On the Variance of Estimates with Prescribed Expectations

By L. Le Cam
University of California, Berkeley
NSF DMS 9001710

1. Introduction.

Let $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ be an experiment given by measures P_θ on a σ-field $\mathcal{A}$. Let $\mathcal{D}$ be a class of real valued functions defined on Θ. What can one say about the variance of estimates T defined on $\mathcal{E}$ and such that the function $\theta \rightsquigarrow \int T dP_\theta$ belongs to $\mathcal{D}$?

This problem and variants of it have been studied at great length in the statistical literature. Perhaps one of the best known result linked to the problem is the Cramér-Rao inequality with its family of variants.

An article of E.W. Barankin (1949) showed that, using a small amount of linear space theory, one can give accurate bounds that, as he shows, imply all the Cramér-Rao type inequalities published at that time.

The main point of Barankin's paper however is that his bound is attainable. That is, under a mild condition, there do exist estimates T that achieve Barankin's bound.

In this paper we reconsider Barankin's formulation extending it somewhat to cover the case where one desires to obtain the minimax variance. This means finding estimates T with expectation in $\mathcal{D}$ as before but such that $\sup_\theta \mathrm{Var}_\theta T$ be minimum. (Here $\mathrm{Var}_\theta T$ is the variance of T under P_θ).

Barankin had considered only the problem of minimizing $\mathrm{Var}_{\theta_0} T$ or other moments for a specified θ_0. As we shall see the minimax problem is somewhat more complex than the problem of minimizing a variance at a point θ_0.

One of the aim of the present paper is to show that whether the problem is one of minimizing $\mathrm{Var}_{\theta_0} T$ or $\sup_\theta \mathrm{Var}_\theta T$, one can, in principle, get a solution by solving a family of two-dimensional problems.

Unfortunately the solutions involve integrals of the form $\int \frac{dPdQ}{dM}$ for measures P, Q and M that are convex combinations of the P_θ's.

These integrals are not easily computable or tractable in practice. Our hope was that some relevant feature would shine through and help in directing the arguments. This has not happened so far.

Specifically, the contents of the paper are as follows. Section 2 sets the problem in a linear space framework in the general spirit of Barankin [1949]. this means introducing the linear span $S(\mathcal{E})$ of the set $\{P_\theta : \theta \in \Theta\}$ in the space of finite signed measures in the σ-field $\mathcal{A}$. One expresses a quality requirement through a subset C of the (algebraic) dual of $S(\mathcal{E})$. Without any further conditions this system yields lower bounds that Barankin had already shown to be better than all lower bounds available publicly in 1949.

To get the *existence* of estimates that satisfy the desiderata, one requires that C contains the origin and that it be closed and convex. In many cases one can show that the proposed set C is closed in the dual of $S(\mathcal{E})$ by showing that it is compact for pointwise convergence on $S(\mathcal{E})$. This is done for several cases in Section 3. Section 4 is about the problem of finding estimates with prescribed expectations for all P_θ and minimum variance under a particular measure Q. Although the solution is implicit in Theorem 1, Section 2, we give a direct solution, assuming that Q dominates all the P_θ and that the densities dP_θ/dQ are square integrable for Q. These are the conditions in Barankin (1949). We give examples showing that without any conditions the minimum variance for Q may be zero but that no estimate can achieve an actual zero. The results can be written in various forms. We give one in the form of a modulus of continuity equivalent to the form prefered by J. Kiefer [1952].

To terminate this section we place the result of Fabian-Hannan [1977] in the context of the present theory to complete Barankin's assertion that his bound covers all known Cramér-Rao type bounds. Of course, neither Kiefer [1952], nor Fabian and Hannan [1977] can claim that their bounds are attainable without using Barankin's arguments.

Section 5 is an aside. It shows that the matrices or kernels that occur in Section 4 also arise in a different problem, that of finding estimates that minimize an expected square deviation. This is done here (as in Le Cam [1986]) in a Bayesian context.

It is also shown that the matrices in question determine the type of the experiment $\mathcal{E}$. Section 6 contains a minimax theorem that reduces the minimax variance problem to two-dimensional ones involving only three probability measures at a time. These measures are finite convex combinations of the initial ones that formed the experiment $\mathcal{E}$.

Section 7 considers more specifically the case where $\mathcal{E}$ is a binary experiment, $\mathcal{E} = \{P_1, P_2\}$. This is of course the simplest case. Even there, a complete solution is not available, but certain usable inequalities are. We further show that in passages to the limit that give inequalities similar to the classical Cramér-Rao ones, one can sometimes, but not always, substitute Hellinger distances for the distances issued from the formula

$$k^2(P,Q) = \frac{1}{2}\int \frac{(dP - dQ)^2}{d(P+Q)}.$$

Section 8 reconsiders the problem of minimizing the variance under P_1 of an estimate with specified expectation at P_1 and P_2. It shows that whenever P_2 is dominated by P_1 but dP_2/dP_1 is not square integrable for P_1 the variance under P_1 can be made as close to zero as one wishes, but, of course, it cannot be made exactly zero.

Section 9 takes up the problem of substituting Hellinger distances for the Hilbert norms of Section 4 for the case where the parameter space Θ is finite.

Section 10 is a mix of history and examples of applications, mostly of the two point formulas of Section 7. It contains some results inspired by a paper of Simons and Woodroofe [1983]. Up to this section we had not used any particular structure on the index set Θ. Here we use a linear structure to relate our previous inequalities to a Cramér-Rao inequality in infinite dimensional linear spaces.

We have not attempted full coverage. In particular we have not said anything about the use of differential inequalities as in Berkson and Hodges [1962] on L. Brown and Gajek [1990] or Brown and Low [1991].

2. A formulation in linear spaces.

Let $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ be an experiment given by probability measures P_θ on a σ-field $\mathcal{A}$ of subsets of a set $\mathcal{X}$. The set Θ can be arbitrary.

Consider a function γ from Θ to $(-\infty, +\infty)$. A problem that has been considered often in the statistical literature can be stated as follows: Do there exist estimates T defined on $\mathcal{E}$ and such that $\int T dP_\theta = \gamma(\theta)$ for all $\theta \in \Theta$? If so, what can one say about their qualities or lack of them?

To study this problem we shall use a linear space formulation imitated from Barankin [1949]. There are several spaces that introduce themselves naturally in the problem. One of them, to be called $S(\mathcal{E})$, for span of $\mathcal{E}$, is the space of finite signed measures that are finite linear combinations $\sum_\theta \alpha_\theta P_\theta$ of the P_θ's of $\mathcal{E}$.

Another space, say $S_0(\mathcal{E})$, is the space of finite linear combinations of the form $\sum \alpha_{s,t}(P_s - P_t)$.

An estimate T defined by $\mathcal{E}$ and such that $\int |T| dP_\theta < \infty$ for all θ determines linear functionals φ_T on $S(\mathcal{E})$ and its subspace $S_0(\mathcal{E})$ by $\langle \varphi_T, \mu \rangle = \int T d\mu$ if $\mu = \sum \alpha_\theta P_\theta \in S(\mathcal{E})$.

The restriction of φ_T to $S_0(\mathcal{E})$ does not differentiate between T and a shift $T + c$ where c is not random. This is a convenience if one wants to study variances or other shift invariant measures of dispersion.

Barankin [1949] noted that if one prescribes an expectation function γ by $\gamma(\theta) = \int T dP_\theta$ for some estimate T this defines a linear functional on the space $S(\mathcal{E})$.

Of course, if one prescribed a function γ on Θ there might not be any linear functional on $S(\mathcal{E})$ that corresponds to it. Even if there is such a functional it might not be obtainable from an estimate. Even if it is, the estimate might be a very poor one.

To take such possibilities into account we shall suppose given a certain set C of estimates T such that $\int |T| dP_\theta < \infty$ for $\theta \in \Theta$. These estimates will serve as standard of quality. In Theorem 1 below we shall measure the difficulty of producing a prescribed expectation function by the smallest $r \geq 0$ such that there is a $T \in rC$ with the prescribed expectations.

For the selected C, the restrictions of its elements to $S(\mathcal{E})$ form a subset C^* of the space S' of linear functionals on $S(\mathcal{E})$. The further restrictions of these to $S_0(\mathcal{E})$ will be denoted C_0^*.

The set C^* has a polar C^o in $S(\mathcal{E})$. This is the subset of $S(\mathcal{E})$ formed by signed measures μ such that $\langle \varphi, \mu \rangle \leq 1$ for all $\varphi \in C^*$. Similarly we have a polar C_o^o in $S_o(\mathcal{E})$. By abuse of language we shall call these polar sets the polars of C. This should entail no confusion.

The following statements are easy to prove.

Lemma 1 *There are linear functionals φ on $S(\mathcal{E})$ such that $\langle \varphi, P_\theta \rangle = \gamma(\theta)$ for all $\theta \in \Theta$ if and only if for every finite linear combination $\sum_\theta \alpha_\theta P_\theta$, the equality $\sum \alpha_\theta P_\theta = 0$ implies $\sum \alpha_\theta \gamma(\theta) = 0$.*

The necessity is visible. The sufficiency can be proved by an application of Zorn's lemma.

Lemma 2 *In order that there be a $\varphi \in C^*$ such that $\langle \varphi, P_\theta \rangle = \gamma(\theta)$ for all $\theta \in \Theta$ it is necessary that $\sum \alpha_\theta \gamma(\theta) \leq 1$ for every finite linear combination such that $\sum \alpha_\theta P_\theta \in C^o$.*

One can write similar statements for $S_0(\mathcal{E})$. For instance Lemma 2 applied to $S_0(\mathcal{E})$ would say that there is a $\varphi \in C_0^*$ such that $\langle \varphi, P_s - P_t \rangle = \gamma(s) - \gamma(t)$, $s, t \in \Theta$ only if $\sum \alpha_{s,t}[\gamma(s) - \gamma(t)] \leq 1$ for all combinations such that $\sum \alpha_{s,t}(P_s - P_t) \in C_0^o$.

Lemma 2 is a simple consequence of the definition of polar sets. Note that it does not say anything about the conditions being sufficient, only about "necessity". Nevertheless, Barankin [1949] observed that such a Lemma implied all the Cramér-Rao type inequalities that were published at that time. It is clear that it also implies the Chapman-Robbins result of [1951] and Kiefer's inequality [1952]. In Section 4 we shall show that the more recent inequality of Fabian and Hannan [1977] is also a consequence of Lemma 2, taking into account the convexity of a polar such as C^o.

This is so for matters of "necessity". However the great merit of Barankin's paper is that it does not stop there. He provides conditions under which the necessary condition of Lemma 2 becomes a sufficient one.

We shall state conditions for C^* on $S(\mathcal{E})$. One can state similar conditions for C_0^* on $S_0(\mathcal{E})$. Consider C^* as a subset of the algebraic dual S' of $S(\mathcal{E})$. The two spaces S' and $S(\mathcal{E})$ are in separated duality for the weak topologies they induce on each other. Consider then the following conditions

Condition (OCC): The set C satisfies (OCC) if it has the following properties

1) $0 \in C^*$,

2) C^* *is convex*

3) C^* *is* $w[S', S(\mathcal{E})]$ *closed in* S'

The third condition is often difficult to verify. We shall often replace it by a stronger condition (3^+)

(3^+) C^* *is compact for pointwise convergence on* $S(\mathcal{E})$.

If (1), (2) and (3^+) holds we shall refer to condition $(\text{OCC})^+$.

One can formulate similar conditions for the restriction C_0^* of the elements of C^* to $S_0(\mathcal{E})$. In this case we shall refer to $(\text{OCC})_0$ and $(\text{OCC})_0^+$.

Note in passing that it may be very much easier to verify the compactness requirement for $(\text{OCC})_0^+$ than for $(\text{OCC})^+$. This yield the following result in which r denotes a non-negative real number

Theorem 1 *Assume that C satisfies condition* (OCC) *and let γ be a real valued function defined on Θ.*

There exist a $T \in rC$ such that $\int T dP_\theta = \gamma(\theta)$ for all θ if and only if $\sum \alpha_\theta \gamma(\theta) \leq r$ for all finite linear combinations such that $\sum \alpha_\theta P_\theta \in C^o$.

Proof We already know from Lemma 2 that the condition is necessary. For a given r, the condition implies, by Lemma 1, that there is a $\varphi \in S'$ such that $\langle \varphi, P_\theta \rangle = \gamma(\theta)$. This φ is well defined and our condition means $\langle \varphi, \mu \rangle \leq r$ for $\mu \in C^o$. Equivalently φ belongs to the polar rC^{oo} of the set $\frac{1}{r}C^o$.

The bipolar theorem says that the bipolar C^{oo} is the closed convex hull of C^* and 0. Since $0 \in C^*$ and since C^* is closed and convex, one has $C^{oo} = C^*$. Since C^* is itself the image of C there is a $T \in C$ such that $\langle \varphi, P_\theta \rangle = \int T dP_\theta$. Hence the result.

Note. If the condition is satisfied for some r, it is satisfied for a smallest one given by

$$r = \sup_\alpha \{\sum \alpha_\theta \gamma(\theta); \sum \alpha_\theta P_\theta \in C^o\}.$$

The preceding theorem admits an extension that would be more relevant for practical purposes. Indeed, in practice, it would be unwise to insist on satisfying the relations $\int T dP_\theta = \gamma(\theta)$ *exactly.* Such insistence can be of value in some theoretical questions, hardly ever in practical ones. One would readily be satisfied to get an expectation function $\theta \rightsquigarrow \int T dP_\theta$ that belongs to some small neighborhood of a desirable γ.

Let us see what happens if we accept every possibility in a certain class $\mathcal{D}$.

Theorem 2 *Let C satisfy condition (OCC). Let $\mathcal{D}$ be a nonempty class of real valued functions defined on Θ. Assume that $\mathcal{D}$ is convex and compact for pointwise convergence on Θ.*

Then there is a $T \in rC$ with an expectation function $\theta \rightsquigarrow \int T dP_\theta$ belonging to $\mathcal{D}$ if and only if

$$\sup_\alpha \inf_f \{\sum \alpha_\theta f(\theta); f \in \mathcal{D}, \sum \alpha_\theta P_\theta \in C^o\} \leq r$$

Remark Note the order "sup inf", not "inf sup". The sup is over the finite linear combinations such that $\sum \alpha_\theta P_\theta \in C^o$.

Proof Let A be the set of finite linear combinations $\alpha = \{\alpha_\theta; \theta \in \Theta\}$ such that $\sum \alpha_\theta P_\alpha \in C^o$. This is a convex set. For each $f \in \mathcal{D}$ define a function φ_f on A by $\varphi_f(\alpha) = \sum \alpha_\theta f(\theta)$.

Let $\mathcal{R}_0$ be the set of functions obtained in that manner for $f \in \mathcal{D}$. It is a convex pointwise compact set. Let $\mathcal{R}$ be the set of functions of the form $\varphi + u$ with $\varphi \in \mathcal{R}_0$, $u \geq 0$. It is a convex set closed for pointwise convergence. According to the Hahn-Banach theorem a function ψ on A belongs to $\mathcal{R}$ if and only if

$$\int \psi d\pi \geq \inf\{\int \varphi d\pi; \varphi \in \mathcal{R}_0\}$$

for all probability measures π with finite support on A. However, since A is convex $\int \psi d\pi$ is just the evaluation of ψ at some point of A. Thus if we define r_0 as $\sup_\alpha \inf_f \{\sum \alpha_\theta f(\theta);$

$f \in \mathcal{D}$, $\sum \alpha_\theta P_\theta \in C^o\} = \sup_\alpha \inf_\psi \{\psi(\alpha);\ \psi \in \mathcal{R},\ \alpha \in A\}$, there is an $f \in \mathcal{D}$ such that $\varphi_f(\alpha) \leq r_0$ for all $\alpha \in A$. For $r_0 \leq r$ the existence of a $T \in C$ with $\int T dP_\theta = f(\theta)$ now follows from Theorem 1.

Remark 1 One can state exact analogues of Theorems 1 and 2 using the space $S_0(\mathcal{E})$ instead of $S(\mathcal{E})$.

For instance, select a set C as "standard" and let C_0^* be its restriction to $S_0(\mathcal{E})$. Let γ be a given function. Assume $(\text{OCC})_0$.

It will be possible to satisfy the equalities $\int T dP_s - \int T dP_t = \gamma(s) - \gamma(t)$ by a $T \in rC$ if and only if

$$\sum_{s,t} \alpha_{s,t}[\gamma(s) - \gamma(t)] \leq r$$

for all finite linear combinations such that

$$\sum_{s,t} \alpha_{s,t}(P_s - P_t) \in C_0^o.$$

Remark 2 To apply a result such as the above, one presumably needs to be able to verify whether a combination such as $\sum_{s,t} \alpha_{s,t}(P_s - P_t)$ belongs to C_0^o. Whether that is difficult or not will depend on how C_0^o can be characterized.

For the linear combinations note the following: Exchanging P_s and P_t if necessary, one can assume $\alpha_{s,t} \geq 0$. Then let $a = \sum_{s,t} \alpha_{s,t}$. The sum $\sum \alpha_{s,t}(P_s - P_t)$ can be written

$$\sum \alpha_{s,t}(P_s - P_t) = a(P_\mu - P_\nu)$$

where $P_\mu = \int P_\theta \mu(d\theta)$, $P_\nu = \int P_\theta \nu(d\theta)$ for probability measures μ and ν with finite support on Θ. The statement in Remark 1 takes then the form: There is a $T \in rC$ such that $\int T(dP_s - dP_t) = \gamma(s) - \gamma(t)$ for all pairs (s, t) of points of Θ if and only if whenever $a(P_\mu - P_\nu) \in C_0^o$ one has $a[\int \gamma(s)\mu(ds) - \int \gamma(t)\nu(dt)] \leq r$.

For a given pair (μ, ν) the problem of verifying whether $a(P_\mu - P_\nu)$ belongs to C_0^o is a one dimensional one. In terms of the given pair (P_μ, P_ν) it is at most a two dimensional problem. We shall return to that feature in Section 4 and 6.

In very many cases the sets C^* or C_0^* are symmetric around the origin. Then, working for instance on $S_0(\mathcal{E})$ one can define a "seminorm" ρ by taking for $\rho(m)$, $m \in S_0(\mathcal{E})$ the infimum of numbers a such that $m \in aC_0^o$. (It will be truly a semi-norm if C satisfies $(\text{OCC})_0^+$).

The system then looks as follows. One takes a signed measure λ with finite support and total mass zero on Θ. One puts $\langle \lambda, \gamma \rangle = \int \gamma d\lambda$ and $M_\lambda = \int P_\theta \lambda(d\theta)$. The desired relation then reads

$$|\langle \lambda, \gamma \rangle| \leq r\rho(M_\lambda).$$

3. Some sets that satisfy the condition (OCC).

Barankin [1949] discusses the situation where one has an experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$, a real valued function $\theta \rightsquigarrow \gamma(\theta)$ and another probability measure Q on the σ-field carrying the P_θ's. He wishes to find estimates T such that $\int T dP_\theta = \gamma(\theta)$ and such that $\int |T|^s dQ$ be as small as possible.

The probability measure Q used here need not be one of the P_θ's. (It is in Barankin's paper). It could be a $P_\mu = \int P_\theta \mu(d\theta)$ as in the end of Section 2 or it could be entirely unrelated.

If necessary for the notation, we shall enlarge Θ, adding to it a point Δ and let $\mathcal{E}^\Delta$ be the corresponding enlarged experiment with $P_\Delta = Q$. We shall also extend γ to Δ, putting $\gamma(\Delta) = 0$. (This may necessitate a recentering of γ).

Barankin imposes the following condition:

Condition BD. *All the P_θ are dominated by the measure Q. The number s satisfies $s > 1$ and for that chosen s, the densities $f_\theta = \frac{dP_\theta}{dQ}$ belong to the space $\mathcal{L}_t(Q)$ of functions such that $\int |f|^t dQ < \infty$ with $\frac{1}{s} + \frac{1}{t} = 1$.*

To apply the results of Section 1 one would take a quality set C that is a ball $C = \{T : \int |T|^s dQ \leq 1\}$.

Such a set is clearly convex for $s \geq 1$. It is symmetric around zero. For $s > 1$ it is also compact for the weak convergence on $\mathcal{L}_t(Q)$. Since (BD) requires that the densities f_θ belong to $\mathcal{L}_t(Q)$ our condition $(\mathrm{OCC})^+$ is satisfied.

In the next section, we shall discuss more particularly the case $s = 2$, where one wants to get estimates with prescribed expectation and minimum variance for Q.

One could wonder whether the condition (BD) is necessary, or whether it was imposed for convenience. We shall see, in Section 8 that when $\Theta \cup \Delta$ consist of exactly two points with $P_\theta << Q$ the minimum variance problem may not have a solution. Put $\gamma(\theta) = 1$ and $\gamma(\Delta) = 0$. then take for Q the Lebesgue measure on $[0, 1]$ and for P_θ the measure that has density $\frac{1}{2\sqrt{x}}$ with respect to Q. It will be shown in Section 8 that for T that have the prescribed expectations, $\inf_T \mathrm{Var}(T|Q) = 0$ but there is of course no T with variance zero and $\int T dP_\theta \neq \int T dQ$. The minimax variance problem mentioned in Section 2 and treated in Section 6 below introduces sets of the form $\{T = \int T^2 dP_\theta \leq m + \gamma^2(\theta), \text{ all } \theta \in \Theta\}$.

More generally consider sets of the form

$$C = \{T : \int T^2 dP_\theta \leq \rho(\theta)\}$$

where ρ is a function from Θ to $(0, \infty)$. Such a set is clearly convex and symmetric. To make sure it is not empty one may require $1 \leq \rho(\theta) < \infty$.

Let us show that, under some restrictions, it also satisfies the compactness condition of $(\mathrm{OCC})^+$.

To do so it is somewhat more convenient to introduce the set $\mathcal{M}$ of probability measures with finite support on Θ. For $\mu \in \mathcal{M}$ one writes $P_\mu = \int P_\theta \mu(d\theta)$. The condition $\int T^2 dP_\theta \leq \rho(\theta)$, all θ is then equivalent to $\int T^2 dP_\mu \leq \rho(\mu) = \int \rho(\theta)\mu(d\theta)$.

To proceed, let us introduce a linear space of measures $W = \cup_\mu \{W_\mu ; \mu \in \mathcal{M}\}$ where for each μ the space W_μ is the space of signed measures Q dominated by P_μ and having a density $\frac{dQ}{dP_\mu}$ that is bounded in absolute value. The space W is clearly a linear space. One could also consider the space $\bar{W}_\mu$ of finite signed measures Q dominated by P_μ and such that $\int |\frac{dQ}{dP_\mu}|^2 dP_\mu < \infty$. On W_μ or $\bar{W}_\mu$ let us use the square norm $\|Q\|_\mu^2 = \int |\frac{dQ}{dP_\mu}|^2 dP_\mu$.

Theorem 1 *Let φ be a linear functional defined on W and such that if $w \in W_\mu$ then $|\langle \varphi, w\rangle| \leq \sqrt{\rho(\mu)}\|w\|_\mu$.*

Then, if $\mathcal{E} = \{P_\theta : \theta \in \Theta\}$ is dominated by a finite measure, there exist a measurable function f such that $\langle \varphi, w\rangle = \int f dw$ and f satisfies $\int f^2 dw \leq \rho(\mu)\|w\|_\mu$ for each $w \in W_\mu$.

Proof According to Halmos and Savage (1949) there exist a sequence $\{\theta(j), c_j\}$ with $\theta(j) \in \Theta$ and $c_j > 0$ such that $\nu = \sum c_j P_{\theta(j)}$ is a finite measure that dominates all the ρ_θ.

(To see this easily, let M be a finite measure dominating all the P_θ. Let M_0 be its component in the band generated by the P_θ. Then M_0 still dominates all the P_θ, but being in their band, it is smaller than some convergent sum $\sum c_j P_{\theta(j)}$)

Now consider the finite measure $\lambda_n = \sum_{j=1}^n c_j \delta_{\theta(j)}$ on Θ and the corresponding integral $Q_n = \sum_{j=1}^n c_j P_{\theta(j)}$. It has the form $a_n P_{\mu_n}$ for $\mu_n = \frac{\lambda_n}{\|\lambda_n\|}$.

By assumption, one has $|\langle \varphi, w\rangle| \leq \sqrt{\rho(\mu_n)}\|w_n\|_{\mu_n}$ for that particular μ_n and for $w \in W_{\mu_n}$. Thus φ extends by continuity to the Hilbert space $\bar{W}_{\mu_n}$ and φ is representable there by an integral

$$\langle \varphi, w\rangle = \int f_n dw \quad \text{with} \quad \int f_n^2 dP_{\mu_n} \leq \rho(\mu_n).$$

One can decompose the band L of the P_θ into two complementary bands, writing $L = L_n + L_n^+$ where L_n is the band of P_{μ_n} and where the elements of L_n^+ are disjoint from those of L_n. Since the whole family is dominated, there is a set $A_n \in \mathcal{A}$ that carries all the measures in L_n and has measure zero for all those in $L_n^\perp$.

One does not change $\int f_n dw$, $w \in W_{\mu_n}$, by putting f_n equal to zero outside A_n.

Proceeding in this manner one gets a sequence $\{f_n : n = 1, 2, \ldots\}$ such that $f_{n+1} = f_n + u_n$ when u_n is zero on A_n and carried by $A_{n+1} \backslash A_n$. The f_n so constructed clearly tend to a limit f pointwise.

Note that f_n is determined up to equivalence on A_n. Thus f is defined up to equivalence for the dominating measure ν.

Let us show that if $\mu_0 \in \mathcal{M}$ is another probability measure, then $\int f^2 dP_{\mu_0} \leq \rho(\mu_0)$. To do so, perform the same construction, but with a sequence $\{\theta'(j); j = 1, 2\ldots\}$ that is the same as the previous one but with the θ's in the support of μ_0 put at its start.

This will give another function, say g instead of f. The two functions f and g differ only on a set of measure zero and $\int g^2 dP_{\mu_0} \leq \rho(\mu_0)$. Therefore $\int f^2 dP_{\mu_0} = \int g^2 dP_{\mu_0} \leq \rho(\mu_0)$.

This concludes the proof of the theorem.

An immediate corollary is as follows:

Lemma 1 *Consider the set C of functions satisfying $\int T^2 dP_\theta \leq \rho(\theta)$ for all $\theta \in \Theta$ for a given function ρ such that $1 \leq \rho(\theta) < \infty$. If $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ is dominated then C is compact for the pointwise convergence on W or on $\cup_\mu \{\bar{W}_\mu; \mu \in \mathcal{M}\}$.*

Proof Each such T defined a linear functional φ_T on W. That functional clearly satisfies $|\langle \varphi_T, W \rangle| \leq \sqrt{\rho(\mu)} \|W\|_\mu$ for $w \in W_\mu$ hence also for $w \in \bar{W}_\mu$. A limit pointwise on W is a linear functional on W, satisfying the same conditions. Hence it comes from some $T \in C$.

Remark Note that pointwise convergence on W implies pointwise convergence on $\cup_\mu \bar{W}_\mu$ of the extensions by continuity. This clearly also implies convergence of the integral $\int T dP_\mu$.

The domination by a finite measure used in Theorem 1 is not entirely necessary,

but some restriction appears to be. One condition that would work would be that $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ be $\sum$-finite. That means that there is a partition $\{A_j; j \in J\}$, usually of high cardinality, such that:

1) The restrictions of the P_θ; $\theta \in \Theta$ to each A_j are dominated.

2) If a function $f \geq 0$ has its restrictions to A_j measurable, it is already measurable.

3) The integrals $\int f dP_\theta$ are equal to $\sum_j \int_{A_j} f dP_\theta$.

It is stated in Le Cam (1986) that any experiment $\mathcal{E}$ has an equivalent version $\mathcal{F}$ that is $\sum$-finite. One obtains $\mathcal{F}$ by enlarging the σ-field $\mathcal{A}$ and the set that carries it.

For an example consider the set of pairs $\theta = (\alpha, \lambda)$. Let P_θ be the distribution of αX_λ where X_λ is an ordinary Poisson variable with $EX_\lambda = \lambda$.

4. Estimates with minimum variance at a point.

In this section we shall consider an experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ given by probability measures a σ-field $\mathcal{A}$ and another probability measure Q defined on that same σ-field.

The problem is then to find estimates with prescribed expectations $\gamma(\theta)$ for the P_θ but with minimum variance for Q.

The measure Q need not be one of the P_θ. It may be a finite convex sum $Q = \sum \alpha_\theta P_\theta$ or it may be entirely different from all such sums. For ease of notation we shall add to Θ an extra point Δ, let $P_s = Q$ and let $\mathcal{E}^\Delta$ be the experiment $\mathcal{E}^\Delta = \{P_\theta; \theta \in \Theta \cup \Delta\}$.

Since the estimates T under consideration are supposed to have small variance at Q, they will also have an expectation $\int TdQ$. We shall set $\int TdQ = 0$, replacing if necessary the prescribed $\gamma(\theta) = \int TdP_\theta$ by $\int TdP_\theta - \int TdQ$.

The problem will be considered only under the following restriction.

(BD) *The P_θ are all dominated by Q and the densities $f_\theta = dP_\theta/dQ$ are such that $\int f_\theta^2 dQ < \infty$.*

This is one of the problems treated by Barankin in [1949] (except that he puts $Q = P_{\theta_0}$ for some $\theta_0 \in \Theta$). The solution is given by Theorem 1, Section 2. However, here, one can proceed more directly to an explicit solution.

We shall work in the Hilbert space $L_2(Q)$ with square norm $\|f\|^2 = \int f^2 dQ$. However, we shall make no notational difference between a function and its equivalence class.

Let $\mathcal{U}$ be the class of functions T such that $\int T^2 dQ < \infty$ and $\int TdP_\theta = \gamma(\theta)$, $\theta \in \Theta \cup \Delta$. Let $\mathcal{U}_0$ be the class of functions u such that $\int U^2 dQ < \infty$ and $\int udP_\theta = 0$ for $\theta \in \Theta \cup \Delta$. These are the unbiased estimates of zero.

The sets $\mathcal{U}$ and $\mathcal{U}_0$ will be treated as subsets of $L_2(Q)$.

Lemma 1 (Lehmann-Scheffé, 1950). *Assume that $\mathcal{U}$ is not empty. Then a $T \in \mathcal{U}$ minimizes $\int T^2 dQ$ if and only if it is orthogonal in $L_2(Q)$ to all unbiased estimates of zero.*

Proof Indeed

$$\int (T + \alpha u)^2 dQ = \int T^2 dQ + \alpha^2 \int u^2 dQ + 2\alpha \int uTdQ.$$

If $\int uTdQ$ is not zero one can select an α small of appropriate sign to make this smaller than $\int T^2 dQ$. Hence the result. To say that $u \in \mathcal{U}_0$ is to say that for $f_\theta = \frac{dP_\theta}{dQ}$ one has $\int u f_\theta dQ = 0$ for all $\theta \in \Theta$ and for $\theta = \Delta$ as well.

Let $g_\theta = f_\theta - 1$. This is an element of the subspace $L_{2,0}$ of functions such that $\int g dQ = 0$. In $L_{2,0}$ the functions g_θ span a certain linear space $\mathcal{G}$ and $\mathcal{U}_0$ is the orthogonal complement of $\mathcal{G}$. Thus, the orthogonal complement of $\mathcal{U}_0$ is the space $\mathcal{H} \subset L_{2,0}$, closure of $\mathcal{G}$.

This leads to the following lemma

Lemma 2 *Let J be a finite subset of Θ. If there is a $T \in \mathcal{H}$ that satisfies $\int T dP_\theta = \gamma(\theta)$, $\theta \in \Theta$, the orthogonal projection T_j of T on the span of $\{g_j; j \in J\}$ satisfies $\int T_j dP_j = \gamma(j)$, $j \in J$ and T is the limit in $\mathcal{H}$ of the T_j as J increases along the directed finite subsets of θ.*

This is clear. Barankin [1949] observes that one needs only to find approximate solutions that became more and more refined as J increases. For simplicity we shall use only exact solutions.

Note that for a given J the g_j, $j \in J$ might not be linearly independent in $L_{2,0}$. Then if $\sum c_j g_j = 0$ one must have $\sum c_j \gamma(j) = 0$. When that occurs we need only to look at subsets J such that the g_j are linearly independent. Now fix a J such that the g_j are linearly independent, and, for ease in notation, drop the subscripts J whenever they would occur. Thus we shall look for a $T \in$ span of g_j; $j \in J$ such that $\int T dP_j = \gamma(j)$, $j \in J$.

Since T is in the span of the g_j it has the form $\sum c_j g_j$ for real numbers c_j, $j \in J$. The equations $\int T dP_k = \gamma(k)$ become

$$\int \sum c_j g_j g_k dQ = \gamma(k)$$

and

$$\text{var}_Q T = \sum_j \sum_k c_j c_k \int g_j g_k dQ.$$

This suggest writing $\{c_j; j = J\}$ as a column vector c and similarly for $\{\gamma(j), j \in J\}$. Introduce also the matrix B whose entries are

$$B_{j,k} = \int g_j g_k dQ; j, k \in J,$$

and the column vector g formed by the g_j, $j \in J$.

Lemma 3 *For the finite set J with g_j, $j \in J$ linearly independent, the minimum variance T_J has the form $c'g$ for $Bc = \gamma$. Its variance under Q is*

$$c'Bc = \gamma' B^{-1} \gamma$$

Note that B is positive definite nonsingular since the g_j are linearly independent in $L_{2,0}$. The variance $\gamma' B^{-1}\gamma$ can also be written in a variety of equivalent forms. For instance

$$\begin{aligned} \gamma' B^{-1}\gamma &= \sup_c \{|c'\gamma|^2;\ c'Bc \leq 1\} \\ &= \sup_{c\neq 0} \frac{|c'\gamma|^2}{c'Bc} = \sup_{c \neq 0} \frac{|c'\gamma|^2}{\|c'g\|^2}. \end{aligned}$$

This corresponds to the polar inequalities of Theorem 1, Section 2.

Still a further form can be obtained as follows. In a ratio such as $\frac{|c'\gamma|^2}{\|c'g\|^2}$ one can always assume that $\sum |c_j| = 1$. Then note that

$$\sum c_j (P_j - Q) = \sum c_j^+ p_j + (\sum c_j^-) Q - (\sum c_j^+ Q + \sum c_j^- P_j)$$

can be put in the form $\sum c_j(P_j - Q) = P_\mu - P_\nu$ where $P_\mu = \int P_s \mu(ds)$ for a probability measure μ with finite support on $\Theta \cup \Delta$ and similarly for ν. Then $\|c'g\|^2 = \int \frac{(dP_\mu - dP_\nu)^2}{dQ}$. Therefore

$$\frac{|c'\gamma|^2}{\|c'g\|^2} = \frac{|\langle \nu, \mu\rangle - \langle \gamma, \nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dQ}}.$$

The inequality

$$\text{var}_Q T \geq \sup_{\mu,\nu} \frac{|\langle \gamma, \mu\rangle - \langle , \nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dQ}}$$

is exactly the inequality given by Kiefer [1952], except that here we use only measures μ and ν with finite support on $\Theta \cup \Delta$.

We have the further result that, under BD, there exist a T for which the inequality is an exact equality. (This was noted by Kiefer [1952] by reference to Barankin's paper).

Note further that the ratio $R(\mu,\nu) = |\langle \nu, \mu\rangle - \langle \gamma, \nu\rangle|^2 \ \{\int \frac{(dP_\mu - dP_\nu)^2}{dQ}\}^{-1}$ is also obtainable as follows:

One seeks estimates T such that $\int T(dP_\mu - dP_\nu) = \langle \nu, \mu\rangle - \langle \gamma, \nu\rangle$. The ratio $R(\mu,\nu)$ is the minimum possible variance of such estimates.

Note especially that in this problem T is restricted to satisfy $\int T(dP_\mu - dP_\nu) = \langle \gamma, \mu - \nu \rangle$ and not the more stringent restriction that $\int T dP_\theta = \gamma(\theta)$ for all θ.

There is an adventitious feature in our creation of the pairs (μ, ν) to get $\|c'g\|^2 = \int \frac{(dP_\mu - dP_\nu)^2}{dQ}$. It is the circumstance that $\mu + \nu$ gives mass at least unity to the point Δ. This is not a serious feature, but it leads to the following remarks.

Take a particular pair (μ, ν) of probability measures with finite support on $\Theta \cup \Delta$. Let δ be the Dirac mass at Δ. For $\epsilon \in (0, 1]$, let $\mu(\epsilon) = \epsilon\mu + (1 - \epsilon)\delta$ and $\gamma(\epsilon) = \epsilon\nu + (1 - \epsilon)\delta$. For any triplet (λ, μ, ν) of probability measures with finite support on $\Theta \cup \Delta$, let

$$R(\lambda, \mu, \nu) = \frac{|\langle \gamma, \mu \rangle - \langle \gamma, \nu \rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}}.$$

Lemma 4 *With the notation just described one has*

$$\begin{aligned} R(\delta, \mu, \nu) &= R[\delta; \mu(\epsilon), \nu(\epsilon)] \\ &\leq \sup_{\mu', \nu'} R[\frac{1}{2}(\mu' + \nu'); \mu', \nu'] \end{aligned}$$

Proof One has $\int \frac{[dP_{\mu(\epsilon)} - dP_{\nu(\epsilon)}]^2}{dQ} = \epsilon^2 \int \frac{[dP_\mu - dP_\nu]^2}{dQ}$ and a similar relation for $|\langle \gamma, \mu(\epsilon) \rangle - \langle \gamma, \nu(\epsilon) \rangle|^2$. Hence the first equality.

For the second, note that

$$\frac{1}{2}[\mu(\epsilon) + \nu(\epsilon)] = \epsilon \frac{1}{2}(\mu + \nu) + (1 - \epsilon)\delta.$$

Thus, writing $\lambda = \frac{1}{2}(\mu + \nu)$ and $\lambda(\epsilon) = \frac{1}{2}[\mu(\epsilon) + \nu(\epsilon)]$, one has

$$P_{\lambda(\epsilon)} = \epsilon P_\lambda + (1 - \epsilon)Q.$$

Hence $P_{\lambda(\epsilon)} \geq (1 - \epsilon)Q$ and therefore

$$\frac{1}{1 - \epsilon} \int \frac{[dP_{\mu(\epsilon)} - dP_{\nu(\epsilon)}]^2}{dQ} \geq \int \frac{[dP_{\mu(\epsilon)} - dP_{\nu(\epsilon)}]^2}{dP_{\lambda(\epsilon)}}.$$

This gives the inequality as stated.

These formulas, and Lemma 3, were obtained under the assumption that the g_j, $j \in J$ were linearly independent. This is necessary if one wants to avoid dividing by zero or writing inverses for matrices that are singular, but it is otherwise immaterial.

For instance, in Section 7, we shall work with a finite set P_j; $j \in J$ and with measures Q of the form $Q = \sum \beta_j P_j$, $\beta_j \geq 0$, $\sum \beta_j = 1$. Then the set $\{P_j; j \in J\} \cup \{Q\}$ is certainly not a linearly independent set. However, the corresponding matrix B satisfies $\beta' B = B\beta = 0$. Thus if one works in the space of vectors orthogonal to β this will not change

$$\sup_c \{|c'\gamma|^2; c'Bc \leq 1\}$$

as long as one keeps the condition $\int T dQ = 0$, which becomes $\beta'\gamma = 0$.

The foregoing equalities and inequalities have the following consequence: *The general problem has been reduced to the solution of a family of two-dimensional problems.* In the following sense. Consider two probability measures (with finite support) on $\Theta \cup \Delta$. Let them be μ and ν and let $w = \mu - \nu$. Consider the problem of finding estimates T such that $\int T dQ = 0$, $\int \int T dP_\theta w(d\theta) = \int \gamma(s) w(ds)$ and $\int T^2 dQ$ be minimized.

Let $V(w)$ be the minimum possible variance. Then the minimum variance for the initial problem is simply $\sup_w V(w)$.

As mentioned earlier Barankin, in his 1949 paper, showed that the bounds so obtained included all the various existing versions of the Cramér-Rao type inequalities. They also give the Chapman-Robbins (1951) and the Kiefer (1952) results. Let us show that the bounds given here also include the Fabian-Hannan (1977) inequality.

These authors consider a case where Q is one of the P_θ's, say P_{θ_0}. They assume that a certain sequence, say $\{g_n\}$, of differences $g_n = \frac{dP_{\theta_n}}{dQ} - 1$ is such that for a certain sequence $\{c_n\}$ of real numbers, the products $c_n g_n$ converge weakly in $L_2(Q)$ to a limit h. The bound is then

$$\text{var}_Q T \geq \frac{1}{\|h\|^2} \lim_n \{c_n[\gamma(\theta_n) - \gamma(\theta_0)]\}^2.$$

Note that if $c_n g_n$ converges weakly to h then, for $T \in L_2(Q)$ one has

$$\lim_n c_n \int g_n T dQ = \int h T dQ = \lim_n c_n[\gamma(\theta_n) - \gamma(\theta_0)].$$

Thus the limit in the bound does exist.

Actually, Fabian and Hannan use a filter instead of a sequence, but it is a filter with countable base and nothing needs to be changed in the arguments.

It can be assumed without loss of generality that $\gamma(\theta_0) = 0$. We shall do so.

The weak convergence of $c_n g_n$ implies that there are numbers $b < \infty$ and N such that $n > N$ implies $\|c_n g_n\|^2 \leq b^2$. Then there will be finite convex combination, say $\sum_{k>N} \alpha_{n,k} c_k g_k$ such that $\|\sum \alpha_{n,k} c_k g_k\| \leq b$ and such that $\sum \alpha_{n,k} c_k g_k$ converges in the sense of the norm to k. The limit of $\sum \alpha_{n,k} c_k \gamma(\theta_k)$ is the same as that of $\gamma(\theta_k)$.

Note then that the $\sum \alpha_{n,k} c_k g_k$ are particular cases of the sums used in Lemma 3. Therefore

$$\lim_n \frac{|c_n \gamma(\theta_n)|^2}{\|h\|^2} \leq \sup_c \frac{|c'\gamma|^2}{\|c'g\|^2},$$

and the Fabian-Hannan bound is not larger than Barankin's.

5. A related problem.

Instead of trying to find estimates such that $\int T dP_\theta = \gamma(\theta)$, with small variance, one could just try to estimate γ with small expected square deviations. For instance one could use a prior π on Θ and try to minimize $\int E_\theta[T-\gamma(\theta)]^2 \pi(d\theta)$.

For practical purposes this may be a better idea than to try to enforce the equalities $\int T dP_\theta = \gamma(\theta)$, even though as shown in Section 2, one can give some leeway to the equalities without too much additional effort.

In this section we shall consider the problem of minimizing $\int E_\theta[T-\gamma(\theta)]^2 \pi(d\theta)$ for a probability measure π with finite support on Θ. We shall see that the solution involves the matrices B with entries $B_{j,k} = \int \frac{dP_j dP_k}{dQ} - 1$ as in Section 4, but with $Q = P_\pi = \int P_\theta \pi(d\theta)$.

We also show that if these matrices are known for enough measures $\pi \in \mathcal{M}$, then the type of the experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ is determined.

We shall restrict ourselves to measures π with finite support to avoid stating measurability conditions. If such conditions are imposed, the result extends to arbitrary probability measures on Θ. We shall give two derivations of the same formula.

Lemma 1 *Let π be a probability measure with finite support on Θ. Then*

$$\inf_T \int E_\theta (T-\theta)^2 \pi(d\theta) = \frac{1}{2} \int\int [\gamma(s)-\gamma(t)]^2 A(s,t) \pi(ds)\pi(dt)$$

with $A(s,t) = \int \frac{dP_s dP_t}{dQ}$, $Q = \int P_\theta \pi(d\theta)$.

Proof For the Bayesian problem with prior π the optimum T is the posterior expectation of γ given the observations. The posterior risk is the variance of that posterior distribution. This variance is also $\frac{1}{2} E|\gamma(s)-\gamma(t)|^2$ for independent versions s and t with common distribution the posterior of θ. The joint distribution of the observation x and θ may be written $P_\theta(dx)\pi(d\theta) = F_x(d\theta)Q(dx)$. so we have to take

$$\frac{1}{2}\int\int\int |\gamma(s)-\gamma(t)|^2 F_x(ds) F_x(dt) Q(dx)$$

$$= \frac{1}{2}\int\int |\gamma(s)-\gamma(t)|^2 \frac{P_s(dx)P_t(dx)}{Q(dx)} \pi(ds)\pi(dt).$$

Hence the result.

One can argue differently as follows. Let $\mathcal{A}$ be the σ-field of the observations but take the joint distribution of x and θ under $P_\theta(dx)\pi(d\theta)$.

Then

$$E\gamma^2(\theta) = EE[\gamma(\theta) - E(\gamma(\theta)|\mathcal{A})]^2 + E\{E[\gamma(\theta)|\mathcal{A}]\}^2.$$

Thus

$$EE[\gamma(\theta) - E[\gamma(\theta)|\mathcal{A}]]^2 = E\gamma^2 - E\{E[\gamma(\theta)|\mathcal{A}]]^2.$$

However $E[\gamma(\theta)|\mathcal{A}] = \int \gamma(\theta)F_x(d\theta)$. Thus

$$E[\gamma(\theta)|\mathcal{A}]^2 = \int\int \gamma(s)\gamma(t)F_x(ds)F_x(dt)$$

and $E\{E[\gamma(\theta)|\mathcal{A}]\}^2 = \int\int\int \gamma(s)\gamma(t)\frac{P_s(dx)P_s(dx)}{Q(dx)}\pi(ds)\pi(dt)$. This gives

$$\inf_T \int E_\theta[T - \gamma(\theta)]^2\pi(d\theta) = \int \gamma^2(\theta)\pi(d\theta) - \int\int \gamma(s)\gamma(t)A(s,t)\pi(ds)\pi(dt).$$

This is obviously equivalent to the formula given in Lemma 1. Let us transform it to make the matrices $B(s,t) = A(s,t) - 1$ appear.

Lemma 2 *The quantity* $\inf_T \int E_\theta(T-\theta)^2\pi(d\theta)$ *is equal to*

$$\mathrm{var}_\pi\gamma - \int\int \gamma(s)\gamma(t)B(s,t)\pi(ds)\pi(dt).$$

(Here $\mathrm{var}_\pi\gamma$ *is* $\int\gamma^2(\theta)\pi(d\theta) - [\int\gamma(\theta)\pi(d\theta)]^2$*. It is the variance of* $\gamma(\theta)$ *for the prior measure).*

Proof Since $A(s,t) = B(s,t) + 1$, one has

$$\int\int \gamma(s)\gamma(t)A(s,t)\pi(ds)\pi(dt) = \int\int \gamma(s)\gamma(t)\pi(ds)\pi(dt) + \int\int \gamma(s)\gamma(t)B(s,t)\pi(ds)\pi(dt).$$

The result follows.

Note that in this form, it does not matter if one replaces $\gamma(\theta)$ by $\gamma(\theta) - \bar{\gamma}$, with $\bar{\gamma} = \int \gamma(\theta)\pi(d\theta)$ since $\int B(s,t)\pi(dt) = \int B(s,t)\pi(ds) = 0$.

Note also that a term of the type $\int\int \gamma(s)\gamma(t)A(s,t)\pi(ds)\pi(dt)$ can be put in the alternate form $\int \frac{(dV)^2}{dQ}$ where V is a certain signed measure $V = aP_\mu - bP_\nu$, just as in Section 4.

The occurence in this section of terms of the type $A(s,t) = \int \frac{dP_s dP_t}{dQ}$ with $Q = \int P_s\pi(ds)$ and the fact that if one wants to find $\inf_T \sup_\theta E_\theta[T - \gamma(\theta)]^2$ one will have to consider several averages $\int P_s\pi(ds)$ for variable π, suggests inquiring whether the $A(s,t)$ can be computed in terms of a few simple characteristics of the experiment $\mathcal{E}$. We shall now show that to compute them one needs to know the type of the experiment $\mathcal{E}$. Specifically, the following result holds.

Proposition 1 *Let Θ be a finite set. For a probability measure π on Θ let $A(s,t;\pi) = \int \frac{dP_s dP_t}{dP_\pi}$ where $P_\pi = \int P_\theta \pi(d\theta)$ as usual.*

If the matrices $A(s,t;\pi)$ are known for all π in the neighborhood of an interior point of the simplex of probability measures on Θ, then the type of $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ is well determined.

Proof It should be clear that the matrices $A(s,t;\pi)$ depend only on the *type* of the experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$. This is usually represented by the Blackwell canonical measure in which one takes for π the uniform distribution on Θ, but one can use just as well any π that charges all the $\theta \in \Theta$.

In this situation we shall write π_θ for the mass given to θ by π. Similarly for other measures. The Blackwell representative under π uses the densities $f_\theta = \frac{dP_\theta}{dQ}$, $Q = \sum \pi_\theta P_\theta$. The vector $f = \{f_\theta; \theta \in \Theta\}$ of such densities sends Q into a probability measure on the simplex $S = \{u : u \in I\!R^\Theta,\ u_\theta \geq 0,\ \sum \pi_\theta u_\theta = 1\}$.

To characterize the type of $\mathcal{E}$ it is enough to give the distribution of f under Q.

Now take another finite positive measure $\beta = \{\beta_\theta; \theta \in \Theta\}$ on Θ. If it is not a probability measure, write $t = \sum \beta_\theta$ and $\mu_\theta = \frac{\beta_\theta}{t}$. This is now a probability measure (unless $t = 0$).

Consider the values

$$A(s,\theta;\mu) = \int \frac{dP_s dP_\theta}{dM_\mu}, \text{ with } M_\mu = \sum \mu_\theta P_\theta.$$

Assume that all the β_θ are strictly positive. Then one can write

$$A(s,\theta;\mu) = \int \frac{dP_s dP_\theta}{dQ} \frac{dQ}{dM_\mu}$$

and

$$\begin{aligned} \sum\sum \pi_s \pi_\theta A(s,\theta;\mu) &= \int \frac{dQ}{dM_\mu} \frac{\sum \pi_s \pi_\theta dP_s dP_\theta}{dQ} \\ &= \int \frac{dQ}{dM_\mu} dQ. \end{aligned}$$

Now if $\int \frac{dQ}{dM_\mu} dQ$ is known, so is $\int \frac{dQ}{dM_\beta} dQ$ for $\beta = t\mu$. Thus one can assume $\int \frac{dQ}{dM_\beta} dQ$ known for β in a neighborhood of $\pi = \{\pi_\theta; \theta \in \Theta\}$ in the space $\mathbb{R}^\Theta$ itself. Note that $\frac{dM_\beta}{dQ} = \sum \beta_\theta f_\theta$, with $f_\theta = \frac{dP_\theta}{dQ}$. Thus $\int \left(\frac{1}{\sum \beta_\theta f_\theta}\right) dQ$ is given for β in a neighborhood of π in $\mathbb{R}^\Theta$.

This is sufficient to determine the distribution under Q of the vector $f = \{f_\theta; \theta \in \Theta\}$. One can see this by writing

$$\int \left(\frac{1}{\sum \beta_\theta f_\theta}\right) dQ = \int_0^\infty (\int e^{-(\sum \beta_\theta f_\theta)x} dQ) dx.$$

Alternately one can take successive derivatives with respect to the β_θ and evaluate them at the point π.

For instance, if $D(\beta) = \int \left(\frac{1}{\sum \beta_\theta f_\theta}\right) dQ$ then

$$\frac{\partial D(\beta)}{\partial \beta_j} = -\int \frac{f_i}{(\sum \beta_\theta f_\theta)^2} dQ,$$

$$\frac{\partial^2 D(\beta)}{\partial \beta_j \partial \beta_k} = 2\int \frac{f_j f_k}{(\sum \beta_\theta f_\theta)^3} dQ,$$

and so forth.

Since $\sum \pi_\theta f_\theta = 1$, when one evaluate such derivatives at π one gets all the mixte moments of $f = \{f_\theta; \theta \in \Theta\}$. These determine the distribution of f for Q since the distribution is carried by a compact subset of $\mathbb{R}^\Theta$.

This gives the result as stated.

The preceding Proposition is a bit of an enthusiasm damper. To solve the minimax problem $\inf_T \sup_\theta E_\theta (T - \theta)^2$ one will have to look at many prior measures such as π or μ. The Proposition says that there is no way to compute exactly these minimax risks, perhaps for several functions γ, unless one knows the type of the experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$.

Note that it does not say that one cannot compute *approximations* to the minimax risks. To obtain those it will be sufficient to have approximations to the type of $\mathcal{E}$.

The above discussion has been given in terms of a real valued function γ. However nothing is changed if it is a function to some Hilbert space with norm $\|\cdot\|$. One just changes $|\gamma(s) - \gamma(t)|^2$ to $\|\gamma(s) - \gamma(t)\|^2$ and $\gamma(s)\gamma(t)$ to the inner product $\langle \gamma(s), \gamma(t) \rangle$.

6. Minimax variance for prescribed expectations.

In this section we consider an experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ given by probability measures P_θ on a σ-field $\mathcal{A}$. There is also a set $\mathcal{D}$ of real valued functions defined on Θ.

The problem is to find estimates T such that, for some $\gamma \in \mathcal{D}$, one has $\int T dP_\theta = \gamma(\theta)$ but in such a way that $\sup_\theta \text{Var}_\theta T$ be as small as possible.

The main result of this section is that the problem is solvable by taking limits of solutions of finite dimensional problems, each reducible to a sequence of problems of the type considered in Section 4.

We shall use throughout the following assumptions:

(A1) *The experiment $\mathcal{E}$ is dominated by a σ-finite measure.*

(A2) *The set $\mathcal{D}$ is a non-empty convex set that is compact for pointwise convergence on Θ.*

Let $\mathcal{F}$ be the class of finite subsets of Θ directed upward by inclusion.

For each $F \in \mathcal{F}$ consider the problem of finding estimates T_F such that the function $s \rightsquigarrow \int T_F dP_s$, $s \in F$ is the restriction to F of some $\gamma \in \mathcal{D}$. The finite dimensional problem is to evaluate

$$b(F) = \inf_{T_F} \sup_{s \in F} \text{Var}_s T_F$$

where the "inf" is taken over estimates T_F such that $s \rightsquigarrow \int T_F dP_s$ belongs to the set $\mathcal{D}_F$, restriction of $\mathcal{D}$ to F.

Note that this problem is a finite dimensional one in the sense that it involves only the P_s for $s \in F$ and the restrictions of the functions in $\mathcal{D}$ to F.

One first result is as follows.

Theorem 1 *Let assumptions* (A1) *and* (A2) *be satisfied and let m be the minimax value*

$$m = \inf_T \sup_{\theta \in \Theta} \text{Var}_\theta T$$

where T runs over the set of estimates such that $\int T^2 dP_\theta < \infty$ and such that $\theta \rightsquigarrow \int T dP_\theta$ belongs to $\mathcal{D}$.

Then

$$m = \sup_{F \in \mathcal{F}} b(F) = \lim_F b(F)$$

Proof Consider first the case where $b = \sup_F b(F)$ is finite. One can certainly assert that $b \leq m$.

Let ρ be a real valued function defined on Θ and such that $\rho(\theta) \geq 1 + b + \sup_\gamma \{\gamma^2(\theta), \gamma \in \mathcal{D}\}$. Since $b < \infty$ and since $\mathcal{D}$ is pointwise compact one sees that $1 \leq \rho(\theta) < \infty$.

Now fix a finite set $F \in \mathcal{F}$. There will exist pairs (T_F, γ_F) such that $\gamma_F \in \mathcal{D}_F$ and such that $\int T_F dP_s = \gamma_F(s)$, $s \in F$ with $\mathrm{Var}_s T_F \leq b$, $s \in F$.

Consider also the spaces $W_{\mu(F)}$ of Section 3, Theorem 1 for $F \in \mathcal{F}$ and $\mu(F)$ uniform on F. It is easily seen that $W = \cup_F \{W_{\mu(F)}; F \in \mathcal{F}\} = \cup_\mu (W_\mu; \mu \in \mathcal{M})$.

The estimate T_F defines a linear functional ψ_F on $W_{\mu(F)}$. It can be extended to the whole space W giving there some linear functional φ_F. Along an ultrafilter $\mathcal{V}$ finer than the filter of tails of the directed set $\mathcal{F}$ the pairs (φ_F, γ_F) will have a limit where the second coordinate $\gamma = \lim_{\mathcal{V}} \gamma_F$ belongs to $\mathcal{D}$.

If F_1 is any finite subset of Θ, as soon as $F \in \mathcal{F}$ satisfies $F_1 \subset F$, the linear functional φ_F will be such that

$$|\langle \varphi_F, w \rangle| \leq \sqrt{\rho[\mu(F_1)]} \|w\|_{\mu(F_1)}$$

for the norms used in the proof of Theorem 1 Section 3. Therefore the limit $\lim_{\mathcal{V}} \langle \varphi_F, w \rangle$ will satisfy the same inequality for all $w \in W_{\mu(F_1)}$. As a consequence $\lim_{\mathcal{V}} \varphi_F$ defines a linear functional φ on W. It is such that

$$|\langle \varphi, w \rangle|^2 \leq \rho[\mu(F)] \|w\|_{\mu(F)} \text{ for } F \in \mathcal{F}.$$

By Theorem 1, Section 3 it arises from an integral $\langle \varphi, w \rangle = \int T dw$ where

$$\int T^2 dw \leq \rho[\mu(F)] \|w\|^2_{\mu(F)}, \text{ all } F \in \mathcal{F}.$$

The functions γ_F converge to γ and, since the topology used in Theorem 1, Section 3, implies convergence of integrals one will have $\gamma(\theta) = \int T dP_\theta$ for all $\theta \in \Theta$.

Since $\mathrm{Var}_\theta T_F \leq b$ for $\theta \in F$, one concludes, by lower semicontinuity, that $\mathrm{Var}_\theta T \leq b$ for all $\theta \in \Theta$.

This concludes the proof since $b = \infty$ would imply $m = \infty$.

One can also obtain a result that involves prior distributions. Here again we shall limit ourselves to priors λ that belong to $\mathcal{M}$.

Theorem 2 *Let assumptions* (A1) *and* (A2) *be satisfied. Then the minimax value m of Theorem 1 is equal to*

$$m = \sup_{\lambda \in \mathcal{M}} \inf_{T_F} \int (\mathrm{var}_\theta T_F) \lambda(d\theta)$$

where F is the support of λ and T_F runs through estimates such that the function $s \rightsquigarrow \int T_F dP_s$ defined on F belongs to $\mathcal{D}_F$.

Proof According to Theorem 1, it is sufficient to prove the result for finite subsets of Θ. That is, for a given finite set F, we need to prove that

$$b(F) = \sup_{\lambda \in \mathcal{M}_F} \inf_{T_F} \int (\mathrm{Var}_\theta T_F) \lambda(d\theta).$$

Thus, fix an F and assume first that $b(F)$ is finite. Take on F a function ρ_F such that for $s \in F$ one has $1 \leq \rho_F(s) < \infty$ and

$$\rho_F(s) \geq [1 + b(F)] + \sup_{\gamma \in \mathcal{D}_F} \gamma^2(s).$$

Any T_F that achieves the minimum or close to it on F will be in the set

$$C_F = \{T_F; \int T^2 dP_s \leq \rho_F(s), s \in F\}$$

Consider the set $\mathcal{R}_0$ of functions defined on F by $s \rightsquigarrow r(s) = \mathrm{Var}_s T$ for $T \in C_F(\mathcal{D})$ where $C_F(\mathcal{D})$ is the set of $T \in C_F$ whose expectation $s \rightsquigarrow \int T dP_s$ belong to $\mathcal{D}_F$.

The variances satisfy the inequalities that for $\alpha \in (0, 1)$ and $T_3 = \alpha T_1 + (1 - \alpha) T_2$ one has $\mathrm{Var}_s T_3 \leq \alpha \mathrm{Var}_s T_1 + (1 - \alpha) \mathrm{Var}_s T_2$. Thus introduce the set $\mathcal{R}$ of functions of the form $\varphi = \psi + u$, $\psi \in \mathcal{R}_0$, $u \geq 0$. According to the above the set $\mathcal{R}$ has the following properties:

1) The set $\mathcal{R}$ is convex

2) The set $\mathcal{R}$ is closed for pointwise convergence on F

3) For $r \in \mathcal{R}$ one has $r(s) \geq 0$, $s \in F$.

The second properties follows from the compactness of C_F.

It follows from a minimax theorem (for which see for instance Le Cam (1964) that

$$\inf_{\varphi \in \mathcal{R}} \sup_{s \in F} \varphi(s) = \sup_{\lambda \in \mathcal{M}_F} \inf_{\varphi \in \mathcal{R}} \int \varphi d\lambda.$$

Furthermore there exists a φ_0 such that

$$\sup_\lambda \int \varphi_0 d\lambda = \sup_\lambda \inf_\varphi \int \varphi d\mu.$$

This gives the desired result if $b(F) < \infty$. If on the contrary $b(F) = \infty$, take any $a \in (0, \infty)$.

Replace the previous C_F by

$$C_{F,a} = \{T; \int T^2 dP_s \leq (1 + a) + \sup_\gamma [\gamma^2(s); \gamma \in \mathcal{D}_F\}.$$

Let $C_{F,a}(\mathcal{D})$ be the subset of $C_{F,a}$ formed by those T's whose expectations $s \rightsquigarrow TdP_s$, $s \in F$ belong to $\mathcal{D}_F$. If $C_{F,a}(\mathcal{D})$ is not empty the previous argument would apply, giving

$$\inf_T \{\sup_{s\in F} \mathrm{Var}_s T, T \in C_{F,a}(\mathcal{D}\}$$

$$= \sup_{\lambda\in\mathcal{M}_F} \inf_T \{\int (\mathrm{Var}_s T)\lambda(ds), T \in C_{F,a}(\mathcal{D})\}.$$

However this value would then be finite, contrary to the assumption that $b(F) = \infty$.

The result follows as stated.

So far we have taken for $\mathcal{D}$ any class of functions that satisfies condition (A2). To describe further implications of Theorem 2, let us look *at the case where* $\mathcal{D}$ *consists of a single function* γ.

That is, for a given finite set F*, one requires* $\int T_F dP_s = \gamma(s)$*,* $s \in F$*.*

In such a case the function $s \rightsquigarrow \gamma^2(s)$ is also known. If $\lambda \in \mathcal{M}_F$ and $P_\lambda = \int P_s \lambda(ds)$ one can write

$$\int T^2 dP_\lambda = \int (\mathrm{Var}_s T)\lambda(ds) + \int (\int TdP_s)^2 \lambda(ds).$$

Hence

$$\mathrm{Var}(T|P_\lambda) = \int (\mathrm{Var}_s T)\lambda(ds) + \int (\int TdP_s)^2 \lambda(ds) - [\int TdP_s \lambda(ds)]^2$$
$$= \int (\mathrm{Var}_s T)\lambda(ds) + \int \gamma^2(s)\lambda(ds) - (\int \gamma(s)\lambda ds)^2.$$

The means that, for given γ, the problem of minimizing $\int (\mathrm{Var}_s T)\lambda(ds)$ is equivalent to the problem of minimizing the variance of T under P_λ. This is a problem that was treated at length in Section 4.

The solution given there suggests the following. Let F be the exact support of λ. The Barankin conditions $\int \left(\frac{dP_s}{dP_\lambda}\right)^2 dP_\lambda < \infty$ are satisfied and for estimates T such that $\int TdP_s = \gamma(s)$, $s \in F$, the minimum variance under P_λ is given by

$$\inf_T \mathrm{Var}(T|P_\lambda) = \sup_{\mu,\nu} \frac{|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}}$$

where μ and ν are arbitrary elements of $\mathcal{M}_F$, such that $P_\mu \neq P_\nu$.

This yields

$$\inf_T \int (\mathrm{Var}_s T)\lambda(ds) = \sup_{\mu,\nu} \frac{|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}} - \{\int \gamma^2(s)\lambda(ds) - [\int (s)\lambda(d)]^2\}$$

As already noted in Section 4, the quantity $R(\lambda,\mu,\nu) = \frac{|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}}$ is the minimum possible variance at P_λ of an estimate T such that $\int T dP_\mu - \int T dP_\nu = \langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle$.

Also, for any T such that $\int T dP_\theta = \gamma(\theta)$ for all θ one has

$$\mathrm{Var}(T|P_\lambda) \geq \sup_{\mu,\nu} \frac{|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}}.$$

In addition, for every $\epsilon \in (0,1)$ and triplet (λ,μ,ν) there exist measures μ' and ν' such that

$$\frac{1}{2}(\mu' + \nu') = \epsilon\frac{(\mu+\nu)}{2} + (1-\epsilon)\lambda$$

and such that $R(\lambda,\mu',\nu') = R(\lambda,\mu,\nu)$. This gives, for fixed λ,

$$\sup_{\mu,\nu} \frac{|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}} \leq \frac{1}{1-\epsilon} \sup_{\mu,\nu} \frac{|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_{\frac{1}{2}(\mu+\nu)}}}$$

This leads to the following Theorem.

Theorem 3 *Let* (A1) *be satisfied and let* $\mathcal{D}$ *be reduced to the single element* γ. *Take pairs of probability measures* (μ,ν) *in* $\mathcal{M}$ *and let* $\lambda = \frac{1}{2}(\mu+\nu)$ *and* $\mathrm{Var}(\gamma|\lambda) = \int \gamma^2(s)\lambda(ds) - |\int \gamma(s)\lambda(ds)|^2$.

Then if the minimax risk of Theorem 1 is finite, it is equal to

$$\sup_{\mu,\nu} \left\{ \frac{|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}} - \mathrm{Var}(\gamma|\lambda) \right\}.$$

Proof One can apply the result of Theorem 2 taking an arbitrary λ. Then the Barankin or Kiefer bound is applicable as long as we work on the exact support of λ. This gives a minimax risk equal to

$$\sup_{\lambda,\mu,\nu} \left\{ \frac{|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}} - \mathrm{Var}[\gamma|\lambda) \right\}$$

where the sup is taken over all triplets (λ,μ,ν) of $\mathcal{M}$ such that $\mu + \nu << \lambda$. This is certainly greater than the value given in the Theorem. However, by the inequalities written before the statement of the theorem, one can restrict oneself to triplets such that $\lambda = \frac{1}{2}(\mu+\nu)$.

Hence the result

The quantity

$$\frac{|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}} - \mathrm{Var}(\gamma|\lambda)$$

can also be written in a different form.

Indeed $\mathrm{Var}(\gamma|\lambda) = \frac{1}{2}[\mathrm{Var}(\gamma|\mu) + \mathrm{Var}(\gamma|\nu)) + \frac{1}{4}|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2$. Also

$$\frac{|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2}{\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}} = \frac{1}{2}(\mathrm{Var}(T|P_\mu) + \mathrm{Var}(T|P_\nu)) + \frac{1}{4}|\langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle|^2$$

for an estimate T such that $\int T(dP_\mu - dP_\nu) = \langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle$ with minimum variance at P_λ.

Therefore the result of Theorem 3 can be rephrased as follows.

Theorem 4 *Let* (A1) *be satisfied and let $\mathcal{D}$ be reduced to the single element γ. For each pair (μ,ν) of probability measures with finite support on Θ and for $\lambda = \frac{1}{2}(\mu+\nu)$ let $T(\mu,\nu)$ be an estimate that minimizes* $\mathrm{Var}(T|P_\lambda)$ *subject to the condition that $\int T(dP_\mu - dP_\nu) = \langle\gamma,\mu\rangle - \langle\gamma,\nu\rangle$.*

Then the minimax risk of Theorem 1 is equal to

$$\sup_{\mu,\nu} \frac{1}{2} \left\{ \begin{array}{l} \mathrm{Var}[T(\mu,\nu)|P_\mu] + \mathrm{Var}[T(\mu,\nu)|P_\nu] \\ -\mathrm{Var}(\gamma|\mu) - \mathrm{Var}(\gamma|\nu) \end{array} \right\}.$$

Note that if one takes T such that $\int T dP_\theta = \gamma(\theta)$ for all θ in the support of λ, one will have

$$\mathrm{Var}[T|P_\mu] - \mathrm{Var}(\gamma|\mu) = \int [\mathrm{Var}_\theta T)\mu(d\theta).$$

Thus the formulation of Theorem 4 amounts to say that one can take $\sup_{\mu,\nu}$ in two steps. First one fixes $\lambda = \frac{1}{2}(\mu+\nu)$ and take the sup for over such (μ,ν). Then one takes the sup over λ.

7. The minimax problem for a binary experiment.

Let $\{P_1, P_2\}$ be a pair of probability measures on a σ-field $\mathcal{A}$. Let $\{\gamma_1,\gamma_2\}$ be a given pair of numbers. The problem considered in this section is that of finding estimates T such that $\int T dP_i = \gamma_i$, $i = 1, 2$ and such that $\max_i \mathrm{Var}_i T$ be as small as possible.

As an aside, we also consider the problem of finding a T with prescribed expectation and minimum variance under P_1. This is similar to the problem treated in Section 4 except that we do not impose the restriction that P_1 dominates P_2 and that $\int \left(\frac{dP_2}{dP_1}\right)^2 dP_1 < \infty$.

It is shown that if $P_2 << P_1$ but the integral $\int \left(\frac{dP_2}{dP_1}\right)^2 dP_1 = \infty$ a T with prescribed expectations and minimum variance under P_1 does not exist, whenever $\gamma_1 \neq \gamma_2$.

Throughout this section we shall assume that $\gamma_1 \neq \gamma_2$ and that $P_1 \neq P_2$, other cases being trivially solvable or unsolvable.

Except perhaps for the version given by Fabian and Hannan (1977), (see also Wolfowitz (1947)), the usual Cramér-Rao inequality is a consequence of inequalities valid for binary experiments. An inequality published by Kholevo (1973) and Pitman (1979) in terms of Hellinger distances already implies such Cramér-Rao inequalities.

Here the square Hellinger distance between measures (P, Q) is

$$h^2(P,Q) = \frac{1}{2}\int (\sqrt{dP} - \sqrt{dQ})^2.$$

We shall also use another square distance k^2 defined by

$$k^2(P,Q) = \frac{1}{2}\int \frac{(dP - dQ)^2}{d(P+Q)}.$$

This distance gives an inequality that is somewhat sharper than Kholevo's.

It is well known that both h and k define distances on the space of probability measures carried by a σ-field $\mathcal{A}$. For the case of k, see Le Cam (1986), page 47. Note also the inequalities

$$0 \leq h^2 \leq k^2 \leq h^2(2 - h^2) \leq 1.$$

If we consider a pair (P_1, P_2) of probability measures and write $E_i T$ and $\mathrm{Var}_i T$ for the expectation and variance of an estimate T under P_i one can prove the following

Lemma 1 *Suppose that $E_i T$ exists for $i = 1, 2$. Then*

$$\begin{aligned} \frac{1}{2}[\mathrm{Var}_1 T + \mathrm{Var}_2 T] &\geq |E_1 T - E_2 T|^2 \frac{1 - k^2}{4k^2} \\ &\geq |E_1 T - E_2 T|^2 \frac{1 - h^2}{4h^2(2 - h^2)}, \end{aligned}$$

with $k^2 = k^2(P_1, P_2)$ and similarly for h^2.

Proof The first inequality is proved directly in Le Cam and Yang (1990) page 127. It also follows from the results of Section 4, as we shall see below. The second inequality follows from the relations between h^2 and k^2.

Note We shall show below that the minimax variance for given (γ_1, γ_2) does not exceed $|\gamma_1 - \gamma_2|^2 \frac{1}{2k^2}$. Thus, in a sense the first inequality of Lemma 1 is in the right spirit. Note that the bound given there is attained for a special T. The second inequality is attainable only in the special case where $|\frac{dP_1-dP_2}{d(P_1+P_2)}|$ takes only one value.

However we shall show that in the case where h^2 is let tend to zero (as for the Cramér-Rao inequality) and under some uniform integrability condition, one can replace k^2 by $2h^2$.

We shall call the second inequality the Kholevo bound. The inequality given in Lemma 1 is a special case of inequalities obtainable by minimizing $\beta \mathrm{Var}_1 T + (1-\beta)\mathrm{Var}_2 T$ for $\beta \in [0,1]$. For symmetry of notation we shall let $\beta = \frac{1}{2}(1+\alpha)$, $(1-\beta) = \frac{1}{2}(1-\alpha)$ where α now ranges over $[-1,+1]$. Also let

$$M_\alpha = \frac{1}{2}(1+\alpha)P_1 + \frac{1}{2}(1-\alpha)P_2.$$

and $K(\alpha) = \int \frac{(dP_1-dP_2)^2}{dM_\alpha}$ so that $K(0) = 4k^2$.

Lemma 2 *Assume that $|\alpha| < 1$. Then for $D = \gamma_1 - \gamma_2$ and for estimates such that $E_i T = \gamma_i$ one has*

$$\inf_T \frac{1}{2}[(1+\alpha)\mathrm{Var}_1 T + (1-\alpha)\mathrm{Var}_2 T] = D^2 \left\{ \frac{1}{K(\alpha)} - \frac{1}{4}(1-\alpha^2) \right\}.$$

Proof This follows from the formulas of Section 6 above applied to the triplet $(P_\lambda, P_\mu, P_\nu)$ with $\lambda = \frac{1}{2}(1+\alpha)\delta_1 + \frac{1}{2}(1-\alpha)\delta_2$, δ_i Dirac mass at i and with $P_\mu = P_1$, $P_\nu = P_2$. It is sufficient to check that $\mathrm{Var}(\gamma|\lambda) = \frac{1}{4}(1-\alpha^2)D^2$.

The estimate T_α that achieves the minimum variance under M_α is given by $T_\alpha = \frac{D}{K(\alpha)}\frac{dP_1-dP_2}{dM_\alpha}$. Thus, for instance

$$\int T_\alpha^2 dP_1 = \frac{D^2}{K^2(\alpha)} \int \left[\frac{dP_1-dP_2}{dM_\alpha}\right]^2 dP_1$$

and

$$\mathrm{Var}_1 T_\alpha = \frac{D^2}{K^2(\alpha)} \int \left[\frac{dP_1-dP_2}{dM_\alpha}\right]^2 dP_1 - \left[\left(\frac{1-d}{2}\right) D\right]^2.$$

Note that $K(\alpha)$ depends only on α and the pair (P_1, P_2) while the numerator in the risk depends only on $D = \gamma_1 - \gamma_2$. This separation does not occur in more general cases although a sort of separation occurs whenever γ takes only two values.

To obtain the minimax risk one should take $\sup_\alpha \left\{\frac{1}{K(\alpha)} - \frac{1}{4}(1-\alpha^2)\right\}$, $|\alpha| < 1$. Note that K is a convex function of α. Equating derivatives to zero gives the relation

$$2\alpha \left(\int \frac{w^2}{1+\alpha w} dM_0\right)^2 = \int \frac{w^3}{2+\alpha w} dM_0$$

where $w = \frac{1}{2}\frac{dP_1 - dP_2}{dM_0}$, $M_0 = \frac{1}{2}(P_1 + P_2)$.

It can be checked that this relation is precisely equivalent to the condition that $\mathrm{Var}_1 T_\alpha = \mathrm{Var}_2 T_\alpha$.

Note that unless $\int w^3 dM_0 = 0$, the maximum risk will not be attained at $\alpha = 0$ and will therefore be larger than the bound in Lemma 1.

There are interesting cases where $\int w^3 dM_0 = 0$. This happens for instance when the distribution of w under M_0 is symmetric around zero, as happens if P_1 and P_2 are obtained by shifts of a symmetric measure. However, in general, one must expect that $\int w^3 dM_0$ will be different from zero. Thus, no matter what α will be, the following result is of some interest.

Lemma 3 *The minimax variance r satisfies the following inequalities*

$$\begin{aligned} D^2\left[\frac{1}{K(0)} - \frac{1}{4}\right] &\leq r = D^2 \sup_{|\alpha|<1}\left\{\frac{1}{K(\alpha)} - \frac{1}{4}(1-\alpha^2)\right\} \\ &\leq D^2 \sup_{0\leq\alpha<1}\left\{\frac{1+\alpha}{K(0)} - \frac{1}{4}(1-\alpha^2)\right\} \\ &\leq \frac{2D^2}{K(0)}. \end{aligned}$$

Proof Note that the left most term is precisely the bound given in Lemma 1 in terms of k^2. The fact that one needs only to look at $\alpha \in (-1+1)$ instead of the closed interval $[-1,+1]$ follows from the convexity of K, even if it is not continuous at the points -1 or +1.

To proceed let us look at the form of $K(\alpha)$ in terms of the function $w = \frac{dP_1 - dP_2}{dM_0}$. One has $K(\alpha) = 4\int \frac{w^2}{1+\alpha w} dM_0$. Now $|\alpha| < 1$ and $|w| \leq 1$. Thus

$$\int \frac{w^2}{1+\alpha w} \geq \int \frac{w^2}{1+|\alpha|} dM_0 = \frac{1}{1+|\alpha|}[\frac{1}{4}K(0)].$$

The inequalities follow.

One could ask whether the upper bound can be improved. It cannot in general according to the next result.

Lemma 4 *For any $\epsilon \in (0,1)$ there exist pairs (P_1, P_2) such that the minimax risk r is larger than $(2-\epsilon)\frac{D^2}{K(0)}$.*

Proof Let us look at the distribution under M_0 of the function w. This is a probability measure S on $[-1,+1]$ such that $\int w dS = 0$. Conversely any S satisfying these conditions can be obtained from a binary experiment. Indeed the measure S is a slightly modified version of Blackwell's canonical measure.

Construct a probability measure S that places a mass $\pi < 1/2$ at 1 and a mass $1-\pi$ at $-\frac{\pi}{1-\pi}$. One will have $\int w dS = 0$ and $\int w^2 dS = \pi\left(1+\frac{\pi}{1-\pi}\right)$

$$\int \frac{w^2}{1+\alpha w} dS = \pi \left\{ \frac{1}{1+\alpha} + \frac{\pi}{1-(1+\alpha)\pi} \right\}$$

for $\alpha > -1$. Thus for $\alpha > 0$ and for π sufficiently small one can make the ratio $\frac{K(\alpha)}{K(0)}$

$$= \left[\frac{1}{1+\alpha} + \frac{\pi}{1-(1+\alpha)\pi} \right] \left[1 + \frac{\pi}{1-\pi} \right]^{-1}$$

as close to $\frac{1}{1+\alpha}$ as desired. This implies the desired result.

Remark 1. The example used in the proof of Lemma 4 is also such that for the distances h and k one has $\frac{k^2}{2h^2}$ as close to $\frac{1}{2}$ as one wishes. In such case, since both k^2 and h^2 are small, the first bound of Lemma 1 is almost twice as large as the Kholevo bound.

Remark 2. The proof of Lemma 4 uses pair (P_1, P_2) such that $K(0)$ tends to zero. This is in the nature of things. One can show that $K(\alpha) - \frac{K(0)}{1+\alpha}$ cannot tend to zero unless $K(0)$ does. Indeed takes two numbers $\alpha \in (0,1)$ and $t \in (0,1)$. One can write that

$$\begin{aligned} K(\alpha) - \frac{K(0)}{1+\alpha} &\geq \int w^2 \left(\frac{1}{1-\alpha t} - \frac{1}{1+\alpha} \right) I[-1,-t] S(dw) \\ &+ \int w^2 \left(\frac{1}{1-\alpha t} - \frac{1}{1+\alpha} \right) I[(-t,t]] S(dw) \\ &\geq \frac{\alpha}{1+\alpha} \frac{1-t}{1+\alpha t} \int w^2 I[-1,t] S(dw). \end{aligned}$$

Thus if $K(\alpha) - \frac{K(0)}{1+\alpha}$ tends to zero, so does $\int w^2 I(-1,t] S(dw)$ and, in particular $\int w^- Sd(w) \to 0$. However $\int w^- S(dw) = \int w^+ S(dw)$. Therefore $\int |w| S(dw)$ and $\int w^2 S(dw)$ tend to zero.

Sequences of pairs $(P_{1,n}, P_{2,n})$ such that the corresponding $K(0) = K_n(0)$ tend to zero occur often. Indeed they are involved in the classical Cramér-Rao inequality, except in the Fabian-Hannan proof which uses weak convergence in Hilbert space. (See however Section 4 above). For such cases one can obtain a special result as follows

Lemma 5 *Let $\{\mathcal{E}_n\}$ be a sequence of experiments $\mathcal{E}_n = \{P_{1,n}, P_{2n}\}$. Let $Q_n = \frac{1}{2}[P_{1,n} + P_{2,n}]$ and $f_{i,n} = \frac{dP_{i,n}}{dQ_n}$.*

Assume that the Hellinger square distances $h_n^2 = \frac{1}{2}\int(\sqrt{dP_{1,n}} - \sqrt{dP_{2,n}})^2$ *are non zero but tend to zero as* $n \to \infty$. *Let* X_n *be the random variable* $X_n = \frac{1}{2}[\sqrt{f_{1,n}} - \sqrt{f_{2,n}}]^2$ *for the distribution induced by* Q_n.

Assume that if F *is any cluster point of the sequence* $\mathcal{L}\left(\frac{X_n}{EX_n}\right)$ *then* $\int xF(dx) = 1$.

Then the quantities

$$\frac{k^2(P_{1,n}, P_{2,n})}{2h_n^2}$$

and $\sup_{|\alpha|\leq 1} \frac{K_n(0)}{K_n(\alpha)}$ *tend to unity.*

Proof Dropping the index n for simplicity, let $\varphi_i = \sqrt{f_i}$. Then one can write

$$k^2 = \frac{1}{4}\int(\varphi_1^2 - \varphi_2^2)^2 dQ \text{ and } 2h^2 = \int(\varphi_1 - \varphi_2)^2 dQ.$$

From this it follows that

$$\begin{aligned} k^2 &= \frac{1}{4}\int(\varphi_1 - \varphi_2)^2(\varphi_1 + \varphi_2)^2 dQ \\ &= \frac{1}{4}\int(\varphi_1 - \varphi_2)^2[4 - (\varphi_1 - \varphi_2)^2]dQ \end{aligned}$$

since $\varphi_1^2 + \varphi_2^2 = 2$. Thus

$$\begin{aligned} k^2 &= 2h^2 - \frac{1}{4}\int(\varphi_1 - \varphi_2)^4 dQ \\ &= 2h^2 - EX^2 \end{aligned}$$

for the variable $X = \frac{1}{2}(\sqrt{f_1} - \sqrt{f_2})^2$. Returning to the sequence $\mathcal{E}_n$ one has

$$\frac{k_n^2}{2h_n^2} = 1 - \frac{1}{2}\frac{EX_n^2}{EX_n} = 1 - \frac{1}{2}E(X_n)\left(\frac{X_n}{EX_n}\right).$$

The condition imposed on the limiting distribution of X_n/EX_n implies that for every $\epsilon > 0$ there is a $b = b(\epsilon)$ and an $N = N(\epsilon)$ such that $n \geq N$ implies

$$E\left[\frac{X_n}{EX_n}\right] I(X_n \geq bEX_n] \leq \epsilon.$$

Since $0 \leq X_n \leq 1$ this gives

$$E\{(X_n)\left[\frac{X_n}{EX_n}\right] I[X_n \geq bEX_n]\} \leq \epsilon.$$

The other part $E(X_n)\left[\frac{X_n}{EX_n}\right] I[X_n < bEX_n)$ is bounded by bEX_n which tends to zero by assumption. This proves the first assertion. For the second one note that the variable w introduced to represent $K_n(\alpha)$ as $4\int \frac{w^2}{1+\alpha w} dQ_n$ is just

$$w_n = \frac{1}{2}[f_{1,n} - f_{2,n}] = \frac{1}{2}(\sqrt{f_{1,n}} - \sqrt{f_{2,n}})(\sqrt{f_{1,n}} + \sqrt{f_{2,n}}).$$

Therefore $w_n^2 \leq 8X_n$.

Since $Ew_n^2 = k_n^2 \geq h_n^2$. The variables $Y_n = \frac{w_n^2}{Ew_n^2}$ will satisfy the same uniform integrability condition as the $\frac{X_n}{EX_n}$.

The ratio $\frac{K_n(\alpha)}{K_n(0)}$ can be written in the form $\frac{K_n(\alpha)}{K_n(0)} = E\frac{Y_n}{EY_n}[1 + Z_n(\alpha)]$ with $Z_n(\alpha) = \frac{1}{1+\alpha w_n} - 1 = -\frac{\alpha w_n}{1+\alpha w_n}$.

If one takes a fixed $\epsilon \in (0, 1/2)$ and look only at values of $\alpha \in [-1+\epsilon, 1-\epsilon]$ one will have $|Z_n(\alpha)| \leq \frac{|\alpha|}{1-|\alpha|}$. The argument applied above to $\frac{X_n}{EX_n}X_n$ applies also to $\frac{Y_n Z_n(\alpha)}{EY_n)}$. Therefore $E\left[\frac{Y_n}{EY_n}Z_n(\alpha)\right]$ tends to zero as $n \to \infty$.

By varying $\alpha \in [-1+\epsilon, 1-\epsilon]$ as n changes or by noting that $K_n(\alpha)$ is a convex function of α, one concludes that $\frac{K_n(\alpha)}{K_n(0)}$ tends to unity uniformly on $[-1+\epsilon, 1-\epsilon]$. Using the convexity again, one sees that if $|\frac{K_n(\alpha)}{K_n(0)} - 1| < \eta$ on $[-1+\epsilon, 1+\epsilon]$ then $\frac{K_n(\alpha)}{K_n(0)} \geq 1 - \frac{\eta}{1-\epsilon}$ on the segments $[-1, -1+\epsilon]$ and $[1-\epsilon, 1]$. This yields $\inf_\alpha \frac{K_n(\alpha)}{K_n(0)} \geq 1 - \frac{\eta}{1-\epsilon}$ and therefore $\sup_\alpha \frac{K_n(0)}{K_n(\alpha)} \leq (1 - \frac{\eta}{1-\epsilon})^{-1}$. This completes the proof of the second assertion.

Remark. If one obtains the Cramér-Rao inequality by passages to the limit in the first inequality of Lemma 1 or in Kholevo's inequality. (see E.J.G. Pitman [1949]) or Simons and Woodroofe [1983]) the result of Lemma 5 implies that one will obtain the same result whether one uses the first or the second inequality of Lemma 1 or if one passes to the limit for the minimax risk.

Note, however, that this is proved here under an equi integrability condition on $\frac{X_n}{EX_n}$. If such a condition is removed the ratios $\frac{k_n^2}{2h_n^2}$ can tend to anything one pleases in $(\frac{1}{2}, 1)$. Similarly $\frac{K_n(0)}{K_n(\alpha)}$ can tend to anything one pleases between $\frac{1}{2}$ and 1 and the quantity $2h_n^2 \sup_\alpha \frac{1}{K_n(\alpha)}$ may tend to 4.

It is true that the equi integrability condition imposed on X_n/EX_n is not quite a necessary condition for the conclusions of Lemma 5. However it is a condition that will often appear automatically. For instance, for an experiment $\mathcal{E} = \{P_\theta; \theta \in \mathbb{R}\}$ and pairs $(P_0, P_{\theta)_n})$ where $\theta_n \to 0$, the condition will be implied by the often used differentiability in quadratic mean of the square roots of densities.

8. Minimum variance at one point.

This section treats the problem of finding estimates T such that $\int T dP_i = \gamma_i$, $\gamma_1 \neq \gamma_2$, for a binary experiment $\mathcal{E} = \{P_1, P_2\}$ with the requirement that $\text{Var}_1 T$ be minimum.

We shall describe the solution in terms of the mixtures $\frac{1}{2}(1+\alpha)P_1 + \frac{1}{2}(1-\alpha)P_2$ as α increases to unity. The limit $\lim_\alpha\{K(\alpha), \alpha < 1, \alpha \uparrow 1\} = K(1)$ will be involved.

One can always assume $\gamma_1 = 0$, $\gamma_2 = -D \neq 0$.

Proposition 1

a) If P_2 possesses a nonzero part that is P_1 singular, then there exists a T with the required expectation and $\text{Var}_1 T = 0$, $\text{Var}_2 T < \infty$.

b) If $K(1) = \infty$ but $P_2 << P_1$, there exist estimates with the prescribed expectation and $\text{Var}_1 T$ *as small as one pleases. However there is no estimate achieving the prescribed expectation and* $\text{Var}_1 T = 0$.

c) If $K(1) < \infty$, then $P_2 << P_1$ and $\int \left(\frac{dP_2}{dP_1}\right)^2 dP_1 < \infty$. An estimate of the form $T = \frac{D}{K(1)} \frac{dP_1 - dP_2}{dP_1}$ will achieve minimum variance under P_1.

Note Under the condition $K(1) < \infty$ the quantity $\text{Var}_2 T$ may be infinite. It will be finite if and only if $\int \left(\frac{dP_2}{dP_1}\right)^3 dP_1 < \infty$.

The proof will be divided into small lemmas.

Lemma 1 *As α increases to 1 (by values $\alpha < 1$) the $K(\alpha)$ tends to $K(1) = 2\int \frac{(dP_1 - dP_2)^2}{dP_1}$.*

Proof In the notation of Section 7 one has

$$K(\alpha) = 4 \int \frac{w^2}{1 + \alpha w} dM_0 \text{ with } M_0 = \frac{1}{2}(P_1 + P_2)$$

and $w = \frac{1}{2} \frac{dP_1 - dP_2}{dM_0}$.

On the set where $w < 0$ the integrand increases as $\alpha > 0$ increases. On the set where $w \geq 0$ the integrand decreases as α increases. The limits are

$$\int_{w<0} \frac{w^2}{1+w} dM_0 \text{ and } \int_{w \geq 0} \frac{w^2}{1+w} dM_0.$$

This last term is always finite. Hence the limit of $K(\alpha)$ is $4\int \frac{w^2}{1+w} dM_0$ whether it is finite or not. This gives the desired result.

Lemma 2 *Let T_α be defined as in Section 7. then as $\alpha < 1$ tends to unity $(1-\alpha)\text{Var}_2 T_\alpha$ tends to zero and $\frac{1}{2}(1+\alpha)\text{Var}_1 T_\alpha$ tends to $\frac{D^2}{K(1)}$.*

Proof One has always

$$\begin{aligned} \frac{1}{2}(1-\alpha)\text{Var}_2 T_\alpha &\leq \left(\frac{1-\alpha}{2}\right)\text{Var}_2 T_\alpha + \left(\frac{1+\alpha}{2}\right)\text{Var}_1 T_\alpha \\ &= D^2 \left\{ \frac{1}{K(\alpha)} + \frac{1}{4}(\alpha^2 - 1) \right\}. \end{aligned}$$

Thus, if $K(1) = \infty$, all those terms must tend to zero. If on the contrary $K(1) < \infty$ then

$$\mathrm{Var}_1 T_\alpha = \frac{D^2}{K^2(\alpha)} \int \left(\frac{dP_1 - dP_2}{dM_\alpha} \right)^2 dP_1 - \left[\frac{1}{2}(1-\alpha)D \right]^2$$

tends to $\frac{D^2}{K(1)}$. This is the same as the limit of $D^2 \left[\frac{1}{K(\alpha)} + \frac{1}{4}(\alpha^2 - 1) \right]$, hence $(1-\alpha)\mathrm{Var}_2 T_\alpha$ must tend to zero.

With this we can pass to the proof of Proposition 1.

Proof (of Proposition 1) Part (a) – Let V be the part of P_2 that is P_1 singular, with $\|V\| > 0$. Take $T = 0$ on a set that carries P_1 and $\gamma_2/\|V\|$ on a set that carries V. This will give a result as stated in (a). For Part (b) note that, since $P_2 << P_1$, T_α may be written

$$\begin{aligned} T_\alpha &= \frac{D}{K(\alpha)} \frac{2(dP_1 - dP_2)}{(1+\alpha)dP_1 + (1-\alpha)dP_2} \\ &= \frac{D}{K(\alpha)} \frac{2(1-f)}{(1+\alpha) + (1-\alpha)f} \end{aligned}$$

with $f = \frac{dP_2}{dP_1}$. If $K(1) = \infty$, but $P_2 << P_1$, T_α will tend to zero in P_1-probability as $\alpha \to 1$. Thus it tends to zero in P_2-probability as well. However $\int T_\alpha dP_2 = -(1+\alpha)\frac{D}{2}$. Therefore the integrals $\int |T_\alpha| dP_2$ cannot be uniformly convergent and $\mathrm{Var}_2 T_\alpha$ will tend to infinity.

On the contrary $\mathrm{Var}_1 T_\alpha \leq \frac{D^2}{K(\alpha)}$ will tend to zero. However there are no functions T such that $\int T dP_1 = \int T^2 dP_1 = 0$ and $\int T dP_2 = \gamma_2 \neq 0$.

For part (c), note that $K(1) < \infty$ if and only if $P_2 << P_1$ and $\int \left(\frac{dP_2}{dP_1}\right)^2 dP_1 < \infty$.

In this case T_α tends in P_1-probability to $T_1 = \frac{D}{K(1)} \frac{dP_1 - dP_2}{dP_1}$. One has

$$\int T_1 dP_1 = 0, \quad \int T_1^2 dP_1 = \lim_\alpha \int T_\alpha^2 dP_1$$

and T_1 is the strong limit in $L_2(P_1)$ of the T_α. Since $(dP_2/dP_1) \in L_2(P_1)$ one has limit $\int T_\alpha dP_2 = \int T_1 dP_2 = \gamma_2$.

This concludes the proof of the Proposition.

For the remark following the statement of the proposition note that $\mathrm{Var}_2 T_1$ will contain a term proportional to

$$\int \left(\frac{dP_1 - dP_2}{dP_1} \right)^2 dP_2$$

which will be finite only if $\int \left(\frac{dP_2}{dP_1}\right)^2 dP_2 < \infty$ or equivalently $\int \left(\frac{dP_2}{dP_1}\right)^3 dP_1 < \infty$.

9. The case of measures that are nearly equal.

It was shown in Lemma 5, Section 7, that, for pairs (P_1, P_2) for which $k^2(P_1, P_2)$ is small, it can happen that $k^2(P_1, P_2)[2h^2(P_1, P_2)]^{-1}$ be close to unity. For such a result to hold h^2 and k^2 must be small since $h^2(2-h^2)/k^2 \geq 1$ with approximate equality only if the variance of $\sqrt{f_1 f_2}$ under $M\frac{1}{2}(P_1+P_2)$ is small. Here $f_i = dP_i/dM$.

The fact that $k^2/2h^2$ tends to unity locally under the conditions of differentiability in quadratic mean (for square roots of densities) often used in the literature gives this very special case an importance that it does not deserve otherwise.

In the present section we give an extension to the case of an m-tuple $\{P_i; i = 1, 2, \ldots, m\}$ dominated by a common probability measure M.

Let $f_i = dP_i/dM$. Recall that the matrices B of Section 4 have entries of the form $B_{i,j} = \int (f_i - 1)(f_j - 1)dM$. We shall also consider the corresponding correlation matrices C with

$$C_{i,j} = B_{i,j}\{\int (f_i - 1)^2 dM \int (f_j - 1)^2 dM\}^{-1/2}.$$

Consider also the matrices with entries $H_{i,j} = \frac{1}{2}\int(\sqrt{f_i} - 1)(\sqrt{f_i} - 1)dM$. This is not a centered covariance matrix since $\int[\sqrt{f_j} - 1]dM = 1 - h_j^2$, but since h_j^2 will be small this will not matter. We shall also consider the corresponding "correlation" matrix R with entries

$$R_{i,j} = H_{i,j}\{\frac{1}{4}\int(\sqrt{f_i} - 1)^2 dM \int(\sqrt{f_j} - 1)^2 dM\}^{-1/2}.$$

Consider then the following assumptions, which refer to sequences that should be written $\{P_{in}\}$, $\{M_n\}$ etc but they will not suffer if one omits the "n".

(A1) $h^2(P_i, M)$ tends to zero for $i = 1, 2, \ldots, m$

(A2) There is a fixed c_1 such that $dP_i/dM \leq c$ for $i = 1, 2, \ldots, m$ and m is fixed

(A3) For $f_i = \frac{dP_i}{dM}$, consider the variables $Z_i = \frac{(\sqrt{f_i}-1)^2}{2h_i^2}$, with $h_i^2 = \frac{1}{2}\int(\sqrt{f_i} - 1)^2 dM$ as usual.

Any weak cluster point F of the distributions $\mathcal{L}[Z_i|M]$ is such that $\int xF(dx) = 1$.

(Here the weak convergence used is that where $E\varphi(Z_i) \to \int \varphi(z)F(dz)$ for bounded continuous functions φ).

Proposition 1 *Assume that (A1) (A2) (A3) hold. Then* $\frac{1}{8h_i^2}\int(f_i - 1)^2 dM$ *tends to unity and the difference between the matrices C and R tends to zero.*

Proof Write

$$\begin{aligned} f_j - 1 &= 2(\sqrt{f_j} - 1) + (\sqrt{f_j} - 1)[\sqrt{f_j} + 1 - 2] \\ &= 2(\sqrt{f_j} - 1) + (\sqrt{f_j} - 1)^2. \end{aligned}$$

This gives

$$(f_j - 1) = 4(\sqrt{f_j} - 1)^2 + (\sqrt{f_j} - 1)^4 + 4(\sqrt{f_j} - 1)^3.$$

It follows then from the argument carried out for Lemma 5, Section 7 that $\int (f_j - 1)^2 dM / 4 \int (\sqrt{f_j} - 1)^2 dM$ tends to unity.

Similarly

$$\begin{aligned} (f_j - 1)(f_i - 1) &= 4(\sqrt{f_j} - 1)(\sqrt{f_i} - 1) \\ &+ 2(\sqrt{f_j} - 1)(\sqrt{f_i} - 1)^2 \\ &+ \ldots \end{aligned}$$

According to Schwarz inequality, one has

$$\begin{aligned} \left|E\frac{(\sqrt{f_i} - 1)}{h_i}\frac{(\sqrt{f_j} - 1)}{h_j}\right|^2 &\leq E\frac{(\sqrt{f_i} - 1)^2}{h_i^2} E\frac{(\sqrt{f_j} - 1)^4}{h_j^2} \\ &= 2E\frac{(\sqrt{f_i} - 1)^4}{h_j^2}. \end{aligned}$$

Similarly for the other terms. Therefore C-R tends to zero. This completes the proof.

Note One can relax the condition (A2) that $P_i \leq c_1 M$ but then (A3) must be strengthened. The restriction that the cardinality m of the index set be fixed is somewhat irrelevant. Note that in Lemma 5, Section 7, condition (A3) was automatic, since M was $\frac{1}{2}(P_1 + P_2)$.

The importance, if any, of this proposition resides in the fact that, for numbers such as $\int \frac{dP_i dP_j}{dM}$, the relevant geometry is that of the Hilbert space $L_2(M)$. By contrast for numbers $\int (\sqrt{f_i} - 1)(\sqrt{f_j} - 1) dM = \int [\sqrt{dP_i} - \sqrt{dM}][\sqrt{dP_j} - \sqrt{dM}]$ the relevant geometry is the Hilbertian geometry induced by the Hellinger distance. This is a “universal” geometry not attached to a particular M.

Here the matrix R is one in which one views the P_i's from the “center” M, but M does not appear in the definition of the distances.

10. Inequalities of Cramér-Rao type.

The purpose of this section is to link the arguments of the previous sections to the common form of the Cramér-Rao inequalities. The most usual forms assume that Θ is a subset of a vector space V. They use various differentiability assumptions.

In the previous sections, our set Θ was just a set without any special structure. However note that in Section 2 we have embedded the family $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ in a linear space $S(\mathcal{E})$. We also used the space $\mathcal{M}$ of probability measures with finite support on Θ. This is a convex subset of the linear space $\mathcal{F}(\Theta)$ of all finite signed measures with finite support on Θ.

In this sense we embedded Θ in a linear space. There is more. In Section 4, Lemma 4 we used measures of the form $\mu(\epsilon) = \epsilon\mu + (1-\epsilon)\delta$ where $\epsilon \in (0,1)$ and where δ is the Dirac mass at a particular point Δ.

The bound of Section 4 could be expressed in the following manner.

One takes $\mu(\epsilon)$ as above and $\nu(\epsilon) = \epsilon\nu + (1-\epsilon)\delta$. Let then

$$b(\mu,\nu) = \lim_{\epsilon \to 0} \frac{1}{\epsilon}[\langle \gamma, \mu(\epsilon)\rangle - \langle \gamma, \nu(\epsilon)\rangle]$$

and

$$\sigma^2(\mu,\nu) = \lim_{\epsilon \to 0} \frac{1}{\epsilon^2} \int \frac{[dP_{\mu(\epsilon)} - dP_{\nu(\epsilon)}]^2}{dP_\Delta}$$

(These limits certainly exist since the quantities involved do not depend on ϵ.)

The bound is then

$$\mathrm{Var}_\Delta T \geq \frac{b^2(\mu,\nu)}{\sigma^2(\mu,\nu)}$$

for all possible pairs (μ,ν).

That looks very much like an ordinary Cramér-Rao inequality and is indeed an immediate consequence of Schwarz inequality.

The added feature of Section 4 is that it is proved that the bound is attainable, since we used the supremum over (μ,ν) and assumed that $\int (\frac{dP_\theta}{dP_s})^2 dP_\Delta < \infty$ for every θ.

The most usual forms of the Cramér-Rao inequality assume that Θ is a subset of a linear space, say V. Then the procedures of Section 4 would imbed V, or at least the linear span of Θ, into our space $\mathcal{F}(\Theta)$ of signed measures, but one can proceed otherwise and use differentiability relations. First assume for simplicity that our special point $\Delta \in \Theta$ is the origin of V.

Then for maps Z defined on Θ and with values in a locally convex topological vector space one can consider various definitions of differentiability.

One of them is Gâteaux differentiability as follows

Definition 1 The map Z is Gâteaux differentiable at zero with derivative D if D is a linear map such that for fixed $v \in V$

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon}[Z(\epsilon v) - Z(0) - \epsilon D v] = 0.$$

For this to make sense one should assume that $\epsilon v \in \Theta$ for ϵ small enough.

There is another concept called quasi-differentiability in Dieudonné (1960) page 151. For this one assume that V itself is a topological vector space. Let then Φ be the set of continuous maps φ from $[0, 1]$ to Θ such that $\varphi(0) = 0$ and such that $\lim_{t \to 0} \frac{1}{t}[\varphi(t) - \varphi(0)] = \varphi'(0)$ exists in V.

Definition 2 The map Z is quasi-differentiable at zero if there is a linear map D such that

$$\lim_{t \to 0} \frac{1}{t}[Z[\varphi(t)] - Z[\varphi(0)] - tD\varphi'(0)] = 0$$

for every $\varphi \in \Phi$.

With this concept comes the idea of a tangent set or space. Let $\mathcal{C}_0$ be the set of derivatives $\varphi'(0)$ for $\varphi \in \Phi$. It is always a closed cone. Following Bickel, Klaassen, Ritov and Wellner one can call it the tangent *set* of Θ at zero. These same authors call the linear span $\mathcal{C}$ of $\mathcal{C}_0$ the tangent *space* of Θ at zero. Pfanzagl [1982] has amply demonstrated the value of such concepts.

There are many other definitions of differentiability. For instance if V is a normed space one can use the concept of Fréchet differentiability. We shall return to that later.

With such definitions one can produce various forms of Cramér-Rao inequalities.

One of them imitated from Fabian and Hannan [1977] is as follows.

Lemma 1 *Assume that V is a topological vector space and that the map $\theta \rightsquigarrow P_\theta$ satisfies the following conditions:*

1) For θ close enough to zero one has $P_\theta << P_0$. Furthermore the density $\frac{dP_\theta}{dP_0}$ is square integrable for P_0.

Let $\theta \rightsquigarrow Z(\theta)$ be the map $\theta \rightsquigarrow Z(\theta) = \frac{dP_\theta}{dP_0}$ considered as a map to $L_2(P_0)$ on a neighborhood of zero.

2) Assume that $\theta \rightsquigarrow Z(\theta)$ is quasi-differentiable at zero with quasi-derivative D for the weak topology of $L_2(P_0)$.

Then if T is an estimate such that $\int T^2 dP_0 < \infty$ and such that $\int T dP_\theta = \gamma(\theta)$ the quasi-derivative Γ of γ at zero exist and satisfies

$$\text{Var}_0 T \geq \frac{|\Gamma v|^2}{\|Dv\|^2}$$

for all v in the tangent space of Θ at zero.

Proof To see that γ is quasi-differentiable note that for $\varphi \in \Phi$

$$\frac{1}{t}[\gamma[\varphi(t)] - \gamma[\varphi(0)]] = \int T\frac{1}{t}[Z[\varphi(t)] - Z(\varphi(0)]]dP_0.$$

By assumption this converges to $\int T[D\varphi'(0)]dP_0$. Thus γ has derivative $v \rightsquigarrow \Gamma v = \int T[Dv]dP_0$. By linearity this relation is valid over the entire linear span of the tangent set $\mathcal{C}_0$.

The stated inequality is obtainable from Schwarz inequality applied to $\int T[Dv]dP_0$.

Remark. Following Fabian and Hannan [1977], see also Wolfowitz [1947], we have assumed quasi-differentiability for the weak topology of $L_2(P_0)$. As shown in Section 4 if the range of $\theta \rightsquigarrow Z(\theta)$ is convex one could use quasi-differentiability for the norm of $L_2(P_0)$ without changing anything in the result.

Note that the assumptions include the domination condition $P_\theta << P_0$ with $\int(\frac{dP_\theta}{dP_0})^2 dP_0 < \infty$ at least for θ close enough to zero. Kholevo's inequality stated in Lemma 1, Section 7, suggests that one could remove such conditions and use instead the quasi-differentiability of the map $\theta \rightsquigarrow \sqrt{\frac{dP_\theta}{dP_0}}$. Indeed one can, but not necessarily for *all* estimates T such that $\int T^2 dP_0 < \infty$.

Let us say that $\theta \rightsquigarrow P_\theta$ satisfies condition DQM at zero if the following two requirements hold.

1) The map $\theta \rightsquigarrow \sqrt{\frac{dP_\theta}{dP_0}}$ from Θ to $L_2(P_0)$ is quasi-differentiable at zero for the norm of $L_2(P_0)$.

Let $\frac{1}{2}D$ be its quasi-derivative.

2) If $\varphi \in \Phi$ then the mass of $P_{\varphi(t)}$ that is P_0 singular tends to zero faster than t^2 as t tends to zero.

The second condition is to insure that $\frac{1}{t^2}\int(\sqrt{dP_{\varphi(t)}} - \sqrt{dP_0})^2$ tends to $\|\frac{1}{2}D\varphi'(0)\|^2$. It will be automatically satisfied if the tangent *set* $\mathcal{C}_0$ is a linear space.

Kholevo's inequality would yield lower bounds for $\frac{1}{2}[\mathrm{Var}_{\varphi(t)}T + \mathrm{Var}_0 T]$ and limits of such quantities. However for arbitrary T such that $\int T^2 dP_0 < \infty$ it does not guarantee differentiability of $\int T dP_\theta$. One can construct simple examples when the mass of the part of P_θ that is P_0 singular is not zero. Here is an example where $P_\theta << P_0$ but where $\int(\frac{dP_\theta}{dP_0})^2 dP_0 = \infty$.

Take for P_0 the Lebesgue measure on $(-\frac{1}{2}, +\frac{1}{2})$. On the interval $(2^{-n}, 22^{-n}]$, $n = 3, 4, \ldots$, let Q_n be the probability measure carried by $(2^{-n}, 22^{-n}]$ that has a density proportional there to $(x - 2^{-n})^{-1/2}$.

Then, according to Section 8, Proposition 1, for any $\epsilon > 0$ and any prescribed a_n there are estimates T_n such that $\int T_n dP_0 = 0$, $\int T_n dQ_n = a_n$ and $\text{Var}_0 T_n < \epsilon 2^{-n}$.

One can assume that T_n (defined on $(-1/2, +1/2)$) is zero outside of $(2^{-n}, 22^{-n}]$.

For $\theta \in (2^{-n}, 22^{-n}]$ let $P_\theta = (1-\theta^3)P_0 + \theta^3 Q_n$. Furthermore let $T = \sum_n T_n$.

Then for $\theta \in (2^{-n}, 22^{-n}]$ one will have $\int T dP_\theta = \theta^3 a_n$ and also $\int T^2 dP_0 \leq \epsilon \sum_{n=3}^{\infty} 2^{-n} < \epsilon$.

Since the number a_n are arbitrary they certainly can be chosen so that $\int T dP_\theta$ behaves quite arbitrarily as θ tends to zero by positive values. Kholevo's inequality will give bounds on $\text{Var}_\theta T + \text{Var}_0 T$ but not on $\text{Var}_0 T$.

The family $\{P_\theta; \theta \in \Theta\}$ used in this example satisfies DQM since for $\theta \in (2^{-n}, 22^{-n}]$,

$$\frac{1}{2}\int(\sqrt{dP_\theta} - \sqrt{dP_0})^2 \leq \int |dP_\theta - dP_0| = \theta^3 \|Q_n - P_0\| \leq 2\theta^3.$$

Note that, according to Lemma 5, Section 7, one would not change anything by replacing Kholevo's inequality by the first inequality of Lemma 1, Section 7.

However one can prove the following

Lemma 2 *Assume that (DQM) is satisfied at zero with quasi-derivative $\frac{1}{2}D$.*

Let T be an estimate such that $\int T dP_\theta = \gamma(\theta)$ and such that for every $\varphi \in \Phi$ one has $\limsup_{t\to 0} \int T^2 dP_{\varphi(t)} < \infty$.

Then γ is quasi-differentiable at zero with quasi-derivative Γ such that $\Gamma v = \int T(Dv)dP_0$. The inequality of Lemma 1 remain valid.

Proof Take a particular $\varphi \in \Phi$ and, for simplicity of notation, let $Q_t = P_{\varphi(t)}$.

Then one can write

$$\frac{1}{t}\int T(dQ_t - dQ_0) = 2\int T\frac{1}{t}[\sqrt{dQ_t} - \sqrt{dQ_0}]\frac{1}{2}[\sqrt{dQ_t} + \sqrt{dQ_0}]$$

$$= \int T(\frac{2}{t})(\sqrt{dQ_t} - \sqrt{dQ_0}]\sqrt{dQ_0}$$

$$+ \int T\frac{1}{t}(\sqrt{dQ_t} - \sqrt{dQ_0})^2.$$

To evaluate this last term, take densities with respect to $M_t = \frac{1}{2}(Q_t + Q_0)$. It becomes $J_t = \int T\frac{1}{t}(\varphi_t - \varphi_0)^2 dM_t$ with

$$\varphi_t = \sqrt{\frac{dQ_t}{dM_t}}, \quad \varphi_0 = \sqrt{\frac{dQ_0}{dM_t}}.$$

Then

$$
\begin{aligned}
J_t^2 &\leq \left(\int T^2 dM_t\right) \int \frac{1}{t^2}(\varphi_t - \varphi_0)^4 dM_t \\
&\leq b \int \frac{1}{t^2}(\varphi_t - \varphi_0)^4 dM_t.
\end{aligned}
$$

According to Lemma 5 Section 7 this term tends to zero.

The first term $2\int T\frac{1}{t}[\sqrt{dQ_t} - \sqrt{dQ_0})\sqrt{dQ_0}$ can be written in the form

$$
2\int T\frac{1}{t}(\sqrt{f_t} - 1)dQ_0
$$

with

$$
f_t = \frac{dQ_t}{dQ_0}.
$$

The result follows since $\frac{2}{t}(\sqrt{f_t} - 1)$ converges in quadratic mean to $D\varphi'(0)$.

It is not unreasonable to insist that $\int T^2 dP_\theta$ be bounded in some neighborhood of zero if one is interested in local minimax results. Then Lemma 2 is usable without the restriction $P_\theta << P_0$, $\int(\frac{dP_\theta}{dP_0})^2 dP_\theta < \infty$.

Lemma 1 or Lemma 2 show that if $\text{Var}_0 T$ is finite then

$$
|\Gamma v| \leq b\|Dv\|
$$

for some $b < \infty$ and all v in the linear span $\mathcal{C}$ of the tangent set $\mathcal{C}_0$. In particular $Dv = 0$ will imply $\Gamma v = 0$.

This implies a restriction on the possible quasi-derivatives of functions such as $\theta \rightsquigarrow \int T dP_\theta$.

Let us show that, under the conditions of Lemma 1, the existence of a $b < \infty$ such that $|\Gamma v| \leq b\|Dv\|$ implies the existence of an estimate T such that $\theta \rightsquigarrow \int T dP_\theta$ has quasi-derivative Γ.

To do this let $\mathcal{H}_0$ be the quotient space of $\mathcal{C}$ identifying u and v if $\|D(u - v)\| = 0$. Put in $\mathcal{H}_0$ the norm $\|Dv\|$ and complete it, obtaining a Hilbert space $\mathcal{H}_1$. The map D can be considered as a one to one map on $\mathcal{H}_0$ that sends $\mathcal{H}_0$ into $L_2(P_0)$ isometrically. It extends to $\mathcal{H}_1$ by continuity and then sends $\mathcal{H}_1$ onto a closed subspace of $L_2(P_0)$.

A linear map Γ satisfying $|\Gamma v| \leq b\|Dv\|$ can also be considered as defined on $\mathcal{H}_0$ and be extended to $\mathcal{H}_1$. Then it can be considered as a continuous linear functional on the image of $\mathcal{H}_1$ by D. Therefore it is given by some square integrable function T such that $\Gamma v = \int T(Dv)dP_\theta$.

The situation under the assumptions of Lemma 2 is somewhat similar: Given any $\epsilon > 0$ and Γ satisfying $|\Gamma v| \leq b\|Dv\|$, there is a Γ_1 such that $|(\Gamma - \Gamma_1)v| \leq \epsilon\|Dv\|$, there is a Γ_1 such that $|(\Gamma - \Gamma_1)v| \leq \epsilon\|Dv\|$ and such that $\Gamma_1 v = \int T_1(Dv)dP_0$ for an estimate T which satisfies the conditions of the lemma.

Indeed one can proceed in the same manner and obtain a P_0 square integrable T as above. Then one approximates T by a bounded T_1 such that $\int |T - T_1|^2 dP_0 < \epsilon$.

(It should perhaps be pointed out that the maps D of Lemma 1 and Lemma 2 are the same. This is because the conditions of Lemma 1 imply that $\theta \rightsquigarrow P_\theta$ is quasi-differentiable for the weak topology of $L_1(P_0)$ while Lemma 2 implies quasi-differentiability for the strong topology of L_1.) It should be clear that neither the conditions of Lemma 1 nor those of Lemma 2 can give upper bounds for the minimax risk on the set Θ itself or even on a neighborhood of zero in Θ.

One can reverse the inequalities of Section 6 to obtain conditions under which a function γ will admit unbiased estimation with finite minimax risk.

If the minimax risk is to be $\leq M$ then γ must satisfy the inequalities

$$|\langle \gamma, \mu \rangle - \langle \gamma, \nu \rangle|^2 \leq \int \frac{[dP_\mu - dP_{nu}]^2}{dP_\lambda}[M + \text{var}(\gamma|\lambda)]$$

for all pairs μ, ν of $\mathcal{M}$ and for $\lambda = \frac{1}{2}(\mu + \nu)$. Except for the occurrence of the term $\text{Var}(\gamma|\lambda)$ on the right side this would be a Lipschitz conditions for the distance

$$\left[\int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda}\right]^{1/2}.$$

If h^2 denotes the Hellinger distance, this give

$$|\langle \gamma, \mu \rangle - \langle \gamma, \nu \rangle|^2 \leq 4h^2(2 - h^2)[M + \text{Var}(\gamma|\lambda)]$$

with $h^2 = h^2(P_\mu, P_\nu) = \frac{1}{2}\int(\sqrt{dP_\mu} - \sqrt{dP_\nu})^2$. Thus γ must satisfy the same sort of modified Lipschitz condition for h.

These are Lipschitzian conditions on the P_μ for $\mu \in \mathcal{M}$.

One cannot omit the terms $\text{Var}(\gamma|\lambda)$ from the right side. To see this note that $0 \leq \int \frac{(dP_\mu - dP_\nu)^2}{dP_\lambda} \leq 4$

It is possible to show that these relations imply the inequalities $|\Gamma v| \leq b\|Dv\|$ under the conditions of Lemma 1 or Lemma 2. However the converse does not seem to be correct.

We shall end on a special subject suggested by the article of Simons and Woodroofe [1983].

Let us consider a sequence of experiments $\mathcal{E}_n = \{P_{\theta,n}; \theta \in \Theta_n\}$ and a corresponding sequence of real valued functions γ_n.

Suppose that μ_n is a probability measure on a σ-field $\mathcal{B}_n$ of subsets of Θ_n. Suppose that $\mathcal{B}_n$ is large enough to make all the relevant functions, such as γ_n, or $\theta \rightsquigarrow P_{\theta,n}(A)$ and so forth measurable.

Fix a particular point, say a_n in Θ_n. Then, under μ_n, the map

$$\theta \rightsquigarrow \frac{|\gamma_n(\theta) - \gamma_n(a_n)|^2}{4k^2(P_{\theta,n}, P_{a_n,n})}$$

is a certain random variable. Let F_n be its distribution under μ_n. Take estimates T_n such that $\int T_n dP_{\theta,n} = \gamma_n(\theta)$ almost everywhere μ_n. Then one can assert the following.

Lemma 3 *In the system just described assume that under μ_n the variables $\frac{1}{2}[\text{Var}_\theta T_n + \text{Var}_{a_n} T_n]$ tend to a constant σ^2.*

Then any cluster point F of the sequence $\{F_n\}$ has its support contained in $[0, \sigma^2]$.

This is an immediate consequence of the inequalities of Lemma 1, Section 7. It is in the same general spirit as Lemma 2.

A lemma of this kind was used by Simons and Woodroofe [1983] for the case where Θ_n is an interval of the line $\Theta_n = (a_n - \epsilon_n, a_n + \epsilon_n)$ with $\epsilon_n > 0$ tending to zero. The measure μ_n was Lebesgue measure on Θ_n normalized to be a probability measure there

Simons and Woodroofe use this to show that the Cramér-Rao inequality "holds almost everywhere".

For this they use conditions similar to those of Lemma 2 but with approximate derivatives instead of derivatives. Extensions to parameter spaces Θ that are infinite dimensional do not seem to have been formulated.

References

Barankin, E.W. (1949). Locally best unbiased estimates. *Ann. Math. Stat.* **20**, 477-501.

Berkson, J. and Hodges, J.L. Jr, (1962). Minimax estimation for the logistic function. *Fourth Berkeley Symposium, Vol. 4*, 77-86.

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. *Semi-parametrics* to appear.

Brown, L. and Gajek, L. (1990). Information inequalities for the Bayes risk. *Ann. Statist.* **18**, 1578-1594.

Brown, L. and Low, M. (1991). Information inequality bounds on the minimax risk (with an application to nonparametric regression). *Ann. Statist.* **19**, 329-337.

Chapman, D. and Robbins, H. (1951). Minimum variance estimation without regularity assumptions. *Ann. Math. Stat.* **22**, 581-586.

Dieudonné, J. (1960). *Foundations of Modern Analysis.* Academic Press.

Fabian, V. and Hannan, J. (1977). On the Cramér-Rao inequality. *Ann. Statist.* **5**, 197-205.

Halmos, P. and Savage, L.J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.* **20**, 225-241.

Kholevo, A.S. (1973). A generalization of the Rao-Cramér inequality. *Theory of Probab. Appl.* Vol 18 #2, 359-362.

Kiefer, J.C. (1952). On minimum variance estimators. *Ann. Math. Stat.* **23**, 627-629.

Le Cam, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Stat.* **35**, 1419-1455.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag.

Le Cam, L. and Yang, G.L. (1990). *Asymptotics in Statistics. Some Basic Concepts.* Springer-Verlag.

Lehmann, E.L. and Scheffé, H. (1950). Completeness, similar regions and unbiased estimation. *Sankhýa* **10**, 305-340.

Pfanzagl, J. (1982). *Contributions to a general asymptotic statistical theory.* L.N.S. #13 Springer Verlag.

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

Simons, G. and Woodroofe, M. (1983). The Cramér-Rao inequality holds almost everywhere. *Recent Advances in Statistics* (Herman Chernoff Festschrift) pp. 69-93.

Wolfowitz, V. (1947). The efficiency of sequential estimates and Wald's equation for sequential processes. *Ann. Math. Statist.* **18**, 215-230.