ON PARTIAL A PRIORI INFORMATION IN STATISTICAL INFERENCE

By J. R. Blum and Judah Rosenblatt

University of New Mexico

1. Introduction. Except in rare situations, information concerning the *a priori* distribution of a parameter is likely to be incomplete. Hence the use of a Bayes rule on some systematically produced choice for an *a priori* distribution, as advocated by the Bayesian school, is difficult to justify. This appears to be the case sometimes even if the *a priori* distribution is known fairly accurately (see Theorem 4). Robbins [3] has suggested that attention be paid to the case in which it is known only that the distribution of the parameter is a member of some given family $\mathfrak{F}$ of distributions. In this note we investigate this idea in several specific contexts—in particular the binomial case in which it is known that p is not less than some given p_0, and the case in which the class $\mathfrak{F}$ consists of distributions close to a given one.

Suppose we are given a fixed sample size statistical decision problem, i.e.,

a positive integer k,

a (parameter) set Θ,

an observable random vector $(X_1, \cdots, X_k) \equiv X$ with density f_θ, relative to some given measure m, where $\theta \varepsilon \Theta$ is unknown,

a set $\mathfrak{D}$ of possible decisions D,

a loss function $\mathfrak{L} \geqq 0$ on $\Theta \times \mathfrak{D}$, and

a σ-algebra S of subsets of $\mathfrak{D}$.

If no more information than that listed above is given, then we feel that it is most reasonable to use the minimax criterion, i.e., to use a rule δ which minimizes

$$\sup_{\theta\varepsilon\Theta} E_\theta \mathfrak{L}(\theta, \delta[X_1, \cdots, X_k]) \equiv \sup_{\theta\varepsilon\Theta} E_\theta[\textstyle\int_{\mathfrak{D}} \mathfrak{L}(\theta, D)\, d\delta(D \mid X_1, \cdots, X_k)]$$

where δ is a mapping from real k dimensional space into the set of probability measures over S, provided that such a rule exists.

In contrast to the case in which no further information is available is the situation in which θ is considered to be the value of a random variable governed by a known distribution function F, over an appropriate σ-algebra T of subsets of Θ. Then one naturally attempts to choose δ so as to minimize the average risk

$$\textstyle\int_\Theta E_\theta \mathfrak{L}(\theta, \delta[X_1, \cdots, X_k])\, dF(\theta) \equiv A(\delta, F).$$

Such a rule is called a Bayes rule relative to F and is usually denoted by δ_F. F itself is called the *a priori* distribution function of θ.

In many problems it is reasonable to assume the existence of an *a priori* distribution function F, but unreasonable to assume perfect knowledge of F. We consider here the problem of decision making when F is known only to be a member of some given class $\mathfrak{F}$.

Received 4 September 1964; revised 11 March 1967.

DEFINITION. Let

$$G(\delta, \mathfrak{I}) = \sup_{F \varepsilon \mathfrak{I}} A(\delta, F).$$

A rule $\delta_0 \varepsilon \mathfrak{I}$ is called $\mathfrak{I}$-minimax if for all $\delta G(\delta_0, \mathfrak{I}) \leqq G(\delta, \mathfrak{I})$.

We note that if $\mathfrak{I} = \{F\}$ then the $\mathfrak{I}$-minimax rules coincide with the Bayes rules δ_F, while if $\mathfrak{I}$ includes the set of all degenerate distributions over Θ then the $\mathfrak{I}$-minimax rules coincide with the minimax rules. (The sets of such rules may, of course, be empty. For some sufficient conditions yielding the existence of Bayes rules see Theorem 3.5, p. 89 of [4].)

2. Some particular cases. We first consider normal densities $N(\theta; x)$ with mean θ and variance 1 with squared error loss. Here, due to the large overlap with results of Cote-Skibinsky [1], we will omit many details. Due to the existence of a sufficient statistic no generality is lost by restricting to the case of a single observation.

THEOREM 1. *Let $p \varepsilon [0, 1]$ be given and for $k > 0$ let $\mathfrak{I}_{[k]} = \{F : F(0) = F(2k) = p\}$. Then the nonrandomized rule for estimating θ given by $\delta_0(x) = x$ is $\mathfrak{I}_{[k]}$-minimax.*

PROOF. We consider the Bayes solutions δ_{G_j} where

$$\begin{aligned} G_j'(\theta) = g_j(\theta) &= p/j && \text{for } -j \leqq \theta \leqq 0 \\ &= (1-p)/j && \text{for } 2k \leqq \theta \leqq 2k + j \\ &= 0 && \text{otherwise.} \end{aligned}$$

The form for δ_{G_j} is well known and can be assumed nonrandomized (see Chapter 4, p. 22 of [2]). It is not difficult to show that

$$\lim_{j \to \infty} A(\delta_{G_j}, G_j) = 1 \quad \text{and} \quad A(\delta_0, F) = 1$$

for all $F \varepsilon \mathfrak{I}_{[k]}$. Hence for arbitrary δ,

$$G(\delta, \mathfrak{I}_{[k]}) \geqq A(\delta, G_j) \geqq A(\delta_{G_j}, G_j) \to_{j \to \infty} 1 = \mathrm{G}(\delta_0, \mathfrak{I}_{[k]}),$$

showing δ_0 to be $\mathfrak{I}_{[k]}$-minimax. Q.E.D.

Thus it appears that even when a good deal seems to be known about F it may not be possible to lower the minimax risk.

We remark that if k is sufficiently large then looking only at θ outside the interval $(0, 2k)$, δ_0 is not admissible in the usual sense. To see this we look at the rule $\delta_{(k)}$ given by

$$\begin{aligned} \delta_{(k)}(x) &= \max(x, 2k) && \text{for } x > k \\ &= \min(0, x) && \text{for } x \leqq k. \end{aligned}$$

This rule only differs from $\delta_0(x)$ for $x \varepsilon (0, 2k)$.

Hence we need only show that for sufficiently large k, all $\theta \not\varepsilon (0, 2k)$.

$$\int_0^{2k} (\delta_{(k)}(x) - \theta)^2 e^{-\frac{1}{2}(x-\theta)^2} \, dx < \int_0^{2k} (x - \theta)^2 e^{-\frac{1}{2}(x-\theta)^2} \, dx.$$

By symmetry we may restrict consideration to $\theta < 0$, in which case it is sufficient to prove that for k sufficiently large, for all $\theta \leqq 0$,

(A) $\theta^2 \int_0^k e^{-\frac{1}{2}(x-\theta)^2} dx + (2k - \theta)^2 \int_k^{2k} e^{-\frac{1}{2}(x-\theta)^2} > dx < \int_0^{2k} (x - \theta)^2 e^{-\frac{1}{2}(x-\theta)^2} dx.$

This is equivalent to showing

$$4k^2 \int_k^{2k} e^{-\frac{1}{2}(x-\theta)^2} dx - 4k\theta \int_k^{2k} e^{-\frac{1}{2}(x-\theta)^2} dx < \int_0^{2k} x^2 e^{-\frac{1}{2}(x-\theta)^2} dx - 2\theta \int_0^{2k} x e^{-\frac{1}{2}(x-\theta)^2} dx$$

for k sufficiently large, all $\theta \leqq 0$. This will surely hold if for all sufficiently large k, for all $\theta \leqq 0$,

$$4k(k - \theta) \int_k^{2k} e^{-\frac{1}{2}(x-\theta)^2} dx < \int_0^{2k} x^2 e^{-\frac{1}{2}(x-\theta)^2} dx,$$

which will hold if

$$4k(k - \theta) \int_k^{2k} e^{-\frac{1}{2}(x-\theta)^2} dx < \int_1^{k+1} e^{-\frac{1}{2}(x-\theta)^2} dx$$

holds for k sufficiently large, all $\theta \leqq 0$. This latter condition is equivalent to

$$4k(k - \theta) \int_1^{k+1} e^{-\frac{1}{2}(u+k-1-\theta)^2} du < \int_1^{k+1} e^{-\frac{1}{2}(u-\theta)^2} du.$$

But for k sufficiently large, all $\theta \leqq 0$, $u \geqq 1$,

$$4k(k - \theta) e^{-\frac{1}{2}(u+k-1-\theta)^2} < e^{-\frac{1}{2}(u-\theta)^2}$$

is easily verified.

From this it follows that for k sufficiently large $A(\delta_{(k)}, F) < A(\delta_0, F)$ all $F \varepsilon \mathfrak{J}_{[k]}$, despite the fact that δ_0 is $\mathfrak{J}_{[k]}$-minimax.

It can be seen that for $k = \frac{1}{2}$ the rule $\delta_{(k)}$ is not minimax even when looking only at $\theta \not\varepsilon (0, 2k)$; in particular it can be verified that

$$\int_{1/2}^1 e^{-\frac{1}{2}x^2} dx > \int_0^1 x^2 e^{-\frac{1}{2}x^2} dx$$

which shows that (A) is reversed for $\theta = 0$, $k = \frac{1}{2}$.

We now examine a case which appears similar to the previous one, but in which the minimax risk is lowered by restriction to $\mathfrak{J}$. The first assertion in the following theorem is a special case of the theorem in Section 6 of [1].

THEOREM 2. *Let F_0 be the distribution function concentrating all mass at* 0 *and let*

$$\mathfrak{J}_p = \{F : F = pF_0 + (1 - p)H, H \quad \textit{arbitrary}\}.$$

Then $\delta_0(x) = x$ is not $\mathfrak{J}_p$-minimax for p sufficiently close to 1. *However the $\mathfrak{J}_p$-minimax risk strictly exceeds $1 - p$ for $0 < p < 1$.*

PROOF. Let $\mathfrak{H}$ denote the class of all distribution functions on the real line, and let E_0 denote expectation relative to the normal distribution with mean 0 and variance 1.

Then since for all δ, $A(\delta, F) = pE_0(\delta^2) + (1 - p)A(\delta, H)$, it follows that

(a) $G(\delta, \mathfrak{J}_p) = pE_0(\delta^2) + (1 - p)G(\delta, \mathfrak{H})$.

If now

(b) $E_0(\delta_1^2) < 1$ and $G(\delta_1, \mathfrak{IC}) < \infty$,
it is easy to verify that whenever p is such that

$$[G(\delta_1, \mathfrak{IC}) - 1]/[G(\delta_1, \mathfrak{IC}) - E_0(\delta_1^2)] < p \leqq 1$$

then $G(\delta_1, \mathfrak{I}_p) < 1 = G(\delta_0, \mathfrak{I}_p)$, since by Theorem 1 we must have $G(\delta_1, \mathfrak{IC}) \geqq 1$. Thus all we need do to prove the first assertion is to exhibit a rule satisfying (b).

Let $F_{(n)}$ denote the normal distribution function with mean 0 and variance n^2. The Bayes rule δ_{G_n} for $G_n = pF_0 + (1 - p)F_{(n)}$ satisfies

$$\lim_{n\to\infty} \delta_{G_n}(x) = 0, \qquad \lim_{|x|\to\infty} [\delta_{G_n}(x)/x] = 1,$$

and hence suggests that we let

$$\begin{aligned} \delta_1(x) &= 0 \quad \text{for} \quad |x| \leqq 1 \\ &= x \quad \text{for} \quad |x| > 1. \end{aligned}$$

The first part of (b) is easily verified, and the second follows from

$$\begin{aligned} A(\delta_1, H) &\leqq \int_{-\infty}^{\infty} [\int_{-1}^{1} \theta^2 N(\theta, x)\, dx + \int_{-\infty}^{\infty} (x - \theta)^2 N(\theta, x)\, dx]\, dH(\theta) \\ &\leqq \int_{-\infty}^{\infty} [k + 1]\, dH(\theta) = k + 1 \end{aligned}$$

where $k = \sup_{|x|, \theta \leqq 1} \theta^2 N(\theta, x) < \infty$. (The rule δ_1 is one of a class of rules satisfying (b) considered in [1].)

We now show that $\inf_\delta G(\delta, \mathfrak{I}_p) > 1 - p$ for $0 < p < 1$. Again using the fact that $G(\delta, \mathfrak{IC}) \geqq 1$ for all δ, it follows from (a) that $\inf_\delta G(\delta, \mathfrak{I}_p) \geqq 1 - p$. However if there were a sequence of rules δ_n such that $G(\delta_n, \mathfrak{I}_p) \to_{n\to\infty} 1 - p$, then we must have both $E_0(\delta_n^2) \to_{n\to\infty} 0$ and $G(\delta_n, \mathfrak{IC}) \to 1$. But the condition $E_0(\delta_n^2) \to_{n\to\infty} 0$ implies $\delta_n \to_{m\to\infty} 0$ in normal mean θ variance 1 probability for all θ, which implies $\lim_{n\to\infty} E_\theta(\delta_n - \theta)^2 \geqq \theta^2$, and hence $G(\delta_n, \mathfrak{IC}) \to_{n\to\infty} \infty$. Thus $G(\delta_n, \mathfrak{I}_p) \to_{n\to\infty} 1 - p$ is impossible. Q.E.D.

In reliability theory it is often assumed that the reliability p of an object (the probability of its functioning properly for a given amount of time) is not less than some given value p_0. We consider the question of whether one can improve over the obvious rule of letting $\delta_0(X) = \max(p_0, \delta(X)$ where $\delta(X)$ is a minimax estimate. Here $X = \sum_{i=1}^{n} X_i$ is the observed number of good units in a sample of size n, and we need only consider rules based on X since it is a sufficient statistic. To avoid completely trivial cases let us restrict consideration to the case in which the minimax risk is finite. We consider the specific case of squared error loss and let $\mathfrak{I}_{(p_0)} = \{F: F$ concentrates all mass on $[p_0, 1]\}$. The usual minimax estimate, as shown by Lehmann [2], is

$$\delta(X) = X/n^{\frac{1}{2}}(1 + n^{\frac{1}{2}}) + 1/2(1 + n^{\frac{1}{2}}),$$

and has constant risk $1/4(1 + n^{\frac{1}{2}})^2$. Let $\mathfrak{R}(p, \delta_0)$ denote the (usual) risk of δ_0 at p, where $\delta_0 = \max(p_0, \delta(X))$, and similarly for $\mathfrak{R}(p, \delta)$. If we let $p_* =$

$1/2(1 + n^{\frac{1}{2}})$, we see that just by making use of the improvement attainable when $X = 0$

(c) $\mathcal{R}(p, \delta) - \mathcal{R}(p, \delta_0) \geqq (1 - p)^n(p_0 - p_*)^2$

whenever $p \geqq p_0 > p_*$.) (Note that $\delta_0 = \delta$ for $p_0 \leqq p_*$.) Hence we see that if $p_* < p_0 < 1 - p_*$ and $p_0 \leqq p < 1 - p_*/2$

(c$'$) $\mathcal{R}(p, \delta) - \mathcal{R}(p, \delta_0) \geqq (p_*/2)^n(p_0 - p_*)^2$.

Let us now assume that

(d) $p_* < p_0 < 1 - p_*$;

(e) $p \geqq p_0$.

Under these conditions we shall now show that one of the rules δ_ϵ given by

$$\delta_\epsilon(X) = (1 + \epsilon)\delta_0(X)$$

for $\epsilon \varepsilon (0, (1 - p_*/2)/(1 - p_*) - 1)$, is better than δ_0 in the minimax sense.

To see this we first note that for $1 - p_*/2 \leqq p \leqq 1$, due to the fact (from (d)) $\delta_0(X) \leqq 1 - p_*$, $\delta_\epsilon(X)$ is closer to p than is $\delta_0(X)$. This, together with the fact (from (e)) that δ_0 is at least as good (for those p) as δ, yields

(f) $\mathcal{R}(p, \delta_\epsilon) < \mathcal{R}(p, \delta_0) \leqq \mathcal{R}(p, \delta)$ for $1 - p_*/2 \leqq p \leqq 1$.

It is easy to see that $\mathcal{R}(p, \delta_0)$ is the uniform (in p) limit of $\mathcal{R}(p, \delta_\epsilon)$ as $\epsilon \rightarrow 0$. Hence we can choose $\epsilon > 0$ sufficiently small so that (under (d))

(g) $|\mathcal{R}(p, \delta_\epsilon) - \mathcal{R}(p, \delta_0)| < \frac{1}{2}(p_*/2)^n(p_0 - p_*)^2$.

We now claim that under (e), for all such ϵ

(h) $\mathcal{R}(p, \delta_\epsilon) < \mathcal{R}(p, \delta) = 1/4(1 + n^{\frac{1}{2}})^2$.

For $1 - p_*/2 \leqq p \leqq 1$ this follows from (f) while for $p_0 \leqq p < 1 - p_*/2$ it follows from (c$'$) and (g). But under (d) we have

$$\lim_{p\to 1} P\{\delta_0(X) = \delta(X) = 1 - p_*\} = 1$$

since $\lim_{p\to 1} P\{X = n\} = 1$. Hence we see that under (d)

$$\lim_{p\to 1} \mathcal{R}(p, \delta_0) = \lim_{p\to 1} \mathcal{R}(p, \delta) = 1/4(1 + n^{\frac{1}{2}})^2.$$

We see therefore that under (d), for those p satisfying (e), δ_0 has the same maximum risk as δ, while for those δ_ϵ given in (h) (since $\mathcal{R}(p, \delta_\epsilon)$ is continuous in p) the maximum risk of δ_ϵ for those p satisfying (e) is less than that for δ. We have thus shown the following:

THEOREM 3. *Let p_0 be given in $(0, 1)$. Let X be a binomial random variable with parameters n and p, where it is given that $p \geqq p_0$, and that n is sufficiently large so that*

$$1/2(1 + n^{\frac{1}{2}}) < p_0 < 1 - 1/2(1 + n^{\frac{1}{2}}).$$

Then (over those $p \geqq p_0$) *the rule* $\delta_0(X) = \max(p_0, \delta(X))$ (*where* δ *is the unrestricted minimax rule*), *is not minimax.*

We remark that it should certainly be feasible to construct a good approximation to an $\mathfrak{I}_{(p_0)}$-minimax rule by discretizing and using linear programming (see for example Weiss [5]).

3. $\mathfrak{I}$ almost equal to J. In this section we discuss the question of when extremely precise knowledge of the *a priori* distribution is almost equivalent to perfect knowledge of this distribution.

Let J be a given *a priori* distribution function and let

$$\mathfrak{I}_{J,p} = \{F : F = pJ + (1 - p)H, H \text{ arbitrary}\}.$$

For $F \varepsilon \mathfrak{I}_{J,p}$, $A(\delta, F) = pA(\delta, J) + (1 - p)A(\delta, H)$. Hence we see that if the usual minimax risk is infinite (as in the case with squared error loss, when the X_i's are normal with unknown mean μ and unknown variance σ^2, i.e., $\mathfrak{L}(\mu, d) = (\mu - d)^2$, where *a priori* distributions are over $\{(\mu, \sigma)\}$ in two dimensions) then for all decision rules δ and all $p < 1$, $G(\delta, \mathfrak{I}_{J,p}) = \infty$.

In the following theorem we give sufficient conditions that

$$\lim_{p \to 1} \inf_\delta G(\delta, \mathfrak{I}_{J,p}) = A(\delta_J, J).$$

THEOREM 4. *Given the* a priori *distribution function* J, *suppose that there exists a corresponding Bayes rules* δ_J. *Suppose further that there is a sequence of rules* $\delta^{(k)}$ *such that*

(j) $\lim_{k \to \infty} A(\delta^{(k)}, J) = A(\delta_J, J)$, with $G(\delta^{(k)}, \mathfrak{K}) < \infty$ *for all* k,

where $\mathfrak{K}$ *is the class of all distributions on the* σ*-field* T *of* Θ. *Then*

$$\lim_{p \to 1} \inf_\delta G(\delta, \mathfrak{I}_{J,p}) = A(\delta_J, J).$$

PROOF. We see at once that

(k) $G(\delta, \mathfrak{I}_{J,p}) = pA(\delta, J) + (1 - p)G(\delta, \mathfrak{K})$.

Let $\epsilon_1, \epsilon_2, \cdots$ be an arbitrary sequence of positive numbers decreasing to 0, and let p_k be chosen in $(0, 1)$ for each k such that

(m) $0 < 1 - p_k \leqq \epsilon_k / G(\delta^{(k)}, \mathfrak{K}) \to_{k \to \infty} 0$.

By (k) and (m) $G(\delta^{(k)}, \mathfrak{I}_{J,p_k}) \leqq A(\delta^{(k)}, J) + \epsilon_k$. Now certainly $A(\delta_J, J) \leqq G(\delta^{(k)}, \mathfrak{I}_{J,p_k})$, hence

$$\lim_{k \to 0} G(\delta^{(k)}, \mathfrak{I}_{J,p_k}) = A(\delta_J, J).$$

Since for $p \geqq p_k$, $G(\delta^{(k)}, \mathfrak{I}_{J,p}) \leqq G(\delta^{(k)}, \mathfrak{I}_{J,p_k})$, it is immediate that for p sufficiently close to 1, $\inf_\delta G(\delta, \mathfrak{I}_{J,p})$ is close to $A(\delta_J, J)$. Q.E.D.

In the following we give some sufficient conditions for satisfaction of (j).

COROLLARY. *Let the random vector* **X** *have density* f_θ, $\theta \varepsilon \Theta$, *relative to the* σ*-finite measure* m. *Suppose that a Bayes rule* δ_J *corresponding to* J *exists, and that the minimax risk* (*both of course relative to the given loss* L) *is finite. Suppose fur-*

ther that there is an increasing sequence $\{B_k\}$ *of measurable subsets of* R^K *such that*

(i) $\sup_{\theta \varepsilon \Theta} \int_{B_k} \mathcal{L}(\theta, \delta_J[\mathbf{x}]) f_\theta(\mathbf{x})\, dm(\mathbf{x}) = c_k < \infty$,

(ii) $\int_{R^K - \lim B_k} f_\theta(\mathbf{x})\, dm(\mathbf{x}) = 0$ *for all* $\theta \varepsilon \Theta$.

Let δ_0 *be a rule for which the minimax risk is finite. Then the rule* $\delta^{(k)}$ *defined by*

(iii) $\delta^{(k)}(\mathbf{x}) = \delta_J(\mathbf{x})$ if $\mathbf{x} \varepsilon B_k$

$= \delta_0(\mathbf{x})$ if $\mathbf{x} \notin B_k$

satisfies (j) *of Theorem* 4.

Proof. For simplicity we omit the argument $\mathbf{x}$ wherever it should appear.

$$A(\delta^{(k)}, J) = \left(\int_\Theta \int_{B_k} + \int_\Theta \int_{R^K - B_k}\right) \mathcal{L}(\theta, \delta)^{(k)} f_\theta\, dm\, dJ(\theta)$$
$$\leqq A(\delta_J, J) + \int_\Theta \left[\int_{R^K - B_k} \mathcal{L}(\theta, \delta_0) f_\theta\, dm\right) dJ(\theta).$$

The second term is finite because δ_0 has finite minimax risk, and the first term is finite because $A(\delta_J, J)$ cannot exceed the minimax risk for δ_0. Now by the additivity of the integral, using (ii) we see that for each θ the inner integral approaches 0 as $k \to \infty$. Hence by the dominated convergence theorem (the dominating function being $\int_{R^K} \mathcal{L}(\theta, \delta_0) f_\theta\, dm$) the final term above approaches 0, thus verifying the first part of (j). To verify the second part

$$A(\delta^{(k)}, H) = \int_\Theta \int_{B_k} \mathcal{L}(\theta, \delta_J) f_\theta\, dm\, dH(\theta) + \int_\Theta \int_{R^K - B_k} \mathcal{L}(\theta, \delta_0) f_\theta\, dm\, dH(\theta).$$

The first integral, by (i) is bounded above by c_k, while the second is bounded above by the minimax risk for δ_0. Hence $G(\delta^{(k)}, \mathcal{K})$ is bounded by the sum of these quantities.

Remark 1. Easily verified conditions yielding (i) are $m(B_k) < \infty$ and $\sup_{\mathbf{x} \varepsilon B_k, \theta \varepsilon \Theta} \mathcal{L}(\theta, \delta_J[\mathbf{x}]) f_\theta(\mathbf{x}) < \infty$ for each k.

Remark 2. It is curious to note that while in the limit as $p \to 1$ the risk $A(\delta_J, J)$ can be achieved for the class $\mathfrak{I}_{J,p}$, it cannot be achieved in general by the "obvious candidate" δ_J. To see this, just let J concentrate all of its mass at 0, where $X_1, \cdots, X_n$ are normal with mean θ, variance 1, $\mathcal{L}(\theta, d) = (\theta - d)^2$. Then $\delta_J(\mathbf{x}) \equiv 0$ and a simple computation shows that $G(\delta_J, \mathfrak{I}_{J,p}) = \infty$ for $p < 1$. Thus we see the reason for working with the rules $\delta^{(k)}$.

Example 1. In the normal mean θ case with square error loss, so long as $\delta_J(x)$ is bounded on each finite interval, we have

$$\sup_\Theta \sup_{-k \leqq x \leqq k} (\theta - \delta_J(x))^2 N(\theta, x) < \infty;$$

hence letting $B_k = [-k, k]$ yields satisfaction of the conditions of Remark 1.

2. Clearly the conditions of the corollary hold for bounded loss functions—in particular 0–1 loss functions.

Let $\mathfrak{A}$ be some given class of distribution functions and let $\mathfrak{I}_{\mathfrak{A},p}$ be defined by

$$\mathfrak{I}_{\mathfrak{A},p} = \{F: F = pN + (1-p)H,\ H \text{ arbitrary},\ N \varepsilon \mathfrak{A}\}.$$

Theorem 4 generalizes to:

Suppose there is a sequence $\delta^{(k)}$ *of rules such that*

(j′) $\lim_{k\to\infty} G(\delta^{(k)}, \mathfrak{A}) = \inf_\delta G(\delta, \mathfrak{A})$

with $G(\delta^{(k)}, \mathfrak{K}) < \infty$ *for all* k. *Then*

$$\lim_{p\to 1} \inf_\delta G(\delta, \mathfrak{J}_{\mathfrak{A},p}) = \inf_\delta G(\delta, \mathfrak{A}).$$

The proof, almost the same as that of Theorem 4, follows from the fact that

$$G(\delta, \mathfrak{J}_{\mathfrak{A},p}) \leqq G(\delta, \mathfrak{A}) + (1 - p)G(\delta, \mathfrak{K}).$$

The corollary to Theorem 4 does not generalize so easily, since we must be able to show that

$$\lim_{k\to\infty} \sup_{N\varepsilon\mathfrak{A}} \int_\Theta \int_{R^K - B_k} \mathfrak{L}(\theta, \delta_0) f_\theta \, dm \, dN = 0.$$

If this can be shown then this corollary generalizes by omitting the reference to J and δ_J in the second sentence, and replacing δ_J in (i) and (iii) by $\delta_{\mathfrak{A},1/k}$ where $G(\delta_{\mathfrak{A},1/k}, \mathfrak{A}) \leqq \inf_\delta G(\delta, \mathfrak{A}) + 1/k$.

REFERENCES

[1] Cote, L. and Skibinsky, M. (1963). On the inadmissibility of some standard estimates in the presence of prior information. *Ann. Math. Statist.* **34** 539–548.

[2] Lehmann, E. Notes on the theory of estimation. University of California. Chapter IV, p. 22.

[3] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

[4] Wald, A. (1950). *Statistical Decision Functions.* Wiley, New York, 89.

[5] Weiss, L. (1961). *Statistical Decision Theory.* McGraw-Hill, New York. p. 106.