (a).

Fix $\vartheta_0, \vartheta \in \Theta$, and choose a continuously differentiable path ϑ_s from ϑ_0 to ϑ, with $s \in [0,1]$. Its derivative with respect to s is denoted by $\dot{\vartheta}_s$. For $s \in [0,1]$ and $x \in \mathscr{X}$ define

$$g_{\mathscr{A}}(s) = \mathscr{A}(\vartheta_s)\dot{\vartheta}_s, \qquad \alpha(\vartheta) = \int_0^1 g_{\mathscr{A}}(s)\,ds,$$

$$g_b(s) = \dot{\vartheta}_s^{\mathsf{T}} b(\vartheta_s), \qquad \kappa(\vartheta) = \int_0^1 g_b(s)\,ds,$$

$$f(x) = \exp\left(\int_0^1 \dot{\vartheta}_s^{\mathsf{T}} \dot{L}_{\vartheta_s}(x)\,ds\right) = \exp\big(\alpha(\vartheta)^{\mathsf{T}} T(x) - \kappa(\vartheta)\big).$$

In view of the continuity assumptions these quantities are well defined, and f is measurable.

We claim that f is a P_{ϑ_0}-density of P_ϑ. Then neither f nor $\kappa(\vartheta)$ will depend on the path ϑ_s that enters into the definition, and the same will be true for $\alpha(\vartheta)$ since the distributions of T do not concentrate on a proper affine subspace. In order to establish our claim we must verify that, for every event $B \in \mathscr{B}$,

$$\int_B f\,dP_{\vartheta_0} = P_\vartheta(B). \tag{6}$$

For arbitrary $\varepsilon > 0$ there exists a partitioning of $\mathbb{R}^k$ into measurable rectangles $R_1, R_2, \dots$ of diameter no greater than ε. Let the set $B \in \mathscr{B}$ be fixed, and define $B_i = B \cap T^{-1}(R_i)$.

First we show that if B_i has positive probability under ϑ_0, then the same is true under ϑ; since ϑ_0 and ϑ are arbitrary, this actually proves that the distributions in $\mathscr{P}$ are pairwise equivalent. The function $s \mapsto P_{\vartheta_s}(B_i)$ is continuously differentiable, compare Witting [(1985), Satz 1.179]. We argue that it cannot start with $P_{\vartheta_0}(B_i) > 0$ and finish with $P_\vartheta(B_i) = 0$. Let us assume this is the case and, without loss of generality, let us assume $P_{\vartheta_s}(B_i)$ to be positive for intermediate values $s \in (0,1)$. Then the function $H(s) = \log P_{\vartheta_s}(B_i)$ is finite and differentiable for $s < 1$, but equals $-\infty$ for $s = 1$. Hence its derivatives $h(s)$ are unbounded, tending to $-\infty$ as s converges to 1.

On the other hand, the differentiability assumptions of part (c) imply that, for $s < 1$,

$$h(s) = \frac{\dot{\vartheta}_s^{\mathsf{T}} \nabla P_{\vartheta_s}(B_i)}{P_{\vartheta_s}(B_i)} = \frac{1}{P_{\vartheta_s}(B_i)} \dot{\vartheta}_s^{\mathsf{T}} \int_{B_i} \dot{L}_{\vartheta_s} \, dP_{\vartheta_s} = m(s|B_i) - g_b(s), \tag{7}$$

where $m(s|B_i) = \int_{B_i} \dot{\vartheta}_s^{\mathsf{T}} \mathscr{A}(\vartheta_s)^{\mathsf{T}} T \, dP_{\vartheta_s} / P_{\vartheta_s}(B_i)$—the conditional expectation under ϑ_s of $\dot{\vartheta}_s^{\mathsf{T}} \mathscr{A}(\vartheta_s)^{\mathsf{T}} T$, given B_i—exists since $\mathscr{A}(\vartheta_s)^{\mathsf{T}} T = \dot{L}_{\vartheta_s} + b(\vartheta_s)$ is integrable. The points $\dot{\vartheta}_s^{\mathsf{T}} \mathscr{A}(\vartheta_s)^{\mathsf{T}} T(x)$, with $x \in B_i$, lie in the image of the set R_i under the mapping $\dot{\vartheta}_s^{\mathsf{T}} A(\vartheta_s)^{\mathsf{T}}$, and hence are bounded by

$$\max_{s \in [0,1]} |\mathscr{A}(\vartheta_s) \dot{\vartheta}_s| \cdot \sup_{t \in R_i} |t| = c \cdot \rho_i, \tag{8}$$

say. It follows that $\sup_{s \in [0,1)} |h(s)| \le c\rho_i + \max_{s \in [0,1]} g_b(s) < \infty$. This contradicts the earlier conclusion that $h(s)$ is unbounded. Hence we have shown that, for $s \in [0,1]$, the probabilities $P_{\vartheta_s}(B_i)$ either stay positive throughout or else vanish identically. In the latter case, (6) evidently holds with B_i in place of B.

Next we verify (6) when $P_{\vartheta_s}(B_i)$ stays positive. In this case, the function $H(s) = \log P_{\vartheta_s}(B_i)$ is the integral of its derivative $h(s)$, for all $s \in [0,1]$. Thus we have

$$\exp\left(\int_0^1 h(s) \, ds \right) = \exp(H(1) - H(0)) = \frac{P_\vartheta(B_i)}{P_{\vartheta_0}(B_i)}.$$

Upon replacing f by its definition and inserting $g_b(s) = m(s|B_i) - h(s)$ from (7), we obtain

$$\begin{aligned} \int_{B_i} f \, dP_{\vartheta_0} &= \int_{B_i} \exp\left(\int_0^1 \left\{ \dot{\vartheta}_s^{\mathsf{T}} \mathscr{A}(\vartheta_s)^{\mathsf{T}} T - g_b(s) \right\} ds \right) dP_{\vartheta_0} \\ &= \int_{B_i} \exp\left(\int_0^1 \left\{ \dot{\vartheta}_s^{\mathsf{T}} \mathscr{A}(\vartheta_s)^{\mathsf{T}} T - m(s|B_i) \right\} ds \right) dP_{\vartheta_0} \frac{P_\vartheta(B_i)}{P_{\vartheta_0}(B_i)}. \end{aligned}$$

Again the points $\dot{\vartheta}_s^{\mathsf{T}}\mathscr{A}(\vartheta_s)^{\mathsf{T}}T(x)$, with $x \in B_i$, lie in the image of the set R_i under the mapping $\dot{\vartheta}_s^{\mathsf{T}}\mathscr{A}(\vartheta_s)^{\mathsf{T}}$, and—being an average—so does $m(s|B_i)$. Therefore, the distance between $\dot{\vartheta}_s^{\mathsf{T}}\mathscr{A}(\vartheta_s)^{\mathsf{T}}T(x)$ and $m(s|B_i)$ is bounded from above by $|\mathscr{A}(\vartheta_s)\dot{\vartheta}_s|\operatorname{diam}(R_i) \leq c\varepsilon$, with the same constant c as in (8). Thus the inner integral is bounded by $\pm c\varepsilon$. Summation over i gives

$$e^{-c\varepsilon}P_\vartheta(B) \leq \int_B f\,dP_{\vartheta_0} \leq e^{c\varepsilon}P_\vartheta(B).$$

Since ε is arbitrary our claim that f is a P_{ϑ_0}-density of P_ϑ is established. Hence $\mathscr{P}$ is an exponential family.

Finally we investigate α. Fixing ϑ_0 and varying ϑ defines α on all of Θ, with $\alpha(\vartheta_0) = 0$. Next we show that α is differentiable at ϑ_0. For a point ϑ close to ϑ_0 we may choose a straight-line path $\vartheta_s = \vartheta_0 + s(\vartheta - \vartheta_0)$, whence $\alpha(\vartheta) = \int_0^1\mathscr{A}(\vartheta_s)(\vartheta - \vartheta_0)\,ds$. Then we have

$$\frac{|\alpha(\vartheta) - \alpha(\vartheta_0) - \mathscr{A}(\vartheta_0)(\vartheta - \vartheta_0)|}{|\vartheta - \vartheta_0|} \leq \int_0^1 \frac{|(\mathscr{A}(\vartheta_s) - \mathscr{A}(\vartheta_0))(\vartheta - \vartheta_0)|}{|\vartheta - \vartheta_0|}\,ds.$$

Since this tends to 0 as ϑ tends to ϑ_0, it follows that α is differentiable at ϑ_0, with Jacobian matrix $\mathscr{A}(\vartheta_0)$.

Fixing ϑ and varying ϑ_0 we similarly know that for $\vartheta_1 \neq \vartheta_0$ the distribution P_ϑ has a P_{ϑ_1}-density proportional to $\exp(\alpha_1(\vartheta)^{\mathsf{T}}T)$, where α_1 is differentiable at ϑ_1 and has Jacobian matrix $\mathscr{A}(\vartheta_1)$. The chain rule $dP_\vartheta/dP_{\vartheta_0} = (dP_\vartheta/dP_{\vartheta_1})(dP_{\vartheta_1}/dP_{\vartheta_0})$ leads to $\alpha(\vartheta) = \alpha(\vartheta_1) + \alpha_1(\vartheta)$. Hence, since α_1 is differentiable at ϑ_1 so is α, and their common Jacobian matrix is $\mathscr{A}(\vartheta_1)$. Thus α is differentiable on Θ and has nonsingular Jacobian matrices $\mathscr{A}(\vartheta)$.

This also implies that α is an open mapping, hence the image set $\alpha(\Theta)$ is an open subset of the canonical parameter domain of the exponential family $\mathscr{P}$. Since T is not concentrated on a proper affine subspace, the family $\mathscr{P}$ must be of order k. □

Note that Theorem 2 provides a particular instance where $\mathbb{L}_1$-differentiability entails $\mathbb{L}_r$-differentiability, for all $r \geq 1$.

Acknowledgment. The authors thank W. R. van Zwet for many constructive suggestions that led to a great improvement of the original manuscript.

REFERENCES

AMARI, S. (1985). *Differential–Geometrical Methods in Statistics. Lecture Notes in Statist.* **28**. Springer, New York.

BARANKIN, E. W. (1949). Locally best unbiased estimates. *Ann. Math. Statist.* **20** 477–501.

ČENCOV, N. N. (1982). *Statistical Decision Rules and Optimal Inference. Trans. Math. Monographs* **53**. Amer. Math. Soc., Providence, R.I.

FABIAN, V. and HANNAN, J. (1977). On the Cramér–Rao inequality. *Ann. Statist.* **5** 197–205.

HÁJEK, J. (1962). Asymptotically most powerful rank-order tests. *Ann. Math. Statist.* **33** 1124–1147.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation. Asymptotic Theory.* Springer, New York.

JOSHI, V. M. (1976). On the attainment of the Cramér–Rao lower bound. *Ann. Statist.* **4** 998–1002.

LE CAM, L. (1966). Likelihood functions for large numbers of independent observations. In *Research Papers in Statistics. Festschrift for Jerzy Neyman* (F. N. David, ed.) 167–187. Wiley, New York.

PITMAN, E. J. G. (1979). *Some Basic Theory for Statistical Inference*. Chapman and Hall, London.

WIJSMAN, R. A. (1973). On the attainment of the Cramér–Rao lower bound. *Ann. Statist.* **1** 538–542.

WITTING, H. (1985). *Mathematische Statistik I. Parametrische Verfahren bei festem Stichprobenumfang.* Teubner, Stuttgart.

ULRICH MÜLLER-FUNK
INSTITUT FÜR MATHEMATISCHE STATISTIK
WESTFÄLISCHE WILHELMS-UNIVERSITÄT
D-4400 MÜNSTER
FEDERAL REPUBLIC OF GERMANY

FRIEDRICH PUKELSHEIM
INSTITUT FÜR MATHEMATIK
UNIVERSITÄT AUGSBURG
D-8900 AUGSBURG
FEDERAL REPUBLIC OF GERMANY

HERMANN WITTING
INSTITUT FÜR MATHEMATISCHE STOCHASTIK
ALBERT–LUDWIGS-UNIVERSITÄT
D-7800 FREIBURG IM BREISGAU
FEDERAL REPUBLIC OF GERMANY