EFFICIENT ESTIMATION UNDER CONSTRAINTS*

CHRISTIAN HIPP

It is well known that in full families of probability measures, the empirical distribution is an efficient estimator for the unknown distribution. When the family is restricted by some finite dimensional linear constraint, then the minimum Pearson distance (MPD) estimator is efficient (Hipp (1988)). Here, for the case of a finite dimensional nonlinear constraint, four methods are given for the construction of efficient estimators. One of these is the MPD method. The simplest method is an approximation of the MPD estimator.

1. INTRODUCTION AND SUMMARY

Let $\mathscr{P}$ be a family of probability measures on d-dimensional Euclidean space $\mathbb{R}^d$, $H: \mathscr{P} \to \mathbb{R}$ a functional, and $X_1, X_2, \ldots$ a sequence of iid random variables with unknown distribution $P \in \mathscr{P}$. We are interested in the construction of estimators $H_n(X_1, \ldots, X_n)$ for $H(P)$ which are efficient in the sense of Hájek and Inagaki's convolution theorem. The family $\mathscr{P}$ will be nonparametric and be given by some finite dimensional constraint.

1.1 Example. $\mathscr{P}$ is the set of all P on the real line with finite variance $\sigma^2(P)$ and with fixed known mean $\mu(P) = \mu_0$, and $H(P) = \sigma^2(P)$. Here,

$$H_n(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i^2(1 - (X_i - \bar{X})(\bar{X} - \mu_0)/s^2) - \mu_0^2$$

with

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad \text{and} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is efficient, its asymptotic variance equals

$$\beta_4 - \sigma^4 - \beta_3^2/\sigma^{-2}$$

* Presented at the "Kolloquium über Mathematische Statistik im Rahmen der Wissenschaftlichen Kolloquien der Universität Hamburg und der Karls-Universität Prag", Hamburg, June 1989.

with $\beta_k = \int (X - \mu(P))^k \, P(\mathrm{d}x)$. Notice that this asymptotic variance is smaller than the asymptotic variance of s^2 whenever $\beta_3 \neq 0$.

1.2 Example. $\mathscr{P}$ is the set of all P with fixed known variance $\sigma^2(P) = \sigma_0^2$, and $H(P) = \mu(P)$. In this case,

$$H_n(X_1, \ldots, X_n) = \bar{X} - \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^3 \, (s^2 - \sigma_0^2)/N$$

with

$$N = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^4 - s^4$$

is efficient with asymptotic variance

$$\sigma_0^2 - \beta_3^2/(\beta_4 - \sigma_0^4)$$

which is smaller than the asymptotic variance of $\bar{X}$ whenever $\beta_3 \neq 0$.

In Example 1.1 we have a linear constraint, in Example 1.2 the constraint is nonlinear. For the case of a linear constraint, the following method yields efficient estimators (see [2]). Let $f = \mathbb{R}^d \to \mathbb{R}^r$, $a \in \mathbb{R}^r$, and $\mathscr{P}$ the family of probability measures with $\int |f(x)|^2 \, P(\mathrm{d}x) < \infty$, $\int f(x) f(x)^{\mathrm{T}} \, P(\mathrm{d}x)$ nonsingular, and $\int f(x) \, P(\mathrm{d}x) = a$. Let P_n be the discrete probability measure with point probabilities

$$p_n(x) = i(x) \left(1 - (f(x) - \bar{f})^{\mathrm{T}} M^{-1} (\bar{f} - a)\right)$$

with $\bar{f} = n^{-1} \sum f(X_i)$, $i(x) = \#\{j \leqq n : X_j = x\}/n$, and

$$M = n^{-1} \sum_{i=1}^{n} (f(X_i) - \bar{f}) (f(X_i) - \bar{f})^{\mathrm{T}} .$$

If M is singular or if in this definition $p_n(x) < 0$ for some $x \in \mathbb{R}^d$, then let P_n be some fixed known element of $\mathscr{P}$. Then for any smooth functional $H: \mathscr{P} \to \mathbb{R}$, $H(P_n)$ is an efficient estimator for $H(P)$. Efficiency is meant in the sense Hájek and Inagaki's convolution theorem (see [3], p. 158, Theorem 9.3.1). Smoothness of H is compact or Hadamard differentiability with respect to a weak topology. For heavy tailed functionals H such as the mean or the variance, it is conveniet to use a weak metric which implies convergence of some moments. One possible choice would be the Mallows metric of some fixed order $k \geqq 1$:

$$d(P, Q) = \inf \left(\mathsf{E}|X - Y|^k\right)^{1/k}$$

where the infimum is taken over all pairs of random variables (X, Y) with marginal distributions P and Q, and P, Q are in the set $\mathscr{Q}$ of all probability measures with finite mean of order k. We prefer to work with a different metric. Let F be non-negative and continuous on $\mathbb{R}^d$ and $\mathscr{Q}$ be the set of all probability measures P satisfying $\int F(x) \, P(\mathrm{d}x) < \infty$. Let $\mathscr{F}$ be the set of all functions $1_C(x)$, $1_C(x) \, F(x)$, C a set of the form $\{x \leqq a\}$ $e \in \mathbb{R}^d$, and for $P, Q \in \mathscr{Q}$ write

$$d(P, Q) = \sup \left|\int f \, \mathrm{d}(P - Q)\right|$$

where the supremum is taken over all $f \in \mathscr{F}$. We have $d(P_n, P) \to 0$ iff $P_n \to P$ in the Kolmogorov metric and

$$\int F(x)\, P_n(\mathrm{d}x) \to \int F(x)\, P(\mathrm{d}x)\,.$$

Write B for the linear space generated by $\mathscr{Q}$, endowed with the topology induced by $\mathscr{F}$, i.e. for $\nu, \mu \in B$

$$d(\mu, \nu) = \sup \left|\int f \,\mathrm{d}(\mu - \nu)\right|\,.$$

Fix $P_0 \in \mathscr{Q}$. A functional $H\colon \mathscr{Q} \to \mathbb{R}$ is compact differentiable at P_0 if there exists $g\colon \mathbb{R}^d \to \mathbb{R}$ with $\int g(x)\, P_0(\mathrm{d}x) = 0$ with the following property: For any family P_t, $t \in (0, 1]$, in $\mathscr{Q}$ for which $t^{-1}(P_t - P_0)$ converges in B, we have

$$H(P_t) = H(P_0) + \int g(x)\,(P_t - P_0)\,(\mathrm{d}x) + o(t)\,.$$

This implies in particular that $\int |g(x)|\, P_t(\mathrm{d}x) < \infty$ for t sufficiently small. The function g is called the derivative of H at P_0 and is denoted by $H'(P_0)$.

1.3 Example. For $F(x) = x^2$ the functional $H(P) = \sigma(P)$ is differentiable at any P with $\sigma(P) > 0$, and

$$H'(P) = ((x - \mu(P))^2 - \sigma^2(P))/(2\sigma(P))\,.$$

1.4 Example. For $w \in \mathbb{R}$ the functional $H(P) = P^{*2}(-\infty, w]$ is differentiable at any P with

$$H'(P)\,(x) = 2(P(-\infty, w - x] - P^{*2}(-\infty, w])\,.$$

In the following we fix F and consider a subfamily $\mathscr{P}$ of $\mathscr{Q}$ which is defined by some smooth constraint $S = (S_1, \ldots, S_r)\colon \mathscr{Q} \to \mathbb{R}^r$:

$$\mathscr{P} = \{P \in \mathscr{Q}\colon S(P) = a\}\,, \quad a \in \mathbb{R}^r\,.$$

The components of S are assumed to be compact differentiable at any $P \in \mathscr{Q}$, with derivatives $S_i'(P)$ satisfying

$$\int S_i'(P)\,(x)^2\, P(\mathrm{d}x) < \infty\,.$$

We consider smooth functionals $H\colon \mathscr{Q} \to \mathbb{R}$ which are compact differentiable at any $P \in \mathscr{P}$, with derivative $H'(P)$ satisfying

$$\int H'(P)\,(x)^2\, P(\mathrm{d}x) < \infty\,.$$

We first compute the lower bounds for the asymptotic variance of regular estimators.

1.5 Proposition. Fix $P_0 \in \mathscr{P}$ and assume that the covariance matrix

$$\Sigma = \int S'(P_0)\,(x)\, S'(P_0)\,(x)^{\mathrm{T}}\, P_0(\mathrm{d}x)$$

of $S'(P_0)$ is regular. Let

$$u_i = \int H'(P_0)\,(x)\, S_i'(P_0)\,(x)\, P_0(\mathrm{d}x)\,, \quad i = 1, \ldots, r$$

and

$$\sigma_0^2 = \int H'(P_0)\,(x)^2\, P_0(\mathrm{d}x) - u^{\mathrm{T}}\Sigma^{-1}u\,.$$

Let H_n be a sequence of estimators with the following property: There exists a probability measure K such that for $P_n \to P_0$ in modified Hellinger distance, $P_n \in \mathscr{P}$, the distribution of

$$\sqrt{(n)}\,(H_n - H(P_n))$$

under P_n^n converges weakly to K. Then there exists a probability measure R such that

$$K = \mathscr{N}(0, \sigma_0^2) * R\,.$$

(The modified Hellinger distance is

$$d_0^2(P_n, P_0) = \int F(x)\,(h_n^{1/2}(x) - h^{1/2}(x))^2\,\mu(\mathrm{d}x)$$

where h_n and h are μ-densities of P_0, respectively.)

If $K = \mathscr{N}(0, \sigma_1^2)$, then necessarily $\sigma_1^2 \geqq \sigma_0^2$.

The proposition is a consequence of a variant of the convolution theorem in Pfanzagl and Wefelmeyer [3], p. 158, Theorem 9.3.1. The details are given in Section 2.

We shall present four methods for the construction of efficient estimators H_n, i.e. for estimators for which the distribution of $\sqrt{(n)}\,(H_n - H(P_0))$ under P_0^n converge weakly to $\mathscr{N}(0, \sigma_0^2)$. The first method is based on the well known improvement procedure for inefficient estimators (see [3], p. 200). Let Q_n be the empirical distribution of $X_1, \ldots, X_n$ and consider estimators of the following kind:

$$G_n(X_1, \ldots, X_n) = H(Q_n) + c^{\mathrm{T}}(S(Q_n) - a)$$

where $c \in \mathbb{R}^r$ is a free vector parameter. This parameter is chosen such that the asymptotic variance of G_n under P_0^n is as small as possible. The optimal $c = c(P_0)$ depends on the unknown distribution P_0. If the map $P \to c(P)$ is smooth, then we obtain an efficient estimator by studentization:

$$H_n(X_1, \ldots, X_n) = H(Q_n) + c(Q_n)^{\mathrm{T}}\,(S(Q_n) - a)\,.$$

1.6 Proposition. Assume that for $P \in \mathscr{Q}$ the covariance matrix

$$\Sigma(P) = \int S'(P)\,(x)\,S'(P)\,(x)^{\mathrm{T}}\,P(\mathrm{d}x)$$

of $S'(P)$ exists, and let

$$u_i(P) = \int H'(P)\,(x)\,S_i'(P)\,(x)\,P(\mathrm{d}x)\,, \quad i = 1, \ldots, r$$

and

$$c(P) = \Sigma(P)^{-1}\,u(P)\,.$$

If $P \to c(P)$ is continuous in the metric d at P_0, and if $\int F(x)^2\,P_0(\mathrm{d}x) < \infty$, then

$$H_n(X_1, \ldots, X_n) = H(Q_n) + c(Q_n)^{\mathrm{T}}\,(S(Q_n) - a)$$

is efficient at P_0.

Proof. Since $\sqrt{(n)}\,(Q_n - P_0)$ is tight in the metric d (see Lemma 2.2), compact differentiability of H and S imply that H_n has the stochastic expansion

$$\sqrt{(n)}\,(H_n - H(P_0)) = n^{-1/2}\sum_{i=1}^{n} H'(P_0)(X_i) + c(Q_n)^{\mathrm{T}}\,n^{-1/2}\sum_{i=1}^{n} S'(P_0)(X_i) + R_n$$

with $R_n \to 0$ in probability. Continuity of c implies $c(Q_n) \to c(P_0)$ in probability and hence H_n has asymptotic variance

$$\mathrm{Var}\,(H'(P_0) - c(P_0)^{\mathrm{T}}\, S'(P_0)) = \sigma_0^2 . \qquad \square$$

The second method is based on the bootstrap. Here, computation of derivatives $H'(P)$ or $S'(P)$ is not necessary: The vector $c(P_0)$ is approximated by a bootstrap estimator c_*. This estimator is a function of bootstrapped covariances. For the convergence of these bootstrapped covariances we need uniform integrability of second moments for statistics

$$\sqrt{(n)}\,(H(Q_n) - H(P))$$

which is uniform for P in a neighborhood of P_0 in $\mathscr{Q}$. Here, $H\colon \mathscr{Q} \to \mathbb{R}$ is some smooth functional, and Q_n is the empirical distribution of n iid observations with distribution P. We impose the slightly stronger condition

$$\sup_P \mathrm{E}_P|\sqrt{(n)}\,(H(Q_n) - H(P))|^{2+\delta} < \infty$$

for some $\delta > 0$, where the supremum is taken over the neighborhood. This last condition is implied by the following condition in which $P_0 \in \mathscr{P}$ is fixed.

Condition A. (i) H is compact differentiable at any $P \in \mathscr{P}$; (ii) There exists $\delta > 0$ and $g\colon \mathbb{R}^d \to \mathbb{R}$ such that

(a) $\sup\{\int g(x)^{2+\delta}\, P(\mathrm{d}x)\colon d(P_0, P) < \delta\} < \infty$;

(b) $P_1, P_2 \in \mathscr{Q}$, $d(P_1, P_2) < \delta$ implies

$$|H'(P_1)(x) - H'(P_2)(x)| \leqq g(x)\, d_2(P_1, P_2)\,, \quad x \in \mathbb{R}^d ;$$

and

(c) $\sup\{\int |H'(P)(x)|^{2+\delta}\, P(\mathrm{d}x)\colon d(P_0, P) < \delta\} < \infty$.

Here, d_2 is the Cramer-v. Mises distance

$$d_2^2(P_1, P_2) = \int (P_1(x, \infty) - P_2(x, \infty))^2\, \mathrm{d}x .$$

Actually, any metric d satisfying

$$\sup_{P,n} \mathrm{E}_P(\sqrt{(n)}\, d^2(Q_n, P))^s < \infty\,, \quad s > 0$$

will serve as well.

1.7 Proposition. Assume that the conditions of the last proposition are satisfied, and assume in addition that $\Sigma(P_0)$ is nonsingular and that $H, S_1, \ldots, S_r$ satisfy Condition A. Let Q_n^* be the empirical distribution of $Y_1, \ldots, Y_n$ which are iid with distribution Q_n, and write E_*, for the expectation with respect to $Y_1, \ldots, Y_n$, given $X_1, \ldots, X_n$. Write

$$u_* = n\, \mathrm{E}_*(H(Q_n^*) - H(Q_n))\,(S(Q_n^*) - S(Q_n))$$

and

$$\Sigma_* = n\, \mathrm{E}_*(S(Q_n^*) - S(Q_n))\,(S(Q_n^*) - S(Q_n))^{\mathrm{T}}$$

and

$$c_* = \Sigma_*^{-1} u_* .$$

For singular Σ_*define c_* as zero. Then

$$H_n(X_1, \ldots, X_n) = H(Q_n) + c_*^{\mathrm{T}}(S(Q_n) - a)$$

is efficient at P_0.

The proof of this proposition will be given in Section 2. □

There are also representations of efficient estimators of the form $H(\hat{Q}_n)$, where $\hat{Q}_n$ is some projection of Q_n onto $\mathscr{P}$. One of these projections is the minimum discriminant information adjusted distribution which is the distribution in $\mathscr{P}$ minimizing the Kullback-Leibler distance to Q_n. For linear constraints, the corresponding estimators have been considered by Habermann [1] and Sheehy [5].

Here we shall deal with the Pearson distance which is defined as

$$\varrho^2(P, Q) = \int (1 - \mathrm{d}Q/\mathrm{d}P)^2 \,\mathrm{d}P$$

if Q is absolutely continuous w.r.t. P, and infinity elsewhere. If the constraint is linear, then the distribution P_n can be considered as the one in $\mathscr{P}$ that minimizes the Pearson distance to Q_n. A similar statement is true also under nonlinear constraints.

Let $\hat{P}_n$ be a distribution in $\mathscr{P}$ that minimizes the distance $\varrho(Q_n, P)$ among all $P \in \mathscr{P}$. We shall assume that $\hat{P}_n$ can be chosen such that $(X_1, \ldots, X_n) \to \hat{P}_n$ is measurable. For finite subsets A of $\mathbb{R}^d$ write

$$\mathscr{E}(A) = \{P \in \mathscr{P}: P(A) = 1\} .$$

A reasonable minimum will be obtained only if $\mathscr{E}(\{X_1, \ldots, X_n\}) \neq \emptyset$. We have to assume that this happens with high probability when n is large. Our condition will be even more restrictive.

1.8 Proposition. Fix $P_0 \in \mathscr{P}$ with $\int F(x)^2 P_0(\mathrm{d}x) < \infty$, and let $H: \mathscr{Q} \to \mathbb{R}$ be compact differentiable at P_0. Assume that the components S_i of $S = (S_1, \ldots, S_r)$ satisfy the following

Condition B. (i) S_i is compact differentiable at any $P \in \mathscr{Q}$;
(ii) The exists $\delta > 0$ and $g: \mathbb{R}^d \to \mathbb{R}$ such that
(a) $\int g(x) P_0(\mathrm{d}x) < \infty$ and
(b) $P_1, P_2 \in \mathscr{Q}$, $d(P_1, P_2) < \delta$ implies

$$|S_i'(P_1)(x) - S_i'(P_2)(x)| \leqq g(x)\, \mathrm{d}_2(P_1, P_2), \quad x \in \mathbb{R}^d .$$

Assume, finally, that $\sqrt{(n)}\,(\hat{P}_n - P_0)$ is tight and that

$$\lim_n P_0^n\{\hat{P}_n\{X_i\} > 0, \quad i = 1, \ldots, n\} = 1 .$$

Then $H(\hat{P}_n)$ is efficient at P_0.

The proof for this proposition will be given in Section 2. □

1.9 Remark. If S is linear, i.e. $S(P) = \int f(x)\, P(\mathrm{d}x)$ with $f\colon \mathbb{R}^d \to \mathbb{R}^r$, and if $\int \|f\|^2 \,\mathrm{d}P_0 < \infty$, then all assumptions of Proposition 1.8 concerning S are satisfied, and

$$\lim_n P_0^n\{\hat{P}_n = P_n\} = 1\,. \tag{1.10}$$

Proof. Condition B is obviously satisfied with $g = 0$. We first show that $\sqrt{(n)}\,.\,(P_n - P_0)$ is tight. Notice first that the processes

$$\mathrm{Cov}_{P_0}(h, f)^{\mathrm{T}}\, \Sigma^{-1} \sqrt{(n)}\,(\bar{f} - a)\,, \quad h \in \mathscr{F}$$

converge weakly to the process

$$\mathrm{Cov}_{P_0}(h, f)^{\mathrm{T}}\, \Sigma^{-1} Y$$

where Y is r-variate normal with zero mean and covariance matrix Σ. It is sufficient to show that

$$\sup_{h\in\mathscr{F}} \|n^{-1} \sum_{i=1}^{n} h(X_i)\,(f(X_i) - \bar{f})^{\mathrm{T}}\, M^{-1} - \mathrm{Cov}_{P_0}(h, f)^{\mathrm{T}}\, \Sigma^{-1}\|$$

converges to zero in probability. We know that

$$\sup_{h\in\mathscr{F}} \|\sum_{i=1}^{n} h(X_i)\,(f(X_i) - \bar{f})\|$$

is bounded in probability (bip). Hence $M^{-1} \to \Sigma^{-1}$ a.e. implies

$$\sup_{h\in\mathscr{F}} \|n^{-1} \sum_{i=1}^{n} h(X_i)\,(f(X_i) - \bar{f})^{\mathrm{T}}\,(M^{-1} - \Sigma^{-1})\| \to 0$$

in probability. We have to show that

$$\sup_{h\in\mathscr{F}} \|n^{-1} \sum_{i=1}^{n} h(X_i)\,(f(X_i) - \bar{f}) - \mathrm{Cov}_{P_0}(F, f)\| \to 0$$

in probability. Using that

$$\sup_{h\in\mathscr{F}} |n^{-1} \sum_{i=1}^{n} h((X_i)\|$$

is bip we can replace $\bar{f}$ by $\mu := \int f(x)\, P_0(\mathrm{d}x)$. Finally

$$\sup_{h\in\mathscr{F}} \|n^{-1} \sum_{i=1}^{n} h(X_i)\,(f(X_i) - \mu) - \mathrm{Cov}_{P_0}(F, f)\| \to 0$$

follows from the Glivenko-Cantelli theorem for the empirical process indexed by the functions $hf\colon h \in \mathscr{F}$ (see Sheehy and Wellner [6], Th. 1.1). We shall now prove (1.10). To simplify notations we assume $X_1, \ldots, X_n$ are distinct. Then the probability measure $\hat{P}_n$ can be identified with the n-vector $(p_1, \ldots, p_n)$ satisfying $p_i \geqq 0$, $\sum p_i = 1$, $\sum p_i f(x_i) = a$ which minimizes

$$\sum_{i=1}^{n} (1 - np_i)^2\,.$$

We shall forget for the moment the condition $p_i \geqq 0$, $i = 1, \ldots, n$, and find the solution with the method of Lagrange multipliers. We obtain that there exists $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^r$ such that

$$-2(1 - np_i) + \alpha + \beta^T f(X_i) = 0\,, \quad i = 1, \ldots, n$$

or, with other α and β

$$p_i = \beta^T f(X_i) + \alpha\,, \quad i = 1, \ldots, n\,.$$

The constraints $\sum p_i = 1$ and $\sum p_i f(x_i) = a$ yield the unique solution

$$p_i = P_n\{X_i\}\,, \quad i = 1, \ldots, n$$

whenever M is regular. Hence $\hat{P}_n = P_n$ on the set

$$A_n = \{M \text{ regular}, (f(X_i) - \bar{f})^T M^{-1}(\bar{f} - a) < 1\,, \quad i = 1, \ldots, n\}.$$

We finally show $\lim_n P_0^n(A_n) = 0$. Since $M \to \Sigma$ a.e. and Σ is regular, it suffices to prove that

$$\lim_n P_0^n\{\sup_{i \leqq n} (f(X_i) - \bar{f})^T M^{-1}(\bar{f} - a) < 1\} = 0\,.$$

Notice that $\bar{f}^T M^{-1}(\bar{f} - a) \to 0$ a.e., so it suffices to prove that

$$\sup_{i \leqq n} \|f(X_i)\|\,\|\bar{f} - a\| \to 0$$

in probability. Since $\sqrt{(n)}\,\|\bar{f} - a\|$ is bip, it is sufficient to show that

$$n^{-1/2} \sup_{i \leqq n} \|f(X_i)\| \to 0$$

in probability. This, however, follows from $\mathsf{E}_P\|f(X_1)\|^2 < \infty$:

$$P_0^n\{\sup_{i \leqq n} \|f(X_i)\| \geqq \varepsilon\sqrt{n}\} \leqq \varepsilon^{-2}\,\mathsf{E}_{P_0}\|f(X_1)\|^2\, 1_{\{\|f(X_1)\| \geqq \varepsilon\sqrt{n}\}} \to 0\,. \qquad \square$$

Finally, there is a simple approximation $\tilde{P}_n$ of the minimum Pearson distance estimator which is efficient. It has point probabilities

$$\tilde{P}_n\{x\} = i(x)\,(1 - S'(Q_n)\,(x)^T M^{-1}(S(Q_n) - a))\,.$$

Notice that $\tilde{P}_n$ is no longer an element of $\mathscr{P}$ in general.

1.10 Proposition. Assume that S is compact differentiable at any $Q \in \mathscr{Q}$ and that the components $S_i'(P)$ satisfy the following condition:

There exist $\delta > 0$ and $g \geqq 0$ with $\int g^2(x)\,P_0(dx) < \infty$ such that for all $P \in \mathscr{Q}$ with $d(P, P_0) < \delta$ and $x \in \mathbb{R}^d$

$$|S_i'(P)\,(x) - S_i'(P_0)\,(x)| \leqq g(x)\,d(P, P_0)\,.$$

Then $\tilde{P}_n$ is efficient at P_0.

Proof. Let H be a smooth functional. We have to show that $H(\tilde{P}_n)$ is efficient for $H(P)$ at P_0. Tightness of $\sqrt{(n)}\,(\tilde{P}_n - P_0)$ implies that $\sqrt{(n)}\,(H(\tilde{P}_n) - H(P_0))$ and

$$\sqrt{(n)} \int H'(P_0)\,(x)\,(\tilde{P}_n - P_0)\,(dx)$$

are asymptotically equivalent (in the sense that the difference converges to zero in probability). The last expression equals

$$\sqrt{(n)} \int H'(P_0)(x)\, Q_n(\mathrm{d}x) -$$

$$- n^{-1} \sum_{i=1}^{n} H'(P_0)(X_i)\, S'(Q_n)(X_i)^{\mathrm{T}} \sqrt{(n)}\,(S(Q_n) - a)\,.$$

It suffices to show that this is asymptotically equivalent to

$$\sqrt{(n)}\,(H(Q_n) - H(P_0)) - \sqrt{(n)}\, c(P_0)\,(S(Q_n) - a)$$

(see the proof of Proposition 1.6). This follows from

$$\lim_n P_0^n\{d(Q_n, P_0) > \delta\} = 0$$

and

$$\left\| n^{-1} \sum_{i=1}^{n} H'(P_0)(X_i)\,(S'(Q_n)(X_i) - S'(P_0)(X_i)\right\| \leqq$$

$$r n^{-1} \sum_{i=1}^{n} |H'(P_0)(X_i)|\, g(X_i)\, d(Q_n, P_0)$$

which converges to zero in probability. □

2. LEMMAS AND PROOFS

In this section we fix F and write $\mathscr{Q}$ for the set of all probability measures with $\int F(x)\, P(\mathrm{d}x)$ finite. Let d be the metric induced by $\mathscr{F}$ and d_1 the Kolmogorov metric:

$$d_1(P, Q) = \sup |P(-\infty, x) - Q(-\infty, x)|\,, \text{ where the sup is over all } x \in \mathbb{R}^d.$$

2.1 Lemma. We have $d(P_n, P) \to 0$ iff $\mathrm{d}_1(P_n, P) \to 0$ and

$$\int F(x)\, P_n(\mathrm{d}x) \to \int F(x)\, P(\mathrm{d}x)\,.$$

Proof. $\Rightarrow$ is obvious.

$\Leftarrow$: By assumption, $F(x)$ is uniformly integrable. Hence for arbitrary $\varepsilon > 0$ we can find $M > 0$ such that for all n

$$\int F(x)\, 1_{\{F(x) > M\}}(x)\, P_n(\mathrm{d}x) < \varepsilon\,.$$

Then

$$d(P_n, P) \leqq \sup_C \left|\int_C \min(F(x), M)\,(P_n - P)(\mathrm{d}x)\right| + \varepsilon\,.$$

The first term on the right hand side converges to zero since F is continuous. □

2.2 Lemma. Let P satisfy $\int F(x)^2\, P(\mathrm{d}x) < \infty$, write B for the linear space generated by $\mathscr{Q}$, equipped with the topology induced by $\mathscr{F}$. Let $X_1, X_2, \ldots$ be iid with distribution P, and Q_n be the empirical distribution of $X_1, \ldots, X_n$. Then $\sqrt{(n)}\,(Q_n - P)$ is tight in B.

Proof. Follows from Pollard's CLT for the empirical process (cf. [4]). □

Proof of 1.5. One can easily rewrite the convolution theorem in [3], p. 158, for the modified Hellinger distance d_0. Clearly H is differentiable at P_0 with respect to d_0. Let $T(\mathscr{P}, P_0)$ be the tangent space of $\mathscr{P}$ at P_0, i.e. the set of all g with $\int g(x)\, P_0(\mathrm{d}x) = 0$, $\int F(x)\, |g(x)|\, P_0(dx) < \infty$, and for some family P_t, $t \in (0, 1]$, in $\mathscr{P}$ we have

$$\int F(x)\, (f_t^{1/2} - 1 - \tfrac{1}{2}tg)^2\, \mathrm{d}P_0 = o(t^2)\,, \quad t \to 0\,.$$

Here f_t is a P_0-density of P_t.

With $\mathscr{L} = \{g\colon \int g(x)\, P_0(\mathrm{d}x) = 0,\ \int F(x)\, |g(x)|\, P_0(\mathrm{d}x) < \infty\}$ we claim that

$$T(\mathscr{P}, P_0) \supseteqq \{g \in \mathscr{L}\colon \int g(x)\, S'(P_0)\,(x)\, P_0(\mathrm{d}x) = 0\}\,. \tag{2.3}$$

If (2.3) holds, then a lower bound σ_0^2 for the asymptotic variance of regular estimators can be computed as the squared length of the projection of $H'(P_0)$ onto the set on the r.h.s. of (2.3). This projection equals

$$H'(P_0) - \mathrm{Cov}\,(H'(P_0), S'(P_0))^{\mathrm{T}}\, \Sigma^{-1}\, S'(P_0)$$

which has squared length $\mathrm{Var}_{P_0}\,(H'(P_0)) - u^{\mathrm{T}}\Sigma^{-1}u$.

To prove (2.3) we first note that $T(\mathscr{P}, P_0)$ is d_0-closed. Hence it suffices to show that $T(\mathscr{P}, P_0)$ contains all bounded g with compact support satisfying $\int g(x)\, P_0(\mathrm{d}x) = \int g(x)\, S'(P_0)\,(x)\, P_0(\mathrm{d}x) = 0$. Choose a bounded function h with compact support satisfying $\int h(x)\, P_0(\mathrm{d}x) = 0$ and

$$\int h(x)\, S'(P_0)\,(x)\, P_0(\mathrm{d}x) \neq 0\,.$$

For $s, t \in (-\varepsilon, \varepsilon)$ let $P_{s,t}$ be the probability measure with P_0-density $x \to 1 + t\,g(x) + s\,h(x)$. We can always choose ε small enough such that all $P_{s,t}$ are nonnegative. The map

$$f(x, t) = S(P_{s,t}) - a$$

is differentiable with $S(0, 0) = 0$ and

$$(\partial/\partial s)\, f(s, t)|_{s=t=0} = \int h(x)\, S'(P_0)\,(x)\, P_0(\mathrm{d}x) \neq 0\,.$$

By the implicit function theorem, for $t \in (-\varepsilon', \varepsilon')$, $0 < \varepsilon' < \varepsilon$ suitably chosen, there exists $s(t) \in (-\varepsilon, \varepsilon)$ for which

$$f(s(t), t) = 0$$

and $s(t) \to 0$, $t \to 0$. We obtain

$$0 = f(s(t), t) = s(t) \int h(x)\, S'(P_0)\,(x)\, P_0(\mathrm{d}x) + o(t + s(t))$$

which implies $s(t) = o(t)$. Let $P_t = P_{t,s(t)}$. Then P_t has P_0-density

$$f_t(x) = 1 + t\,g(x) + s(t)\,h(x)$$

and hence

$$\int F(x)\, (f_t^{1/2}(x) - 1 - \tfrac{1}{2}t\,g(x))^2\, P_0(\mathrm{d}x) =$$
$$= \int F(x)\, (1 + \tfrac{1}{2}t\,g(x) + \tfrac{1}{2}\,s(t)\,h(x) - 1 - \tfrac{1}{2}t\,g(x))^2\, P_0(\mathrm{d}x) + o(t^2) = o(t^2).$$

This concludes the proof of 1.5. □

Proof of 1.7.

We shall deal with the numerator only. The denominator can be dealt with similarly. From Sheehy and Wellner [6] we obtain that

$$Z_n^* := n(S(Q_n^*) - S(Q_n))(S(Q_n^*) - S(Q_n))^{\mathrm{T}}$$

converges a.e. to YY^{T} where Y is $\mathcal{N}(0, \Sigma(P_0))$. To prove that $\mathsf{E}_* Z_n^*$ converges to $\Sigma(P_0)$ weakly it suffices to show that for all $\varepsilon > 0$

$$\lim_M \sup_n P_0^n\{\mathsf{E}_*|Z_n^*| \, 1_{\{|Z_n| > M\}} > \varepsilon\} = 0 .$$

To simplify notation we assume that S is univariate, i.e. $r = 1$. Let $\delta > 0$ be the quantity in Condition A. Since for fixed n

$$\lim_M P_0^n\{\mathsf{E}_*|Z_n^*| \, 1_{\{|Z_n| > M\}} > \varepsilon\} = 0$$

and

$$\lim_n P_0^n\{d(Q_n, P_0) > \delta\} = 0$$

it is sufficient to show that

$$\sup\{\mathsf{E}_P|\sqrt{(n)}(S(Q_n) - S(P))|^{2+\delta/2} : d(P, P_0) \leqq \delta, n \in \mathbb{N}\} < \infty .$$

There exists P_1 on the line connecting Q_n and P such that

$$\sqrt{(n)}(S(Q_n) - S(P)) = \sqrt{(n)} \int S'(P_1)(x)(Q_n - P)(\mathrm{d}x) .$$

If we had P instead of P_1, our assertion would follow from (c) in Condition A. So we are left with the problem to give a bound for

$$\mathsf{E}_P \int (\sqrt{(n)} |S'(P_1)(x) - S'(P)(x)|)^{2+\delta/2} (Q_n + P)(\mathrm{d}x)| .$$

We first consider integration with respect to P. From our Condition A, (b), we obtain the upper bound (notice that $d_2(P_1, P) \leqq d_2(Q_n, P)$)

$$\int g(x)^{2+\delta/2} P(\mathrm{d}x) \, \mathsf{E}_P[\sqrt{(n)} \, d_2(Q_n, P)]^{2+\delta/2} =: I .$$

We know that for all $s > 0$

$$\sup_P \mathsf{E}_P[\sqrt{(n)} \, d_2(Q_n, P)]^s < \infty \tag{2.4}$$

which, together with Condition A, (a), implies local boundedness of I.

We now consider integration with respect to Q_n. With

$$p = (2 + \delta)/(2 + \delta/2)$$

and $q = 1/(p - 1)$ we obtain by Hölder's inequality

$$\mathsf{E}_P \int g(x)^{2+\delta/2} d(Q_n, P) \, Q_n(\mathrm{d}x) \leqq (\mathsf{E}_P[n^{-1} \sum_1^n g(x_i)]^{2+\delta})^{1/p} (\mathsf{E}_P[d(Q_n, P)]^q)^{1/q}$$

which is locally bounded because of (2.4), the inequality

$$\mathsf{E}_P[n^{-1} \sum_1^n g(X_i)]^{2+\delta} \leqq C_{2+\delta} \int g(x)^{2+\delta} P(\mathrm{d}x)$$

and because of (a) in Condition A. □

Proof of 1.8.

Let P_n be the minimum Pearson distance estimator for P_0 under the constraint $\int S'(P_0)(x)\,P(dx) = 0$, i.e. with

$$M = \sum_{i=1}^{n} (S'(P_0)(X_i) - \overline{S'(P_0)})\,(S'(P_0)(X_i) - \overline{S'(P_0)})^{\mathrm{T}}$$

and

$$\overline{S'(P_0)} = n^{-1} \sum_{i=1}^{n} S'(P_0)(X_i)$$

and

$$p(x) = i(x)\,(1 - (S'(P_0)(x) - \overline{S'(P_0)}))^{\mathrm{T}} M^{-1}\, \overline{S'(P_0)})$$

we obtain on $\{M \text{ regular}, p(x) \geqq 0 \text{ for all } x \in \mathbb{R}^d\}$ the identity

$$P_n\{x\} = p(x)\,, \quad x \in \mathbb{R}^d\,.$$

We shall now derive a similar representation for $\hat{P}_n$. To simplify the notation we shall only treat the case of distinct $X_1, \ldots, X_n$. Let

$$D = \{(p_1, \ldots, p_n) \in \mathbb{R}^n \colon p_i \geqq 0\,, \ \sum p_i = 1\}$$

and

$$G\colon D \to \mathbb{R}^r\colon p = (p_1, \ldots, p_n) \to S(P(p))$$

where $P(p)$ is the probability measure giving mass p_i to X_i, $i = 1, \ldots, n$. The computation of $\hat{P}_n$ is equivalent to the solution to the following problem: Find $p \in D$ with $G(p) = a$ such that

$$\sum_{1}^{n} (1 - np_i)^2$$

is minimized. The map G is differentiable with partial derivatives

$$(\partial G/\partial p_i)\,(p) = S'(P(p))\,(X_i)\,, \quad i = 1, \ldots, n\,.$$

By assumption, there exists a sequence of sets A_n with $\lim P_0^n(A_n) = 0$ such that outside A_n we have the minimum attained in the interior of D, i.e. the solution $\hat{p}$ satisfies $\hat{p}_i > 0$, $i = 1, \ldots, n$. Hence the method of Lagrange multipliers works. To simplify the notation we denote

$$t(x) = S'(\hat{P}_n)\,(x)$$

$$\bar{t} \ = n^{-1} \sum t(X_i)$$

$$K \ = n^{-1} \sum_{i=1}^{n} S'(P_0)\,(X_i)\,(t(X_i) - \bar{t})^{\mathrm{T}}\,.$$

There exist $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^r$ such that

$$-\,(1 - n\,\hat{P}_n\{X_i\}) + \alpha + \beta^{\mathrm{T}}\, t(X_i) = 0\,, \quad i = 1, \ldots, n\,.$$

We shall show that for any measurable solution $\hat{P}_n$ of this equation we have

$$\sqrt{(n)}\,(H(\hat{P}_n) - H(P_n)) \to 0$$

in probability. Then efficiency of $H(\hat{P}_n)$ follows along the lines given in Remark 1.9.

The constraint $\sum \hat{p}_i = 1$ yields

$$\hat{p}_i = n^{-1}(1 - \beta^{\mathrm{T}}(t(X_i) - \bar{t})), \quad i = 1, \dots, n .$$

Let $R_n = \int S'(P_0)\, \hat{P}_n(\mathrm{d}x)$. Then

$$R_n = \overline{S'(P_0)} - K\beta .$$

Using Condition B we obtain $K \to \Sigma$ in probability, and since Σ is regular, K will be regular, too, for large n with high probability. Let us assume that K is regular in the following. Then

$$-\beta = K^{-1}(R_n - \overline{S'(P_0)}),$$

and so

$$\hat{p}_i = n^{-1}(1 - (t(X_i) - \bar{t})^{\mathrm{T}} K^{-1}(\overline{S(P_0)} - R_n)), \quad i = 1, \dots, n .$$

Tightness of $\sqrt{(n)}\,(\hat{P}_n - P_0)$ and differentiability of S imply that $\sqrt{(n)}\, R_n \to 0$ in probability:

$$0 = S(\hat{P}_n) - S(P_0) = \int S'(P_0)(x)\,(\hat{P}_n - P_0)(\mathrm{d}x) + o(n^{-1/2}) .$$

Write $a_n \doteq b_n$ if $a_n - b_n \to 0$ in probability. We have

$$\sqrt{(n)}\,(H(\hat{P}_n) - H(P_n)) \doteq \sqrt{(n)} \int H'(P_0)(x)\,(\hat{P}_n - P_n)(\mathrm{d}x) =$$

$$= \sqrt{(n)} \sum_{i=1}^{n} H'(P_0)(X_i)\,\{(t(X_i) - \bar{t})\, K^{-1}(\overline{S'(P_0)} - R_n) -$$

$$- (S'(P_0)(X_i) - \overline{S'(P_0)})\, M^{-1}\, \overline{S'(P_0)}\} := I_1 .$$

Since $|n^{-1} \sum_{i=1}^{n} H'(P_0)(X_i)\,(S'(P_0)(X_i) - \overline{S'(P_0)})\, M^{-1}|$ is bip, we obtain

$$I_1 \doteq n^{-1/2} \sum_{i=1}^{n} H'(P_0)(X_i)$$

$$\{(t(X_i) - \bar{t})\, K^{-1} - (S'(P_0)(X_i) - \overline{S'(P_0)})\, M^{-1}\}\, \{\overline{S'(P_0)} - R_n\}$$

$$=: I_2 .$$

Since

$$n^{-1/2} \sum_{i=1}^{n} H'(P_0)(X_i)\,(t(X_i) - \bar{t})\, \{\overline{S'(P_0)} - R_n\}$$

is bip and $K - M \to 0$, we obtain

$$I_2 \doteq n^{-1/2} \sum_{i=1}^{n} H'(P_0)(X_i)$$

$$\{(t(X_i) - \bar{t}) - (S'(P_0)(X_i) - \overline{S'(P_0)})\}\, M^{-1}(\overline{S'(P_0)} - R_n)$$

$$=: I_3 .$$

Finally, $\sqrt{(n)}\, M^{-1}(\overline{S'(P_0)} - R_n)$ is bip and

$$|n^{-1} \sum_{i=1}^{n} H'(P_0)(X_i) \{(t(X_i) - S'(P_0)(X_i)\}| \leqq$$

$$n^{-1} \sum_{i=1}^{n} |H'(P_0)(X_i)| \, g(X_i) \, d_2(\hat{P}_n, P_0)$$

where we used Condition B, ii (b) appropriately. Therefore

$$I_3 \doteq 0 .$$

This proves Proposition 1.8. □

(Received December 12, 1989.)

REFERENCES

[1] S. J. Haberman: Adjustment by minimum discriminant information. Ann. Statist. *12* (1984), 971—988.

[2] C. Hipp: Efficient estimators for ruin probabilities. In: Proc. 4th Prague Symp. Asympt. Statist. (P. Mandl, M. Hušková, eds.), Charles University, Prague 1989, pp. 259—268.

[3] J. Pfanzagl and W. Wefelmeyer: Contributions to a General Asymptotic Statistical Theory. (Lecture Notes in Statistics 13.) Springer-Verlag, Berlin—Heidelberg—New Nork 1982.

[4] D. Pollard: A central limit theorem for empirical processes. J. Austral. Math. Soc. *A33*, (1982), 235—248.

[5] A. Sheehy: Kullback-Leibler constrained estimation of probability measures. Preprint, Dept. of Statist., Stanford University 1988.

[6] A. Sheehy and J. A. Wellner: Uniformity in P of Some Limit Theorems for Empirical Measures and Processes. Tech. Rep. 134, Dept. Statist., University of Washington, Seattle, WA 1988.

Prof. Dr. Christian Hipp, Institut für Mathematische Stochastik, Universität Hamburg, Bundesstrasse 55, D-2000 Hamburg 13. Federal Republic of Germany.