Sankhyā : The Indian Journal of Statistics
1975, Volume 37, Series A, Pt. 4, pp. 502–513.

INFERENCE ABOUT SEPARATED FAMILIES IN LARGE SAMPLES*

By J. K. GHOSH
Indian Statistical Institute
and
K. SUBRAMANYAM
University of Dar es Salaam

SUMMARY. A separated parametric space is defined and properties of the m.l.e. (maximum likelihood estimator) studied. It is shown that the m.l.e. is asymptotically minimax but not in general asymptotically admissible. The main result of Kraft and Puri (1974) is obtained as a corollary.

1. Introduction

Cox (1962) has pointed out that in some problems the model consists of two separated families of densities $f(\cdot,\ \theta,\ \eta)$, $\eta \in \Omega_\theta$, $\theta = 0, 1$ and one is required to test H_0 $(\theta = 0)$ against H_1 $(\theta = 1)$. These families are assumed to be separated in the sense that no density $f(\cdot, 0, \eta)$ can be obtained as a limit of a sequence $\{f(\cdot, 1, \eta_i)\}$ and vice versa. Cox has not explained the sense in which this limit is to be taken but the following seems to be adequate for most purposes. Let $M = \{f(\cdot, \theta, \eta),\ \eta \in \Omega_\theta,\ \theta = 0, 1\}$ be thought of as a metric space with metric

$$d(f^{(1)}, f^{(2)}) = \int |f^{(1)} - f^{(2)}|\, d\mu$$

where μ is a σ-finite measure with respect to which the densities are taken. Then $f^{(n)} \to f$ if $d(f^{(n)}, f) \to 0$. Thus H_0 and H_1 are separated iff.

$$\inf_{\eta, \eta'} d(f(\cdot, 0, \eta),\ f(\cdot, 1, \eta')) > 0.$$

An example given by Cox involves deciding whether one should fit a linear regression with or without a logarithmic transformation of the variables. Another problem of this sort occurs if one has to choose between a normal and a Cauchy both with unknown location and scale parameters; in this problem,

$$\eta = (\eta_1, \eta_2), f(x, 0, \eta) = \frac{1}{\sqrt{2\pi}\,\eta_1}\, e^{-\frac{1}{2\eta_1^2}(x-\eta_2)^2},\ f(x, 1, \eta) = \frac{1}{\pi\eta_1}\, \frac{1}{1+\left\{\frac{(x-\eta_2)}{\eta_1}\right\}^2}.$$

* Paper presented at the International Symposium on Recent Trends of Research in Statistics held at the Indian Statistical Institute, Calcutta, December 16–27, 1974.

Suppose we modify the last problem a bit by assuming that a double exponential—with unknown location and scale parameters—is also a possibility. More generally let us consider decision problems where we have a countably many separated families $f(\cdot, \theta, \eta)$, $\eta \in \Omega_\theta$, $\theta \in \Theta$, Θ being a countable set. By "separated" we mean that for each θ.

$$\inf_{\substack{\eta \in \Omega_\theta \\ \eta' \in \Omega_{\theta'} \\ \theta' \neq \theta}} d(f(\cdot, \theta, \eta), f(\cdot, \theta', \eta')) > 0.$$

For simplicity we shall consider in this paper the case where we have only one parameter θ. The general case, which can be treated in a similar way, will appear elsewhere.

Suppose then we have a countable separated family of densities $f(\cdot, \theta)$, $\theta \in \Theta$ with respect to a σ-finite measure μ;

$$\inf_{\theta' \neq \theta} d(f(\cdot, \theta'), f(\cdot, \theta')) > 0 \;\forall\; \theta.$$

We make the homogeneity assumption that the set $\{x; f(x, \theta) > 0\}$ is independent of θ.

Let

$$\rho(\theta', \theta) = \inf_{t \geq 0} \int [f(x, \theta')/f(x, \theta)]^t f(x, \theta) d\mu,\; \rho(\theta) = \sup_{\theta' \neq \theta} \rho(\theta', \theta). \quad \ldots \quad (1.1)$$

Then $1-\rho(\theta', \theta)$ is a measure of divergence introduced by Chernoff (1952). It can be shown vide Proposition 2.2, that Θ is separated iff $1-\rho(\theta) > 0 \,\forall \theta$. Note that Θ, or more precisely the family $\{f(\cdot, \theta), \theta \in \Theta\}$, is separated iff it is a discrete metric space under the metric $d(\theta', \theta) = d(f(\cdot, \theta'), f(\cdot, \theta))$ introduced earlier. The statistical problem is to pick the correct value of θ given a sample $X_1, \ldots, X_n$. In the rest of the paper we tacitly assume this set-up.

Let $\hat{\theta}_n$ be a maximum likelihood estimator (m.l.e). Let the loss function $W(\hat{\theta}_n, \theta) = 0$ if $\hat{\theta}_n = \theta$ and positive otherwise. Let $R(\hat{\theta}_n, \theta)$ be the risk under θ. Theorem 3.1 provides an asymptotic estimate of $R(\hat{\theta}_n, \theta)$; it is shown that under some conditions $\lim n^{-1} \log R(\hat{\theta}_n, \theta)$ exists and equals $\log \rho(\theta)$. Theorem 3.2 shows $\hat{\theta}_n$ has an asymptotic minimax property. From this we derive a new proof of the main result of Kraft and Puri (1974). In general there are maximum weighted likelihood estimators (m.w.l.e.) which are asymptotically better than $\hat{\theta}_n$, so that $\hat{\theta}_n$ is not in general asymptotically admissible. For a precise definition of these terms see Section 3.

In Section 4 following Hammersley (1950) we assume θ to be an integer and the loss to be squared error. We develop analogues of Cramer-Rao bounds and show that if an estimator attains it asymptotically at some θ then its risk at some other point tends to infinity !

For the special case of the normal there is a slight overlap between our paper and that of Laderman (1955).

A sequential approach to these problems is considered by Khan (1973b).

2. Some properties of the divergence function

We write $\phi(t) = E_\theta\{f(X_1, \theta')/f(X_1, \theta)\}^t$ suppressing the dependence of ϕ on θ' and θ.

Proposition 2.1 : (a) $\rho(\theta', \theta) = \phi(t_0)$ for some $0 < t_0 < 1$

(b) $\rho(\theta', \theta) = \rho(\theta, \theta')$.

Proof : (a) This follows from convexity of ϕ and the fact $\phi(t) = 1$, $t = 0, 1$.

(b) $$\rho(\theta', \theta) = \inf_{0 \leqslant t \leqslant} E_\theta\{f(X_1, \theta')/f(X_1, \theta)\}^t \text{ by (a)}$$

$$= \inf_{0 \leqslant t \leqslant 1} E_{\theta'}\{f(X_1, \theta)/f(X_1, \theta')\}^{1-t}$$

$$= \rho(\theta, \theta').$$

Proposition 2.2 : $2\{1-\rho^2(\theta', \theta)\}^{1/2} \geqslant d(\theta', \theta) \geqslant \{1-\rho(\theta', \theta)\}/2$.

Proof : Khan (1973b, Lemma 1) has shown that

$$2(1-\phi^2(1/2))^{1/2} \geqslant d(\theta', \theta).$$

This implies $$2\{1-\rho^2(\theta', \theta)\}^{1/2} \geqslant d(\theta', \theta).$$

For the other inequality in Proposition 2.2 we proceed as follows. Let $y \geqslant 0$, $0 \leqslant t \leqslant 1$.

Then

$$(1+ty)-(1+y)^t = 1+ty-\{1+ty(1+\xi)^{t-1}\} \qquad \dots \quad (2.1)$$
$$\leqslant y.$$

and for $0 \leqslant y \leqslant 1$, $0 \leqslant t \leqslant 1$, by concavity of $(1-y)^t$,

$$(1-y)^t \geqslant (1-y)\cdot(1)^t + y\cdot(0)^t = 1-y \geqslant 1-y-ty.$$

So

$$1-ty-(1-y)^t \leqslant y. \qquad \dots \quad (2.2)$$

Let

$$y = \begin{cases} \{f_{\theta'}(x)/f_\theta(x)\}-1 & \text{if } f_{\theta'}(x) \geqslant f_\theta(x) \\ 1-\{f_{\theta'}(x)/f_\theta(x)\} & \text{otherwise.} \end{cases}$$

Then

$$\phi(t) \geqslant 1 + \int_{f_{\theta'}(x) > f_\theta(x)} ty f_\theta(x) d\mu - \int_{f_{\theta'}(x) < f_\theta(x)} ty f_\theta(x) d\mu$$
$$- \int y f_\theta(x) d\mu$$
$$= 1 - d(\theta', \theta).$$

Hence $d(\theta', \theta) \geqslant 1-\rho(\theta', \theta)$, completing the proof.

We shall occasionally need the Monotone Likelihood Ratio (MLR) Assumption : Θ is a set of real numbers and $f(x, \theta')/f(x, \theta)$ is an increasing function of x whenever $\theta' > \theta$.

Proposition 2.3 : Suppose the MLR assumption holds. Then

$$\rho(\theta_1, \theta) > \rho(\theta_2, \theta) \text{ if } \theta < \theta_1 < \theta_2 \text{ or } \theta_2 < \theta_1 < \theta.$$

Proof : We prove a stronger result, namely, that for all t such that $0 \leqslant t \leqslant 1$,

$$E_\theta\{f(X_1, \theta_2)/f(X_1, \theta)\}^t \leqslant E_\theta\{f(X_1, \theta_1)/f(X_1, \theta)\}^t \quad \text{if } \theta < \theta_1 < \theta_2 \text{ or } \theta_2 < \theta_1 < \theta. \quad \ldots \quad (2.3)$$

We consider first $\theta < \theta_1 < \theta_2$. Let

$$A = \int \{f(x, \theta_2)/f(x, \theta_1)\}^t \; f_2(x) d\mu$$

where

$$f_2 = \frac{\{f(x, \theta_1)/f(x, \theta)\}^t f(x, \theta)}{\int \{f(x, \theta_1)/f(x, \theta)\}^t f(x, \theta) d\mu}.$$

Since $0 \leqslant t \leqslant 1$ and MLR assumption holds, $f(x, \theta_1)/f_2(x)$ is an increasing function of x. Also $\{f(x, \theta_2)/f(x, \theta_1)\}$ is an increasing function of x. Hence

$$A \leqslant \int \{f(x, \theta_2)/f(x, \theta_1)\}^t f(x, \theta_1) \leqslant 1$$

which is just (2.3) by definition of A. We can deal with the case $\theta_2 < \theta_1 < \theta$ in a similar way. This completes the proof.

3. Properties of M.L.E.

Let $x = (x_1, \ldots, x_n)$ and $\hat{\theta}_n(x)$ be an m.l.e., i.e., $\hat{\theta}_n$ is any element of the set $\{\theta'; \sup_{\theta''} f(x, \theta'') = f(x, \theta')\}$ if this set is non-empty and $\hat{\theta}_n$ is any element of Θ otherwise. Let $Z_i(\theta', \theta) = \log\{f(X_i, \theta')/f(X_i, \theta)\}$ and $S(\theta', \theta) = Z_1 + \ldots + Z_n$. Let $W(\theta', \theta)$ be the loss in estimating the true value θ by θ'. We assume $W(\theta', \theta) = 0$ if $\theta' = \theta$ and > 0 if $\theta' \neq \theta$.

In Theorem 3.1 we find an asymptotic value of the risk $R(\hat{\theta}_n, \theta) = E_\theta\{W(\hat{\theta}_n, \theta)\}$ and use it in Theorem 3.2 to prove a weak minimax result. The idea behind the proof of Theorem 3.1 is to show that $R(\hat{\theta}_n, \theta)$ behaves asymptotically like $W(\theta_1, \theta)$. $[P_\theta\{\hat{\theta}_n = \theta_1\}]$ for a suitably chosen $\theta_1 \neq \theta$, and so $n^{-1} \log R(\theta_n, \theta)$ behave like $n^{-1} \log P_\theta\{\hat{\theta}_n = \theta_1\}$.

Theorem 3.1 : *Suppose*

(i) $\sum_{\theta'\neq\theta} \rho(\theta', \theta) < \infty$ *and* (ii) $\sum_{\theta'\neq\theta} W(\theta', \theta)\, \rho(\theta', \theta) < \infty$.

Then

$$\lim \; n^{-1} \log R(\hat{\theta}_n, \theta) = \log \rho(\theta).$$

Proof: Let $B_{\theta'}$ be the set $\{S(\theta', \theta) \geqslant 0\}$. By Chernoff's (1952) inequality

$$\sum_{\theta'\neq\theta} P_\theta(B_{\theta'}) \leqslant \sum_{\theta'\neq\theta} \{\rho(\theta', \theta)\}^n \leqslant \sum_{\theta'\neq\theta} \{\rho(\theta', \theta)\} < \infty \quad \text{by (i)}.$$

So by the Borel-Cantelli lemma the probability that only finitely many $B_{\theta'}$'s occur is one, i.e., with probability one

$$\hat{\theta}_n \in \{\theta';\ \sup_{\theta''} f(x, \theta'') = f(x, \theta')\}.$$

Hence

$$P_\theta(\hat{\theta}_n = \theta') \leqslant P_\theta(B_{\theta'}) \leqslant \{\rho(\theta', \theta)\}^n$$

and so

$$R(\hat{\theta}_n, \theta) \leqslant \sum_{\theta'\neq\theta} W(\theta', \theta)\{\rho(\theta', \theta)\}^n. \quad \ldots \quad (3.1)$$

Because of (i), the supremum $\rho(\theta)$ of $\rho(\theta', \theta)$, $\theta' \neq \theta$ is attained at a finite number of points $\theta' = \theta_1, \ldots, \theta_k$. Since

$$\sum W(\theta', \theta) \left\{\frac{\rho(\theta', \theta)}{\rho(\theta)}\right\}^n \leqslant \sum W(\theta', \theta) \left\{\frac{\rho(\theta', \theta)}{\rho(\theta)}\right\}$$

is convergent uniformly in n, we get,

$$\lim \sum_{\theta'\neq\theta} W(\theta', \theta)\{\rho(\theta', \theta)/\rho(\theta)\}^n = \sum_{\theta'\neq\theta} W(\theta', \theta) \lim \{\rho(\theta', \theta)/\rho(\theta)\}^n$$

$$= \sum_{\theta'=\theta_1, \ldots, \theta_k} W(\theta', \theta). \quad \ldots \quad (3.2)$$

It follows from (3.1) and (3.2) that

$$\overline{\lim}\; n^{-1} \log R(\hat{\theta}_n, \theta) \leqslant \log \rho(\theta), \ldots . \quad \ldots \quad (3.3)$$

Let

$$C = \{S(\theta', \theta) > 0 \text{ from some } \theta' = \theta_1, \ldots, \theta_k\},$$

$$D = \bigcap_{\theta'\neq\theta_1, \ldots, \theta_k, \theta} B^c_{\theta'}, \quad W = \min\{W(\theta_1, \theta), \ldots, W(\theta_k, \theta)\}.$$

Then using Chernoff's (1952) theorem

$$R(\hat{\theta}_n, \theta) \geqslant W P_\theta(C \cap D)$$

$$\geqslant W\{P_\theta(C) - \sum_{\theta'\neq\theta_1, \ldots, \theta_k, \theta} P_\theta(B_{\theta'})\}$$

$$\geqslant W[\{\rho(\theta)-\epsilon\}^n - \sum_{\theta'\neq\theta_1, \ldots, \theta_k, \theta} \{\rho(\theta', \theta)\}^n] \quad \ldots \quad (3.4)$$

for any preassigned $\varepsilon > 0$ and n sufficiently large. Since by (i),

$$\sum_{\theta' \neq \theta} \left\{\frac{\rho(\theta', \theta)}{\rho(\theta)}\right\}^n$$

is uniformly convergent, we get as before

$$\lim \sum_{\theta' \neq \theta_1, \ldots, \theta_k, \theta} \{\rho(\theta', \theta)/\rho(\theta)\}^n = \sum_{\theta' \neq \theta_1, \ldots, \theta_k, \theta} \lim \{\rho(\theta', \theta)/\rho(\theta)\}^n = 0. \quad \ldots \quad (3.5)$$

By (3.4) and (3.5), $\underline{\lim}\, n^{-1} \log R(\hat{\theta}_n, \theta) \geqslant \log\{\rho(\theta) - \varepsilon\}$ which, taken with (3.3), completes the proof.

Note that the limiting value is independent of $W(\theta', \theta)$. If the loss is $0-1$ then obviously (i) and (ii) are same.

Theorem 3.2 : *Suppose the conditions of Theorem 3.1 hold and T_n is an estimator such that $\lim n^{-1} \log R(T, \theta)$ exists for all θ. Then*

$$\sup_\theta \lim n^{-1} \log R(T_n, \theta) \geqslant \sup_\theta \lim n^{-1} \log R(\hat{\theta}_n, \theta).$$

Proof : We consider a fixed value of θ say θ_0 and define θ_1 as in the proof of Theorem 3.1, i.e., $\rho(\theta_1, \theta_0) = \rho(\theta_0)$. Consider a prior π which assigns positive probability $\pi_i > 0$ to θ_i, $i = 0, 1$. Then the average risk is minimised by the Bayes estimator B_n which equals θ_0 if $S(\theta_1, \theta_0) < \log \{\pi_0 W(\theta_1, \theta_0)/(\pi_1 W(\theta_0, \theta_1))\}$ and equals θ_1 otherwise. Since $n^{-1} \log \{\pi_0 W(\theta_1, \theta_0)/(\pi_1 W(\theta_0, \theta_1))\}$ tends to zero it follows from Chernoff's (1952) theorem that $\lim n^{-1} \log R(B_n, \theta_i) = \log \rho(\theta_i, \theta_j)$, $i \neq j = 0, 1$. Now using the definition of θ_1 and Proposition 2.1(b), we get $\lim n^{-1} \log R(B_n, \pi) = \log \rho(\theta_0)$ where $R(B_n, \pi)$ is the average risk $\pi_0 R(B_n, \theta_0) + \pi_1 R(B_n, \theta_1)$. Since B_n is Bayes, we get

$$\max_{i=0,1} \lim n^{-1} \log R(T_n, \theta_i) \geqslant \lim n^{-1} \log R(B_n, \pi) = \log \rho(\theta_0).$$

Hence,

$$\sup_\theta \lim n^{-1} \log R(T_n, \theta) \geqslant \sup_\theta \log \rho(\theta).$$

An appeal to Theorem 3.1 now completes the proof.

If $\sup_\theta \rho(\theta) = 1$, the result is not useful. For then any estimator with $\lim R(T_n, \theta) \leqslant K \;\forall\; \theta$ has this weak minimax property even though T_n need not even be consistent.

It would be interesting to find conditions under which

$$\lim n^{-1} \sup_\theta \log R(\hat{\theta}_n, \theta) = \sup_\theta \lim n^{-1} \log R(\hat{\theta}_n, \theta). \quad \ldots \quad (3.6)$$

When (3.6) is true, we get from Theorem 3.2

$$\begin{aligned}\lim n^{-1} \sup \log R(T_n, \theta) &\geqslant \sup \lim n^{-1} \log R(T_n, \theta)\\ &\geqslant \sup \lim n^{-1} \log R(\hat{\theta}_n, \theta)\\ &= \lim n^{-1} \sup \log R(\hat{\theta}_n, \theta) \qquad \ldots \quad (3.7)\end{aligned}$$

which is a more meaningful minimax result than Theorem 3.2.

Theorem 3.3 : *Suppose conditions* (i), (ii) *of Theorem* 3.1 *hold. If, moreover,* (3.6) *holds and for each n there exists* T_n^0 *such that*

$$\sup_{\theta} R(T_n^0, \theta) = \inf_{T_n} \sup_{\theta} R(T_n, \theta), \qquad \ldots \quad (3.8)$$

then

$$\lim n^{-1} \sup_{\theta} \log R(T_n^0, \theta) = \sup_{\theta} \log \rho(\theta). \qquad \ldots \quad (3.9)$$

Proof : (3.9) follows from (3.8) and (3.7).

If Θ is finite all conditions of the theorem hold and so we get the main result of Kraft and Puri (1974).

If we make the additional assumption in Theorem 3.2 that $\rho(\theta_0) = \rho(\theta_1)$, $\forall \theta_0$ and some θ_1 depending on θ_0, then we can make the stronger assertion that $\lim n^{-1} \log R(T_n, \theta_0) < \lim n^{-1} \log R(\hat{\theta}_n, \theta_0)$ at any point θ_0 implies the reverse inequality at the corresponding θ_1, i.e., $\hat{\theta}_n$ is asymptotically admissible. This stronger result follows from the proof of Theorem 3.2.

Unfortunately, as the following simple example shows, we can, in general, do better asymptotically by using some m.w.l.e. instead of $\hat{\theta}_n$. Suppose Θ consists of just three points 1, 2 and 3. Suppose $\rho(1) = \rho(2, 1) > \rho(3, 1)$, $\rho(2) = \rho(3, 2) > \rho(1, 2)$ and $\rho(3) = \rho(2, 3) > \rho(1, 3)$. Let

$$\pi_\theta = \begin{cases} 1 & \text{if } \theta = 1 \\ \lambda^n & \text{if } \theta = 2, 3 \end{cases}$$

where $0 < \lambda < 1$ is to be chosen later. Consider an estimator T_n which maximises the weighted likelihood $\pi_\theta f(x, \theta)$. Let

$$\gamma(\theta', \theta) = E_\theta[f(X_1, \theta')\pi_{\theta'}^{1/n}/\{f(X_1, \theta)\pi_\theta^{1/n}\}]^t, \quad \gamma(\theta) = \sup_{\theta' \neq \theta} \gamma(\theta', \theta).$$

Then it can be shown as in the proof of Theorem 3.1 that $\lim n^{-1} \log R(T_n, \theta) = \log \gamma(\theta)$. If we choose λ sufficiently close to one so that (i) $\gamma(2) = \gamma(3, 3)$ and (ii) $\gamma(3) = \gamma(2, 3)$, then $\gamma(1) < \rho(1)$ and $\gamma(\theta) \leqslant \rho(\theta)$, $\theta = 2, 3$. Thus T_n is asymptotically better than $\hat{\theta}_n$.

If Θ is not finite but the MLR assumption holds, then too usually one can construct an asymptotically better m.w.l.e. Asymptotic admissibility of $\hat{\theta}_n$ seems to be an exception rather than the rule.

4. Estimation of an integer valued parameter

In this section we assume Θ is the set of integers and allow an estimate to be any real number, not necessarily an ingeter. The loss in estimating θ by a is $(a-\theta)^2$.

Both Theorems 3.1 and 3.2 are still applicable. But the proof of Theorem 3.2 needs some change, since the form of the Bayes estimator B_n in the present set-up will be quite different from that in Section 3. However, one can still show that $\lim n^{-1} \log R(B_n, \pi) = \log \rho(\theta_0)$ and so the proof goes through. We omit the details.

Following Hammersley (1950) we develop an analogue of the Cramer-Rao bound. Let, $\lambda = \{f(X, \theta_1)/f(X, \theta_0)\}$, $\theta_1 \neq \theta_0$ and assume $A(\theta_1) = E_{\theta_0}\{f(X_1, \theta_1)/f(X_1, \theta_0)\}^2 < \infty$. Then λ has finite variance $A^n(\theta_1)-1$ under θ_0. By the Cauchy-Schwartz inequality, we have for any estimator T_n

$$\text{var}(T_n \mid \theta_0) \geqslant \{E_{\theta_1}(T_n)-E_{\theta_0}(T_n)\}^2/\{A^n(\theta_1)-1\}$$

$$= (\theta_1-\theta_0)^2/\{A^n(\theta_1)-1\} \qquad \ldots \quad (4.1)$$

if $E_\theta(T_n) = \theta$ for $\theta = \theta_0, \theta_1$. Note that $A(\theta_1) > 1$. Clearly the best bound of this type is obtained by maximising (4.1) with respect to θ_1. When the MLR assumption holds, the maximum occurs either at $\theta_1 = \theta_0-1$ or θ_0+1, answering partly a question of Hammersley (1950).

The following result shows that if an asymptotically unbiased estimator attains a Cramer-Rao bound (4.1) asymptotically at θ_0 then its variance tends to infinity under θ_1. Thus there does not exist an estimator attaining a Cramer-Rao bound asymptotically at all θ_0. This solves a problem raised by Hammersley (1950).

Theorem 4.1 : *Suppose T_n is an estimator such that $\lim E_\theta(T_n) = \theta$ if $\theta = \theta_1, \theta_0$. Let $A(\theta_1) < \infty$. If $\lim n^{-1} \log R(T_n, \theta_0) = -\log A(\theta_1)$ then $\lim R(T_n, \theta_1) = \infty$.*

Proof : Assume $\theta_1 > \theta_0$ without loss of generality. We use T_n to construct a test of $H_0(\theta = \theta_0)$ vs $H_1(\theta = \theta_1)$ as follows.

If $\quad T_n-\theta_0 > k$ accept H_1

$\qquad\qquad\quad \leqslant k$ accept H_0

where $0 < k < \theta_1-\theta_0$. Let α_n, β_n be the errors of first and second kind of this test. Let α_n' be the error of first kind of the most powerful test of H_0 vs H_1 which has error of second kind equal to β_n. Note that

$$\alpha_n' \leqslant \alpha_n \leqslant E_{\theta_0}(T_n-\theta_0)^2/k^2. \qquad \ldots \quad (4.2)$$

So by our assumption on T_n,

$$\overline{\lim}\ n^{-1} \log \alpha'_n \leqslant -\log A(\theta_1). \qquad \dots \quad (4.3)$$

Suppose, if possible, $\underline{\lim}\ \beta_n < 1$. Then by Stein's lemma—see Rao (1962, Lemma 4.2)—we can choose a subsequence n_i such that

$$\lim\ n_i^{-1} \log \alpha'_{n_i} = -I \qquad \dots \quad (4.4)$$

where $$I = E_{\theta_1} \{\log(f(X_1, \theta_1)/f(X_1, \theta_0))\}.$$

But $I < \log A(\theta_1)$ and so (4.4) contradicts (4.3). So $\underline{\lim}\ \beta_n = 1$ and hence

$$\lim\ \beta_n = 1. \qquad \dots \quad (4.5)$$

We now show (4.5) implies $\lim\ E_{\theta_1}(T_n-\theta_1)^2 = \infty$.

Let

$$\lambda_1 = E_{\theta_1}(T_n | T_n-\theta_0 \leqslant k),\ \lambda_2 = E_{\theta_1}(T_n | T_n-\theta_0 > k).$$

Then,

$$\beta_n\lambda_1+(1-\beta_n)\lambda_2 = E_{\theta_1}(T_n) = \theta_1+b(n)$$

where $b(n)$ is the bias.

Note that $\lambda_1 \leqslant \theta_0+k$ and $b(n) \to 0$. Hence

$$\underline{\lim}\ (\lambda_1-\theta_1-b(n))^2 > 0. \qquad \dots \quad (4.6)$$

Now,

$$E_{\theta_1}(T_n-\theta_1)^2 \geqslant E_{\theta_1}(T_n-\theta_1-b(n))^2$$

$$\geqslant \beta_n[\lambda_1-\theta_1-b(n)]^2+(1-\beta_n)[\lambda_2-\theta_1-b(n)]^2$$

by Jenson's inequality

$$= \beta_n[\lambda_1-\theta_1-b(n)]^2/(1-\beta_n)$$

which tends to infinity by (4.5) and (4.6).

5. Two examples

5.1. *Normal with integral mean.* Let X_i's be normal with known variance σ^2 and mean θ, $\theta = 0, \pm 1, \pm 2, \dots$. The m.l.e. $\hat{\theta}_n$ = nearest integer to the sample mean $\bar{X}_n$. By Theorem 3.1, $\lim\ n^{-1}(\log\ E_\theta(\hat{\theta}_n-\theta)^2) = \frac{1}{8}\sigma^2$. This was shown in a different way by Hammersley (1950). By one of the remarks following Theorem 3.2, $\hat{\theta}_n$ is asymptotically admissible. In the class of all translation invariant estimators T_n satisfying

$$T_n(x_1+i, \dots, x_n+i) = T_n(x)+i, \quad i = 0, \pm 1, \pm 2, \dots,$$

the best estimator with respect to the squared error loss is given by

$$T_n^0 = \psi(\bar{X}_n - i) + i, \quad \text{if} \quad i - \frac{1}{2} \leqslant \bar{X}_n < i + \frac{1}{2}$$

where
$$\psi(x) = -\Big\{\sum_i i e^{-n(i+x)^2/2\sigma^2}\Big\} \Big/ \Big\{\sum_i e^{-n(i+x)^2/2\sigma^2}\Big\}.$$

In particular T_n^0 is better than the m.l.e. $\hat{\theta}_n$ so that $\hat{\theta}_n$ is neither minimax nor admissible. Khan (1973a) has shown this in a somewhat different way. Of course T_n^0 is minimax and probably admissible. It can be shown that

$$\lim n^{-1} \log E_\theta(T_n^0 - \theta)^2 = \lim n^{-1} \log E_\theta(\hat{\theta}_n - \theta)^2.$$

One would not recommend T_n if one wants integer valued estimators. In the class of integer valued translation invariant estimators, $\hat{\theta}_n$ is best and hence minimax as stated by Stein in the discussion following Hammersley (1950). Probably $\hat{\theta}_n$ is admissible among integer valued estimators.

Theorems 3.1 and 3.2 hold also for the zero-one loss. For this loss function $\hat{\theta}_n$ is the best among all translation invariant estimators and hence minimax. Khan (1973a) has proved $\hat{\theta}_n$ is admissible.

5.2. *Poisson with integral mean.* Let $f(x, \theta) = e^{-\theta}\theta^x/x\,!$, $\theta = 1, 2, \ldots$. Proposition 2.3 and some computations show

$$\begin{aligned} \rho(\theta) = \rho(\theta+1, \theta) &= \exp\{-t_0 - \theta + \theta(1+1/\theta)^{t_0}\} \\ &= \exp\{-t_1/\log(1+1/\theta) + \theta(e^{t_1} - 1)\} \qquad \ldots \quad (5.1) \end{aligned}$$

where $t_0 = -\log\{\theta \log(1+1/\theta)\}/\log(1+1/\theta)$, $t_1 = t_0 \log(1+1/\theta)$.

Theorem 3.1 applies. Theorem 3.2, though true is not useful since $\sup \rho(\theta) = 1$. However one can prove that

$$\sup_\theta \lim n^{-1} \log E_\theta(T_n - \theta)^2/|\log \rho(\theta)| \geqslant \sup \lim n^{-1} \log E_\theta(\hat{\theta}_n - \theta)^2/|\log \rho(\theta)|.$$

Using the technique explained in Section 3 one can construct an m.w.l.e. which is asymptotically better than $\hat{\theta}_n$. But it may be of interest to consider instead a natural competitor T_n = nearest integer to $\bar{X}_n$ and show it is asymptotically better. Using Chernoff's (1952) theorem, we can show

$$\lim n^{-1} \log E_\theta(T_n - \theta)^2 = \inf_{t \geqslant 0} \psi(t)$$

where $\psi(t) = \theta(e^t - 1) - t(\theta + 1/2)$. By (5.1),

$$\begin{aligned} \log \rho(\theta) - \inf_{t \geqslant 0} \psi(t) &\geqslant -t_1\{\log(1+1/\theta)\}^{-1} + \theta(e^{t_1} - 1) - \psi(t_1) \\ &= t_1\{(\theta + 1/2)\log(1+1/\theta) - 1\}/\log(1+1/\theta). \qquad \ldots \quad (5.2) \end{aligned}$$

Let
$$f(x) = \log\left(1+\frac{1}{x}\right) - \frac{1}{x+\frac{1}{2}}.$$

Then
$$f'(x) = -1/[x(x+1)(2x+1)^2] < 0 \; \forall \; x > 0. \qquad \ldots \quad (5.3)$$

Also $\lim_{x \to \infty} f(x) = 0$. So $f(x) > 0 \; \forall \; x > 0$.

Hence
$$\left(x+\frac{1}{2}\right)\log\left(1+\frac{1}{x}\right) - 1 = \left(x+\frac{1}{2}\right) f(x) > 0 \; \forall \; x > 0. \quad \ldots \quad (5.4)$$

From (5.2) and (5.4) it follows that T_n^0 is asymptotically better than $\hat{\theta}_n$ under all θ. Some explanation of this phenomenon is provided below.

Note that T_n maximises the weighted likelihood $p_\theta f(x, \theta)$ where p_θ's are defined recursively as follows :

$$p_1 = 1$$

$$n^{-1}\log p_{\theta+1} - n^{-1}\log p_\theta = 1 - \left(\theta+\frac{1}{2}\right)\log(1+1/\theta), \quad \theta \geqslant 1.$$

By (5.4) p_θ is a decreasing function of θ. So it is to be expected that under θ, T_n takes the value $\theta+1$ with smaller probability than $\hat{\theta}_n$. Since the biggest contribution to variance of T_n and $\hat{\theta}_n$ come from $P_\theta\{T_n = \theta+1\}$ and $P_\theta\{\hat{\theta}_n = \theta+1\}$ respectively, we have here a simple explanation of the better performance of T_n.

6. Miscellaneous remarks

It is of interest to compare the limiting value of $n^{-1}\log_{10} E_\theta(\hat{\theta}_n - \theta)^2$ with its exact value for large but finite n. Some calculations with the normal $N(\theta, 1)$ are given below.

n	36	64	81	∞
$n^{-1}\log_{10}E_\theta(\hat{\theta}_n-\theta)^2$	−0.068	−0.065	−0.064	−0.054

It is not difficult to improve the approximations to $n^{-1}\log_{10}E_\theta(\hat{\theta}_n-\theta)^2$ by using better estimates for large deviation probabilities. See for example Theorem 2 of Bahadur and Rao (1960).

Theorem 3.1, but not the other results, is true without the homogeneity assumption.

In most practical cases one has several discrete and continuous parameters. It can be shown that a result similar to Theorem 3.1 holds in these cases also for the discrete parameters.

Possibly a more realistic reformulation of "separation" is to allow the extent of separation to depend on n. Let $(\Theta)_n$ be the parameter space when sample size is n and suppose for each $\theta \epsilon (\Theta)_n$,

$$\inf_{\substack{\theta' \neq \theta \\ \theta' \epsilon (\Theta)_n}} d(\theta', \theta) = d_n(\theta) > 0.$$

Depending on the behaviour of d_n as a function of n, one would have many asymptotic theories. If one has also estimates of errors in these asymptotic theories, one would be in a position to embed a given problem of separated parameters in the "right" sequence of problems.

These and some other problems will be considered elsewhere.

References

Bahadur, R. R. and Rao, R. R. (1960): On deviations of the sample mean. *Ann. Math. Statist.*, 31, 1015–1027.

Chernoff, H. (1952): A measure of asymptotic efficiency for tests of hypothesis based on the sums of observations. *Ann. Math. Statist.*, 23, 493–507.

Cox, D. R. (1962): Further results on tests of separate families of hypothesis. *J.R. Statist. Soc.*, Series B, 24, 406–424.

Hammersley, J. M. (1950): On estimating restricted parameters. *J.R. Statist., Soc.*, Series B, 12, 192–240.

Khan, R. A. (1973a): On some properties of Hammersley's estimator of an integer mean. *Ann. Statist.*, 1, 756–762.

—— (1973b): On sequential distinguishability. *Ann. Statist.*, 1, 838–850.

Kraft, O. and Puri, M. L. (1974): The asymptotic behaviour of minimax risk for multiple decision problems. *Sankhyā*, Series A, 36, 1–12.

Laderman, L. (1955): On the asymptotic behaviour of decision procedures. *Ann. Math. Statist.*, 26, 551–575.

Rao, C. R. (1962): Efficient estimates and optimum inference procedures in large samples. *J. R. Statist. Soc.*, Series B, 24, 46–72.