

BAYES AND MINIMAX ESTIMATES FOR QUADRATIC LOSS FUNCTIONS

M. A. GIRSHICK AND L. J. SAVAGE
STANFORD UNIVERSITY, UNIVERSITY OF CHICAGO

1. Introduction

This paper is concerned with certain technical aspects of Wald's theory of minimax risk and cursory familiarity with that theory, such as may be had by reading the introductory sections of [1] and [16], is accordingly assumed.

In the present stage of this theory it seems appropriate to invest some activity in the exploration of minimax problems of an intermediate level of generality. We in particular thought it might be fruitful to explore the problem of estimating a real parameter θ where the loss as a function of θ and the estimate g is of the special form

$$(1.1) \quad W(g, \theta) = \lambda(\theta)(g - \theta)^2; \quad \lambda(\theta) > 0.$$

Restriction to a single parameter is of course a convenience of the moment, for it would be surprising if any results obtained here could not be extended directly to quadratic forms. Interest in (1.1) is aroused by the power series consideration that any smooth nonnegative risk function can be expressed approximately in that form. We were further motivated to study (1.1) because a few very general results concerning it (given in section 2 below) led us to hope that we might discover a considerable body of theory at that level of generality. But thus far we have progressed beyond the results of section 2 only in two special contexts. The first of these special contexts (treated in section 3) is that in which θ is a parameter of translation and λ is constant. It is shown that such decision problems are closed, have minimax solutions with the symmetry which would be expected, and which are explicitly computable in terms of conditional expectations. There is every reason to believe that these results can be extended to many other loss functions for which the loss depends only on $g - \theta$.

The second special context in which we have studied (1.1) (treated in sections 4 and 5) pertains to families of distributions on the real line of the form $\beta(\tau) \exp[x\tau] d\psi(x)$, with $\theta = E(x|\tau)$, called exponential families. These exponential families are more versatile than would appear at first glance, as is explained early in section 4. If for one of these families $\lambda(\theta)$ is so chosen that $\lambda(\theta)V(x|\theta) = 1$ and if ψ is such that the range of τ can be the whole real line, then x is an admissible minimax estimate of θ if the range of τ is indeed taken to be the whole real line. If, further,

Part of the research contained in this paper was completed in the summer of 1949 while both authors were in the employ of the RAND Corporation, Santa Monica, California, and the remainder was completed under the sponsorship of Office of Naval Research projects at Stanford University and the University of Chicago.

ψ is confined to a finite interval of the real line, then (as is shown in section 5) under a very general assumption about the cost of observations, the minimax sequential estimate of θ is one of constant sample size and Nature's least favorable distribution for θ is rectangular over the finite interval.

2. Properties of Bayes estimates for quadratic loss functions

Let $dF(x|\theta)$ be a family of probability measures on a space X , depending on a real parameter θ . The problem is to estimate θ on the basis of an observation on x when it is known that the *a priori* distribution of θ is $G(\theta)$, where $G(\theta)$ is a probability measure on the real numbers.

An estimate of θ is a real valued measurable function g on X which associates with every sample point x a value $g(x)$. Let $W[g(x), \theta]$ be the loss to the statistician, if θ is the true parameter, x the observed sample points and g the estimate employed. The function W is a nonnegative measurable function of g and θ . For any choice of g and for any G , the risk $R_0(g, \theta)$ and the expected risk $R(g, G)$ are defined by

$$(2.1) \quad R_0(g, \theta) = \int W[g(x), \theta] dF(x|\theta) = E\{W[g(x), \theta]|\theta\}.$$

$$(2.2) \quad R(g, G) = \int R_0(g, \theta) dG(\theta) = E_G[R_0(g, \theta)] = E_G\{W[g(x), \theta]\}.$$

An estimate g^* minimizing $R(g, G)$ for a given G is, following Wald, called a Bayes estimate for that G . Wald shows that g^* , if it is Bayes for G , satisfies

$$(2.3) \quad r[g^*(x), x] = \inf_h r(h, x),$$

where

$$(2.4) \quad r(h, x) = E_G[W(h, \theta)|x],$$

and that if g^* satisfies (2.2) it is Bayes for G . This characterization is immediately clear if $R(g, G)$ is expressed in the form

$$(2.5) \quad R(g, G) = E_G\{r[g(x), x]\} = E_G\{E_G[W(g(x), \theta)|x]\},$$

and provided it is assumed, as it shall be for the rest of this section, that conditional expectation with respect to x is a *bona fide* expectation.

While (2.3), measurability difficulties notwithstanding, gives a general method for constructing Bayes estimates, specific properties and characterizations of such estimates are generally difficult to obtain except for relatively simple types of loss functions. Of these, the quadratic loss function defined by (1.1) appears, for reasons given in section 1, to be of special interest and the characterization of the Bayes estimates arising from it is the subject of this section.

Let

$$(2.6) \quad r(h, x) = E_G[\lambda(\theta)(h - \theta)^2|x].$$

For a given G and for a given sample point x , $r(h, x)$ may be finite for no real h . If the set $A_G \subset X$ for the elements of which $r(h, x) = \infty$ for all h has positive probability, then it is clear from (2.5) that $R(g, G) = \infty$ for all g , that is,

THEOREM 2.1. *A necessary¹ condition for the existence of a g for which $R(g, G) < \infty$ is that $P(x \in A_G) = 0$, or equivalently $P[\min_h r(h, x) = \infty] = 0$.*

If $x \in X - A_G$, then $r(h, x)$ is finite for some h , but if W is of the form (1.1) more can be said, namely,

THEOREM 2.2. *For each x , $r(h, x) < \infty$ either for no h , for exactly one h , or for every h . The second case implies $E[\lambda(\theta) | x] = \infty$, and the third $E[\lambda(\theta) | x] < \infty$.*

PROOF. In the first place, the following argument orally communicated to us by David Blackwell shows that if $a < b$ and $r(a, x) < \infty$ and $r(b, x) < \infty$, then $r(h, x) < \infty$ for all h .

Since $\lambda(\theta)(h - \theta)^2$ is a convex function of h , we have for all h in the closed interval (a, b)

$$(2.7) \quad \lambda(\theta)(h - \theta)^2 \leq \frac{b-h}{b-a} \lambda(\theta)(a - \theta)^2 + \frac{h-a}{b-a} \lambda(\theta)(b - \theta)^2,$$

and hence

$$(2.8) \quad r(h, x) \leq \frac{b-h}{b-a} r(a, x) + \frac{h-a}{b-a} r(b, x),$$

so that $r(h, x)$ is finite for all h in (a, b) . Now let $h_1 \neq 0$ be in (a, b) and let h_2 be any arbitrary value of h different from h_1 . Then, since

$$(2.9) \quad (h_1 - h_2)^2 + 2(h_1 - h_2)(h_2 - \theta) \leq (h_1 - \theta)^2,$$

$$(2.10) \quad E\{[(h_1 - h_2)\lambda(\theta) + 2(h_2 - \theta)\lambda(\theta)] | x\} < \infty.$$

Setting $h_2 = 0$,

$$(2.11) \quad E\{[h_1\lambda(\theta) - 2\theta\lambda(\theta)] | x\} < \infty.$$

Hence $E[\lambda(\theta) | x] < \infty$ and therefore $E[\theta\lambda(\theta) | x] < \infty$. The finiteness of $E[\theta^2\lambda(\theta) | x]$ now follows from the finiteness $r(h_1, x)$. Thus $r(h, x) < \infty$ for all h .

Next if $r(a, x) < \infty$ and $E[\lambda(\theta) | x] < \infty$ then $r(h, x) < \infty$ for all h .

Since $\lambda(\theta) > 0$, Schwartz's inequality yields

$$(2.12) \quad |E[\lambda(\theta)(a - \theta) | x]| \leq E[\lambda(\theta) | a - \theta | | x] \leq r(a, x)E[\lambda(\theta) | x] < \infty.$$

It follows therefore that $E[\theta\lambda(\theta) | x] < \infty$ since

$$(2.13) \quad E[\theta\lambda(\theta) | x] = E\{[\lambda(\theta)a - \lambda(\theta)(a - \theta)] | x\}.$$

The finiteness of $E[\theta^2\lambda(\theta) | x]$ follows from the finiteness of $r(a, x)$. Thus $r(h, x) < \infty$ for all h . Finally, if $r(h, x) < \infty$ for all h , $E[\lambda(\theta) | x] < \infty$. This follows directly from the proof of Blackwell's result mentioned above.

¹ That the condition is not sufficient is easily seen from the following example:

Let

$$W(g, \theta) = (g - \theta)^2; \quad 1 \leq |\theta| < \infty.$$

$$X = \{x; \quad 0 < x \leq 1\}.$$

$$dF(x | \theta) = |\theta| \quad \text{for } 0 < x \leq |\theta|^{-1},$$

$$= 0 \text{ elsewhere.}$$

$$dG(\theta) = \frac{1}{2} |\theta|^{-2} \quad \text{for } 1 \leq |\theta| < \infty,$$

$$= 0 \text{ elsewhere.}$$

Then $g^*(x) = 0$ while $R(g^*, G) = \infty$, as can easily be verified.

From theorem 2.2 we see that for all $x \in X - A_G$ if $E[\lambda(\theta) | x] < \infty$, or equivalently if $r(h, x)$ exists for at least two values of h , then $r(h, x)$ exists for all h and, moreover, $|E[\theta^i \lambda(\theta) | x]| < \infty$ for $i = 0, 1, 2$. Hence $r(h, x)$ can be written as

$$(2.14) \quad r(h, x) = h^2 E[\lambda(\theta) | x] - 2h E[\theta \lambda(\theta) | x] + E[\theta^2 \lambda(\theta) | x]$$

which, for a given x , is a quadratic function of h and therefore has a unique minimum at $h = h^*$ where²

$$(2.15) \quad h^* = h^*(x) = \frac{E[\theta \lambda(\theta) | x]}{E[\lambda(\theta) | x]},$$

a measurable function of x on the measurable set where $E[\lambda(\theta) | x] < \infty$. Thus we have established the following

THEOREM 2.3. *If G admits any estimate of finite expected risk then it admits a Bayes estimate, and essentially only one, defined thus:*

$$(2.16) \quad \begin{aligned} h^*(x) &= 0, & \text{if } r(h, x) = \infty \text{ for all } h \\ &= h_0, & \text{if } r(h, x) < \infty \text{ only for } h_0 \\ &= \frac{E[\theta \lambda(\theta) | x]}{E[\lambda(\theta) | x]}, & \text{if } r(h, x) < \infty \text{ for all } h. \end{aligned}$$

THEOREM 2.4. *If for a given a priori distribution $G(\theta)$, the expectation of $\lambda(\theta)$ exists, then the Bayes estimate $h^*(x)$ is biased, or the expected risk is zero.³*

PROOF. The existence of $E[\lambda(\theta)]$ implies the existence of $E[\lambda(\theta) | x]$ for almost all x . Thus $h^*(x)$ is given by (2.15). Assume that $h^*(x)$ is unbiased. That is

$$(2.17) \quad E[h^*(x) | \theta] = \theta \text{ for all } \theta.$$

Consider a probability measure H defined by

$$(2.18) \quad dH(\theta) = \frac{\lambda(\theta) dG(\theta)}{\int \lambda(\theta) dG(\theta)} = \frac{\lambda(\theta) dG(\theta)}{E_G[\lambda(\theta)]}.$$

Then $h^*(x)$ is given by

$$(2.19) \quad h^*(x) = E_H(\theta | x) = E_H[\theta | h^*(x)],$$

as can easily be verified from the definition of conditional probability. But equation (2.19) in conjunction with (2.17) implies that $h^*(x) = \theta$ with probability one, according to a theorem by J. Doob which will appear in his forthcoming book on the theory of probability.⁴

If $E[\lambda(\theta)]$ does not exist then theorem 2.4 no longer holds, as can be seen from the results in section 5.

² This form of the Bayes estimate with $\lambda(\theta) = 1$ was independently obtained by J. L. Hodges, Jr. and E. L. Lehmann [5].

³ This theorem is an oral communication from David Blackwell.

⁴ If the second moments of h^* and θ are assumed to exist then Doob's theorem follows immediately from the fact that $E(\theta h^*) = E(h^*)^2 = E(\theta^2)$ which implies that $E(h^* - \theta)^2 = 0$ with probability 1.

3. Minimax estimates for a quadratic loss function in the case of distributions depending on a location parameter

While Bayes estimates, as a class, possess important properties [15], in any given situation the assumption of the existence of an *a priori* distribution may not be valid, or if such a distribution exists, it may be unknown. In such cases, a procedure for selecting an estimate, which has certain merits, is the minimax procedure. The minimax method of decisions is due to A. Wald [14] and is mathematically closely akin to the theory of zero sum two person games, developed by J. von Neumann [7], [8].

An estimate of a parameter θ of a family of probability measures $dF(x|\theta)$ on a space X is called minimax for a given loss function $W[g(x), \theta]$ if $\sup_G R(g^*, G) = \inf_g \sup_G R(g, G)$ where G ranges over the set of all probability measures on the real numbers, g ranges over all estimates on X and $R(g, G)$ is given by (2.2).

It is well known and almost obvious that for any R

$$(3.1) \quad \inf_g \sup_G R(g, G) \geq \sup_G \inf_g R(g, G).$$

Under rather general conditions (see, for example, [16]) equality obtains in (3.1), or even more, there may actually exist a G^* such that

$$(3.2) \quad \sup_G R(g^*, G) = R(g^*, G^*).$$

Under such circumstances R is called closed and G^* if it exists is known as Nature's least favorable distribution.

Another concept introduced by Wald, which is of statistical importance, is that of the admissibility of an estimate. An estimate g^* is called admissible if there exists no uniformly better estimate, more exactly, if there exists no other estimate g such that $R_0(g, \theta) \leq R_0(g^*, \theta)$ for all θ with the inequality holding for at least one θ , where R_0 is the risk as defined in (2.1).

Neither a Bayes nor a minimax estimate need necessarily be admissible unless they are unique. However, it is easy to show that if g^* is Bayes with respect to an *a priori* distribution G , and if there exists a uniformly better estimate g , then g is also Bayes with respect to G and the inequality $R_0(g, \theta) < R_0(g^*, \theta)$ will hold only on a set θ of G measure zero.⁵

In this section we consider a class of estimates of a location parameter which have certain optimum properties and prove that these estimates are minimax and closed with respect to the risk function $\lambda[g(x) - \theta]^2$, where λ is a positive constant which may as well be and will be taken as unity.

To describe the family of distributions and class of estimates formally, let $x = \{x_1, \dots, x_n\}$ denote a point in the cartesian n -space X , and adopt the convention $x + \nu\theta = \{x_1 + \theta, \dots, x_n + \theta\}$. The *translation families* are characterized by the identity,

$$(3.3) \quad dF(x|\theta) = dF(x - \nu\theta|0).$$

⁵ If g also is not admissible then another Bayes estimate g' can be found which is uniformly better than g and so forth. In an unpublished result, Wald shows that under some conditions there exists a sequence of estimates which converges to a limit which is an admissible Bayes estimate.

One dimensional translation families need no introduction to this audience. As for n -dimensional ones, they arise (though not in full generality) as cartesian products, that is, distributions of n independent observations.

The function $u(k)$ will be said to possess the translation property if $u(x + \nu k) = u(x) + k$ for all real k . For any function $f(x)$, the symbol $E_0[f(x)]$ will stand for $E[f(x)|\theta = 0]$.

THEOREM 3.1. *If $u(x)$ has the translation property, $E\{f[u(x) - \theta]|\theta\} = E_0\{f[u(x)]\}$ for arbitrary f . In particular the bias and variance of u are independent of θ .*

If $u(x)$ is any estimate with the translation property and finite bias, let

$$(3.4) \quad z(x) = x - \nu u(x),$$

$$(3.5) \quad u^*(x) = u(x) - E_0[u(x)|z(x)].$$

THEOREM 3.2. *The estimate u^* derived from u :*

1. *Has the translation property,*
2. *Is unbiased,*
3. *Has as small a variance as any estimate having the translation property,*
4. *Is independent of the choice of u , provided some u has finite variance,⁶ except for sets of probability 0 for all θ .*

PROOF.

1. Obvious from invariance of $z(x)$.

2. From 1 and theorem 3.1, the bias of u^* is constant and equal to

$$(3.6) \quad E_0(u^*) = E_0(u) - E_0[E_0(u|z)] = 0.$$

3. If t has the translation property

$$(3.7) \quad V(t) = V(t - u^*) + V(u^*) + 2E_0[(t - u^*)u^*].$$

Now $t - u^*$ depends only on z , therefore

$$(3.8) \quad \begin{aligned} E_0[(t - u^*)u^*] &= E_0 E_0[(t - u^*)u^*|z] \\ &= E_0[(t - u^*)E_0(u^*|z)] = 0. \end{aligned}$$

4. By (3.7) and (3.8) it is clear that two really different estimates cannot have the variance $V(u^*)$, provided $V(u)$ is finite.

COROLLARY. $E\{[t(x) - \theta]^2|\theta\} = E_0[E^2(x)] + E_0^2[t(x)] \geq E_0[u^{*2}(x)]$.

THEOREM 3.3. *The estimate $u^*(x)$ is a minimax estimate for the risk function under consideration if any u has finite variance. If not, the minimax value of the problem is infinite. The problem is closed in any event.*

In view of part 4 of theorem 3.2, it is sufficient to prove the result for a particular choice of a function $u(x)$ possessing the translation property, say for definite-

$$\text{ness}^7 \bar{x} = \sum_{i=1}^n x_i / n.$$

⁶ This proviso could presumably be removed, but this seems worth no effort at the moment.

⁷ Strictly speaking, u must have finite bias so that if \bar{x} fails in this respect, some other u would have to be substituted in the subsequent argument.

Before actually proving this theorem it is expedient to prove some results for more general loss functions of the form

$$(3.9) \quad W[g(x), \theta] = W[g(x) - \theta].$$

Letting $s(x) = g(x) - \bar{x}$, the risk $R_0(s, \theta)$ may conveniently be expressed as a functional ρ of s and θ , thus,

$$(3.10) \quad \rho(s, \theta) = \int W[s(x + \nu\theta) + \bar{x}]dF(x).$$

We shall sometimes write $\rho(s, \theta, F, W)$ to indicate ρ 's dependence on the risk function W and on the measure under consideration. On the other hand, symbols W and $W(\theta)$ will sometimes be used as abbreviations for $W[s(x + \nu\theta) + \bar{x}]$.

Let \mathcal{S} be the set of real valued measurable functions on X , and \mathcal{Q} be the subset of \mathcal{S} such that for $f \in \mathcal{Q}$,

$$(3.11) \quad f(x + \nu k) = f(x),$$

for all real k . Note that for $f \in \mathcal{Q}$, $\rho(f, \theta)$ in (3.10) is independent of θ and may justifiably be written as $\rho(f)$.

Let

$$(3.12) \quad V = V(F, W) = \inf_f \sup_{\theta} \rho(f, \theta, F, W) = \inf_f \rho(f, F, W),$$

$$(3.13) \quad \bar{V} = \bar{V}(F, W) = \inf_s \sup_{\theta} \rho(s, \theta, F, W),$$

$$(3.14) \quad \underline{V} = \underline{V}(F, W) = \sup_G \inf_s \int \rho(s, \theta, F, W)dG(\theta),$$

for $f \in \mathcal{Q}$, $s \in \mathcal{S}$.

Since \mathcal{Q} is a subset of \mathcal{S} , it follows from (3.12) and (3.13) that $V \geq \bar{V}$, and from elementary game theoretic considerations, $\bar{V} \geq \underline{V}$. The general minimax problem under discussion can now be formulated as that of finding conditions under which $V = \underline{V}$. While we believe that this result holds under quite general conditions, we have thus far succeeded only in proving some lemmas leading to theorem 3.3, and theorem 3.4 below.

It is not true that $V = \bar{V}$, much less that $V = \underline{V}$ without any restrictions at all on the form of F , or of W in (3.9). The following example shows the sort of pathology which can occur even for a convex W , if it is monotonic.

Let X be one dimensional so that $x = \bar{x}$, and the elements of \mathcal{Q} are real constants. Define

$$(3.15) \quad \begin{aligned} W(\theta) &= \theta \text{ for } \theta \geq 0, \\ &= 0 \text{ for } \theta < 0, \end{aligned}$$

$$(3.16) \quad \begin{aligned} dF(x) &= x^{-2}dx \text{ for } x \geq 1, \\ &= 0 \text{ for } x < 1. \end{aligned}$$

Now, since for all f

$$(3.17) \quad \rho(f) = \int_m^{\infty} (x + f) x^{-2}dx = \infty,$$

where $m = \max(-f, 1)$, it follows that $V = \infty$. On the other hand, for any $a > 1$, let $s(x) = -a|x|$, then

$$(3.18) \quad \rho(s, \theta) = \int_1^\infty W(x - a|x + \theta|) x^{-2} dx,$$

where

$$(3.19) \quad \begin{aligned} W(x - a|x + a|) &= -(a - 1)x - a\theta && \text{if } x \geq -\theta, \\ &= (a + 1)x + a\theta && \text{if } x \leq -\theta. \end{aligned}$$

This function of x attains an absolute maximum of $-\theta$ at $x = -\theta$ and it is negative except in the interval $-a\theta/(a + 1)$ to $-a\theta/(a - 1)$ for nonpositive θ . Therefore

$$(3.20) \quad \begin{aligned} \rho(s, \theta) &= 0 && \text{for } \theta \geq 0, \\ &\leq \int_{-a\theta/(a+1)}^{-a\theta/(a-1)} \theta x^{-2} dx = \frac{2}{a} && \text{for } \theta < 0. \end{aligned}$$

Thus we have an example where $V = \infty$, $\bar{V} = \underline{V} = 0$ and $W(\theta)$ is convex.

LEMMA 1. Let F and W be such that for every $\epsilon > 0$ there exists a F' and a W' satisfying the conditions⁸

1. $F'(S) \leq F(S)$ for every $S \subset X$,
2. $W'(\theta) \leq W(\theta)$ for every θ ,
3. $V(F', W') \geq V(F, W) - \epsilon$,
4. $V(F', W') = \underline{V}(F', W')$,

then $V(F, W) = \underline{V}(F, W)$.

PROOF. Consider

$$(3.21) \quad \underline{V}(F, W) \geq \underline{V}(F', W') = V(F, W) - \epsilon.$$

LEMMA 2. Let W be bounded, say by W_0 , and let F be such that for some $m > 0$, $F\{x \mid |\bar{x}| > m\} = 0$, then $V(F, W) = \underline{V}(F, W)$.

PROOF. For every $\epsilon > 0$ and $T > m$, there exists an s such that

$$(3.22) \quad \begin{aligned} \underline{V} + \epsilon &\geq \frac{1}{2T} \int_{-T}^T d\theta \int W[s(x + \nu\theta) + \bar{x}] dF(x) \\ &= \frac{1}{2T} \int dF(x) \int_{-T}^T W[s(x + \nu\theta) + \bar{x}] d\theta, \end{aligned}$$

by Fubini's theorem. Now

$$(3.23) \quad \begin{aligned} \int_{-T}^T W[s(x + \nu\theta) + \bar{x}] d\theta &= \int_{-T-\bar{x}}^{T-\bar{x}} W d\theta + \int_{-T}^{-T-\bar{x}} W d\theta + \int_{T-\bar{x}}^T W d\theta \\ &= \int_{-T}^T W\{s[x + \nu(\varphi - \bar{x})] + \bar{x}\} d\varphi + \int_{-T-\bar{x}}^{-T-\bar{x}} W d\theta + \int_{T-\bar{x}}^T W d\theta, \end{aligned}$$

⁸ The measures F considered in this context though bounded need not be probability measures, that is, of total measure 1.

and

$$(3.24) \quad \frac{1}{2T} \int \left| \int_{-T}^{-T-\bar{x}} W d\theta + \int_{T-\bar{x}}^T W d\theta \right| dF(x) \leq \frac{W_0}{T} \int |\bar{x}| dF(x) \leq \frac{mW_0F(X)}{T}.$$

Therefore

$$(3.25) \quad \underline{V} + \epsilon \geq \frac{1}{2T} \int_{-T}^T d\varphi W \{ s[x + \nu(\varphi - \bar{x})] + \bar{x} \} d\varphi - \frac{mW_0F(X)}{T} \geq \inf_{\varphi} W \{ s[x + \nu(\varphi - \bar{x})] + \bar{x} \} d\varphi - \frac{mW_0F(X)}{T}.$$

Letting $T \rightarrow \infty$ and observing that $s[x + \nu(\varphi - \bar{x})] \in \mathcal{Q}$ for all φ , we see that $\underline{V} \geq V$, and the lemma is proved.

THEOREM 3.4.⁹ *If W is bounded then $V = \underline{V}$.*

PROOF. Let W_0 be an upper bound for W . For any $\epsilon > 0$ there is an m so large that $F[|x| > m] \leq \epsilon/W_0$. Define $dF'(x) = dF(x)$, for $|x| \leq m$ and $dF'(x) = 0$, for $|x| > m$. Then, according to lemma 2, $V(F', W) = V(F, W)$. On the other hand

$$(3.26) \quad V(F, W) = \inf_f \rho(f, F, W) \leq \inf_f \rho(f, F', W) + \epsilon = V(F', W) + \epsilon.$$

Therefore, lemma 1 applies to prove the theorem.

LEMMA 3. *Let X be one dimensional, γ a probability measure on X , $d\gamma_m(x) = d\gamma(x)$ for $|x| \leq m$ and $d\gamma_m(x) = 0$ for $|x| > m$. Let $W[s(x + \nu\theta) + \bar{x}] = [s(x + \nu\theta) + \bar{x}]^2$. Then $V(\gamma_m, W)$ monotonically approaches $V(\gamma, W)$.*

PROOF. By definition

$$(3.27) \quad V(\gamma_m, W) = \inf_f \int_{-m}^m (f + x)^2 d\gamma(x),$$

so that

$$(3.28) \quad V(\gamma_m, W) = a_2(m) - \frac{a_1^2(m)}{a_0(m)},$$

where $a_i(m) = \int_{-m}^m x^i d\gamma(x)$, $i = 0, 1, 2$, and provided that m is large enough so that $a_0(m) \neq 0$.

Case 1. $V(\gamma, W) < \infty$. In this case, in notation parallel to that used above,

$$(3.29) \quad V(\gamma, W) = a_2 - a_1^2.$$

But by the Lebesgue convergence theorem $\lim_{m \rightarrow \infty} a_i(m) = a_i$. Therefore case 1 is established.

Case 2. $V(\gamma, W) = \infty$. In this case, $a_2(m) \rightarrow \infty$. Since $a_0(m)$ approaches but does not reach 1 there is an m_0 such that $1 > a_0(m_0) > \frac{1}{2}$ and an arbitrarily large $m > m_0$ such that $a_0(m) > a_0(m_0)$. Consider the measure $\beta_m = \gamma_m - \gamma_{m_0}$.

$$(3.30) \quad V(\beta_m, W) = a_2(m) - a_2(m_0) - \frac{[a_1(m) - a_1(m_0)]^2}{a_0(m) - a_0(m_0)} \geq 0,$$

⁹ Under some additional restrictions A. Wald [13, pp. 318-320] proved the slightly weaker result $\bar{V} = \underline{V}$.

and since $a_0(m) - a_0(m_0) < \frac{1}{2}$,

$$(3.31) \quad a_2(m) > 2[a_1(m) - a_1(m_0)]^2.$$

Therefore

$$(3.32) \quad a_2(m) - \frac{3}{2}a_1^2(m) > \frac{1}{2}[a_1(m) - 4a_1(m_0)]^2 - 6a_1^2(m_0).$$

Let $a_2(m) \geq 2T^2$, and $T \geq 4|a_1(m_0)|$. Then

$$(3.33) \quad a_2(m) - \frac{3}{2}a_1(m) \geq 2T^2 - \frac{3}{2}T \quad \text{if } |a_1(m)| \leq T, \\ \geq \frac{1}{2}[T - 4|a_1(m_0)|]^2 - 6a_1^2(m_0) \quad \text{if } |a_1(m)| > T.$$

Therefore $a_2(m) - 3a_1^2(m)/2$ approaches infinity with $a_2(m)$ and consequently $V(\gamma_m, W) = a_2(m) - a_1^2(m)/a_0(m) \rightarrow \infty$. Finally the monotonicity is immediate from (3.27).

Return now to the proof of theorem 3.3, where, with the new notation $W = [s(x + \nu\theta) + \bar{x}]^2$, and the expectation symbol E used below stands for E_0 . Let $z = x - \nu\bar{x}$. Note that $E[s(x)|z] = E[s(z - \nu\bar{x})|z]$, and since for $f \in \mathcal{Q}$, $f(x) = f(z)$,

$$(3.34) \quad E[f(x)s(x)|z] = E[f(z)s(x)|z] = f(z)E[s(x)|z].$$

Moreover, the conditional expectation $E[s(x)|z]$ can, in this case, be so constructed as to be a probability integral for almost all z [4, p. 210, exercise 5] so that we can write,¹⁰

$$(3.35) \quad E[s(x)|z] = E[s(z + \nu\bar{x})|z] = \int s(z + \nu\bar{x})d\gamma_z(\bar{x}).$$

Case 1. $E(\bar{x}^2|z) < \infty$ for almost all z . In this case, let

$$(3.36) \quad f_0(x) = f_0(z) = -E(\bar{x}|z),$$

and

$$(3.37) \quad w(z) = E[(f_0 + \bar{x})^2|z] = E(\bar{x}^2|z) - f_0^2(z) \\ = \inf_r E[(r + \bar{x})^2|z] = \inf_r \int (r + \bar{x})^2 d\gamma_z(\bar{x}).$$

Now

$$(3.38) \quad \rho(f + f_0) = E\{E[f(x) + f_0(x) + \bar{x}]^2|z\} = E[w(z) + f^2(z)],$$

so that

$$(3.39) \quad V = \rho(f_0) = E[w(z)],$$

whether V is finite or not. (This result also follows from theorem 3.2.)

Let m be a positive number and define a measure $F_m(x) = F(x)$ for $|\bar{x}| \leq m$ and $F_m(x) = 0$ for $|\bar{x}| > m$. Let $h(x, m)$ be the characteristic function of the set $\{x | |\bar{x}| \leq m\}$, and let

$$(3.40) \quad f_m(z) = -\frac{E[h(x, m)\bar{x}|z]}{E[h(x, m)|z]},$$

where meaningful, and $f_m(z) = 0$ elsewhere. Further, let

$$(3.41) \quad w_m(z) = E[h(x, m)(f_m + \bar{x})^2|z] = E[h(x, m)\bar{x}^2|z] = f_m^2(z)E[h(x, m)].$$

¹⁰ This restriction could be foregone and would have to be if the work were much generalized.

Then by an argument similar to that employed above,

$$(3.42) \quad V(F_m, W) = \rho(f_m, F_m, W) = E[w_m(z)],$$

and

$$(3.43) \quad w_m(z) = \inf_r E[h(x, m)(r + \bar{x})^2 | z] = \inf_r \int_{-m}^m (r + \bar{x})^2 d\gamma_z(\bar{x}).$$

Now, by lemma 3, $\lim_{m \rightarrow \infty} w_m(z) = w(z)$, $w_m(z)$ is a nonnegative and nondecreasing function of m . Consequently, the Lebesgue theorem on integration of monotone sequences applies [12, theorem 12.6] to yield $\lim_{m \rightarrow \infty} V(F_m, W) = V(F, W)$.

Now, by lemma 2, $V(F_m, W) = \underline{V}(F_m, W)$ so that condition 3 of lemma 1 is satisfied. The remaining conditions of lemma 1 are satisfied as well and thus the theorem is established for case 1 whether V is finite or not.

Case 2. $E(\bar{x}^2 | z) = \infty$ on a set z of positive probability. At such values of z , $w_m(z) \rightarrow \infty$ by lemma 3 so that $V(F_m, W) \rightarrow \infty$. Therefore, $\infty = V(F, W) = \lim_{m \rightarrow \infty} V(F_m, W)$ and the proof of theorem 3.3 is completed.

We conjecture, but as yet have been unable to establish, that the minimax estimate $\bar{x} - E_0(\bar{x} | x - \nu\bar{x})$ derived here is unique for almost all x and hence is admissible.

$$\text{Let } dF(x - \theta) = \prod_{i=1}^n f(x_i - \theta) dx_1 \dots dx_n, \quad dG(\theta) = d\theta / (2T), \quad (-T \leq \theta \leq T)$$

and $W[u(x) - \theta] = [u(x_1, \dots, x_n) - \theta]^2$. Then the Bayes estimate $u_T^*(x_1, \dots, x_n) = E(\theta | x_1, \dots, x_n)$. If we let $T \rightarrow \infty$ we get generally

$$(3.44) \quad u^* = \lim_{T \rightarrow \infty} u_T^* = \frac{\int_{-\infty}^{\infty} \theta \prod_{i=1}^n f(x_i - \theta) d\theta}{\int_{-\infty}^{\infty} \prod_{i=1}^n f(x_i - \theta) d\theta}.$$

Setting $y_i = x_i - x_1, i = 2, \dots, n$ and letting $\theta = x_1 - \psi$,

$$(3.45) \quad u^* = x_1 - \frac{\int_{-\infty}^{\infty} \psi f(\psi) \prod_{i=2}^n f(y_i + \psi) d\psi}{\int_{-\infty}^{\infty} f(\psi) \prod_{i=2}^n f(y_i + \psi) d\psi} = x_1 - E_0(x_1 | x_2 - x_1, \dots, x_n - x_1).$$

Thus u^* is a minimax estimate of the type considered in this section. It will actually exist, provided some translation estimate has finite variance.

The estimate u^* given by (3.44) was originally introduced by E. J. G. Pitman [11] who showed, among other things, that it is of minimum variance in the class of estimates possessing the translation property. That the minimax estimates

under consideration are not necessarily maximum likelihood estimates is exemplified by the Cauchy family

$$(3.46) \quad f(x - \theta) = \frac{1}{\pi [1 + (x - \theta)^2]},$$

where for $n > 2$, u^* is a symmetric rational function of the x_i (no neat expression for it has as yet been obtained by us), while the maximum likelihood estimate is irrational.

It should be pointed out that there is a slight modification of translation families which might be called multiplicative. The following example will serve in lieu of a formal definition. Consider the estimation of the scale parameter θ in the family of incomplete Γ functions:

$$(3.47) \quad f(x, \theta) = \frac{x^{n-1} e^{-x/\theta}}{\Gamma(n) \theta^n}.$$

In some contexts it might seem reasonable to measure loss by the relative error of the estimate, that is,

$$(3.48) \quad W(g, \theta) = (\ln g - \ln \theta)^2 = \ln^2 \frac{g}{\theta}.$$

Taking then $\eta = \ln \theta$ as a parameter and $y = \ln x$ as the variable, the distribution density of y is given by

$$(3.49) \quad h(y, \eta) = \frac{e^{n(y-\eta)} e^{-e^{y-\eta}}}{\Gamma(n)}.$$

Hence the minimax estimate u^* for η is given by

$$(3.50) \quad \begin{aligned} u^* &= y - E_0(y) \\ &= \ln x - E(\ln x | \theta = 1) = \ln x - \frac{d}{dn} \ln \Gamma(n), \end{aligned}$$

which leads to $x \exp[-\Gamma'(n)/\Gamma(n)]$ as an estimate for θ . So far as we know this estimate has never been discussed. Still another (besides the maximum likelihood estimate x/n) estimate for this parameter is mentioned in section 4 of this paper.

4. Admissible minimax estimates for certain families of distributions which admit a sufficient statistic

Consider a family of probability measures defined by

$$(4.1) \quad dF(x|\theta) = \beta(\theta) \exp[u(x)\tau(\theta)]d\psi(x),$$

where x ranges over a measurable space X , $\psi(x)$ is a measure on X (not necessarily bounded), θ ranges over a set Θ , $u(x)$ and $\tau(\theta)$ are real valued functions, $u(x)$ is measurable, and for $\theta \in \Theta$,

$$(4.2) \quad \beta^{-1}(\theta) = \int e^{u(x)\tau(\theta)} d\psi(x) < \infty.$$

This type of family is a slight extension of that introduced by B. O. Koopman [6] and E. J. G. Pitman [10] in the investigation of sufficient statistics. Because of the Fisher-Neyman factorization theorem [9], [3], $u(x)$ is a sufficient statistic for

such a family, therefore, in discussing the estimation of θ , it will suffice to consider the distribution of u instead of x . Also, whether or not distinct values of θ lead to distinct values of τ , any question about the estimation of θ can be referred to one about the estimation of τ . In view of these considerations, there is no loss in generality in considering only distributions $dF(u|\tau)$ (to be referred to as exponential families) defined by

$$(4.3) \quad dF(u|\tau) = \beta(\tau)e^{u\tau}d\psi(u),$$

where u is real, ψ is a measure on the real numbers, τ ranges over a nonempty set T of real numbers, on which

$$(4.4) \quad \beta^{-1}(\tau) = \int_{-\infty}^{\infty} e^{u\tau}d\psi(u) < \infty.$$

Prior to discussing estimates for the parameter of exponential families, a few relevant theorems concerning the convolutions of two or more such families will be given.

THEOREM 4.1. *Let $dF(u|\tau)$, for each $\tau \in T$ be the distribution of the sum of n random variables u_i distributed according to*

$$(4.5) \quad dF_i(u_i|\tau) = \beta_i(\tau) e^{u_i\tau} d\psi_i(u_i), \quad i = 1, 2, \dots, n$$

over the same range T . Then

(a) $dF(u|\tau)$ is an exponential family over T such that

$$(4.6) \quad \beta(\tau) = \prod_{i=1}^n \beta_i(\tau); \quad d\psi(u) = \frac{e^{-u\tau_0}dF(u|\tau_0)}{\beta(\tau_0)},$$

for any $\tau_0 \in T$, and

(b) u is a sufficient statistic for the multivariate family $\prod dF_i(u_i|\tau)$.

PROOF (a): Let $f(u)$ be any bounded measurable function over the real numbers. Then

$$\begin{aligned} (4.7) \quad \int f(u) dF(u|\tau) &= \int \dots \int f\left(\sum_{i=1}^n u_i\right) \prod dF_i(u_i|\tau) \\ &= \prod \beta_i(\tau) \int \dots \int f\left(\sum u_i\right) e^{(\sum u_i)\tau} \prod dF_i(u_i|\tau) \\ &= \frac{\prod \beta_i(\tau)}{\prod \beta_i(\tau_0)} \int \dots \int f\left(\sum u_i\right) e^{(\sum u_i)(\tau-\tau_0)} \prod dF_i(u_i|\tau) \\ &= \frac{\prod \beta_i(\tau)}{\prod \beta_i(\tau_0)} \int f(u) e^{u(\tau-\tau_0)} dF(u|\tau_0), \end{aligned}$$

where τ_0 is any element of T . Hence we have

$$(4.8) \quad dF(u|\tau) = \beta(\tau)e^{u\tau}d\psi(u),$$

which proves part (a). Part (b) follows from the Fisher-Neyman factorization theorem [9], [3].

Note: If zero is in T , $\psi(u)$ is the convolution of the $\psi_i(u_i)$ and therefore if zero is not in T , $\psi(u)$ may be considered a sort of generalization of the convolution which in the strict sense may not exist.

A partial converse of part (a) of theorem 4.1 is given by

THEOREM 4.2. *If $dF(u|\tau)$ is an exponential family over T where $u = \sum u_i$ is the sum of n independent observations from a single family, $dF_0(u_0|\tau)$, then $dF_0(u_0|\tau)$ is an exponential family over T .*

PROOF. Let $\varphi(\lambda, \tau)$ and $\varphi_0(\lambda, \tau)$ be the moment generating functions of the two families, more precisely, let

$$(4.9) \quad \varphi(\lambda, \tau) = \int e^{\lambda u} dF(u|\tau); \quad \varphi_0(\lambda, \tau) = \int e^{\lambda u_0} dF_0(u_0|\tau),$$

for real λ and $\tau \in T$, infinite values being admitted. As the expected value of a product of independent random variables, $\varphi(\lambda, \tau) = \varphi_0^n(\lambda, \tau)$. On the other hand,

$$(4.10) \quad \begin{aligned} \varphi(\lambda, \tau) &= \beta(\tau) \int e^{(\lambda+\tau)u} d\psi(u) = \frac{\beta(\tau)}{\beta_0(\tau)} \int e^{(\lambda+\tau-\tau_0)u} dF(u|\tau_0) \\ &= \frac{\beta(\tau)}{\beta_0(\tau)} \varphi(\lambda + \tau - \tau_0; \tau_0), \end{aligned}$$

where τ_0 is any element of T . Therefore, since $\varphi(\lambda, \tau) > 0$, we have

$$(4.11) \quad \varphi_0(\lambda, \tau) = \left[\frac{\beta(\tau)}{\beta_0(\tau)} \varphi(\lambda + \tau - \tau_0; \tau_0) \right]^{1/n}.$$

But this is exactly the moment generating function $\varphi_1(\lambda, \tau)$ of the exponential family,

$$(4.12) \quad dF_1(u_1|\tau) = \beta_1(\tau) e^{u_1(\tau-\tau_0)} dF_0(u_1|\tau_0),$$

since

$$(4.13) \quad \varphi_1(\lambda, \tau) = \beta_1(\tau) \int e^{u_1(\lambda+\tau-\tau_0)} dF_0(u_1|\tau_0)$$

and

$$= \beta_1(\tau) \varphi_0(\lambda + \tau - \tau_0; \tau_0) = \beta_1(\tau) [\varphi(\lambda + \tau - \tau_0; \tau_0)]^{1/n}$$

$$(4.14) \quad \beta_1^{-1}(\tau) = \int e^{u_1(\tau-\tau_0)} dF_0(u_1|\tau_0) = \varphi_0(\tau - \tau_0; \tau_0)$$

$$= [\varphi(\tau - \tau_0; \tau_0)]^{1/n} = \left[\frac{\beta_0(\tau_0)}{\beta_0(\tau)} \right]^{1/n}.$$

Now, $\varphi(\lambda, \tau)$ is finite for λ between zero and $\tau_0 - \tau$, as is easily seen from equation (4.10) and an elementary property of the Laplace transform [17, p. 240]. This implies [17, p. 243] that $dF_0(u_0|\tau) = dF_1(u_0|\tau)$, which proves the theorem.

Any measure ψ defines an exponential family over any subset T' of the set T of values of τ for which

$$(4.15) \quad \omega(\tau) = \int e^{u\tau} d\psi(u) = \beta^{-1}(\tau) < \infty.$$

We shall henceforth work only with exponential families defined over the whole of the natural range T , given by (4.13). Elementary theorems about the Laplace transform [17, p. 240] say that T is connected, $\omega(\tau)$ is positive and analytic in the interior of T , and the first equality in (4.13) admits any number of differentiations under the integral sign.

Let

$$(4.16) \quad \theta(\tau) = \frac{d}{d\tau} \ln \omega(\tau) = \frac{\omega'(\tau)}{\omega(\tau)},$$

$$(4.17) \quad \sigma^2(\tau) = \frac{d\theta(\tau)}{d\tau} = \frac{d^2 \ln \omega(\tau)}{d\tau^2} = \frac{\omega(\tau)\omega''(\tau) - \omega'(\tau)^2}{\omega^2(\tau)}.$$

THEOREM 4.3. (a) $\theta(\tau) = E(u|\tau)$: (b) $\sigma^2(\tau) = E\{[u - \theta(\tau)]^2|\tau\}$.

PROOF. Straightforward calculations.

THEOREM 4.4. If T is the entire real line, then u is an admissible minimax estimate of $\theta(\tau)$ for the risk function.

$$(4.18) \quad W[g, \theta(\tau)] = \frac{[g(u) - \theta(\tau)]^2}{\sigma^2(\tau)}.$$

PROOF.¹¹ Since

$$(4.19) \quad \frac{1}{\omega(\tau)} \int_{-\infty}^{\infty} \frac{[u - \theta(\tau)]^2}{\sigma^2(\tau)} e^{u\tau} d\psi(u) = 1$$

for all τ , the minimax property of u will follow, if it is proven that u is admissible.

Assume that u is not admissible. Then there exists an estimate $g(u)$ uniformly better than u .

Let

$$(4.20) \quad \varphi(\tau) = E[g(u)|\tau] = \omega^{-1}(\tau) \int g(u) e^{u\tau} d\psi(u).$$

Then

$$(4.21) \quad \begin{aligned} \varphi'(\tau) &= \omega^{-1}(\tau) \left[\int u g(u) e^{u\tau} d\psi(u) - \frac{\omega'(\tau)}{\omega(\tau)} \int g(u) e^{u\tau} d\psi(u) \right] \\ &= \omega^{-1}(\tau) \int g(u) [u - \theta(\tau)] e^{u\tau} d\psi(u) \\ &= E\{g(u)[u - \theta(\tau)]|\tau\} \end{aligned}$$

where the differentiation under the integral sign is justified by an elementary theorem about the Laplace transform referred to above.

Now by the Schwartz inequality applied to equation (4.21),

$$(4.22) \quad [\varphi'(\tau)]^2 = \{Eg(u)[u - \theta(\tau)]|\tau\}^2 \leq \sigma^2(\tau)V(g|\tau).$$

Let $\varphi(\tau) = b(\tau) + \theta(\tau)$, then $\varphi'(\tau) = b'(\tau) + \sigma^2(\tau)$. Also, $V[g(\tau)] = \{E[g(u) - \theta(\tau)]^2|\tau\} - b^2(\tau)$. Substituting these in (3.22)

$$(4.23) \quad \sigma^2(\tau)b^2(\tau) + [b'(\tau) + \sigma^2(\tau)]^2 \leq \sigma^2(\tau)E[g(u) - \theta(\tau)]^2.$$

¹¹ The essential features of this proof are due to J. L. Hodges, Jr. and E. L. Lehmann and are contained in their paper, "Some applications of the Cramér-Rao inequality" appearing in this volume. Their results overlap with ours insofar as they deal with several specific cases of the exponential family of distributions. We are also indebted to H. Rubin for his assistance on some of the details of this proof.

Now by hypothesis $E[g(u) - \theta(u)]^2 \leq \sigma^2(\tau)$. Therefore

$$(4.24) \quad \sigma^2(\tau)b^2(\tau) + [b'(\tau) + \sigma^2(\tau)]^2 \leq \sigma^4(\tau),$$

it follows that $b'(\tau)$ is never positive, so that $b(\tau)$ is nonincreasing in τ . Now dropping the term $[b'(\tau)]^2$ from the left hand side of (4.24) and simplifying

$$(4.25) \quad b^2(\tau) + 2b'(\tau) \leq 0; \quad \text{or} \quad \frac{d}{d\tau} \left[\frac{1}{b(\tau)} \right] \geq \frac{1}{2},$$

where $b(\tau) \neq 0$.

Since $b'(\tau) \leq 0$, it is either true that $b(\tau) = 0$ for all sufficiently large τ , or that above some value τ_0 , $b(\tau) \neq 0$. In the latter event $b^{-1}(\tau) \geq \frac{1}{2}\tau - \frac{1}{2}[\tau_0 - b^{-1}(\tau_0)]$. Therefore in either event $b(\tau) \rightarrow 0$ as $\tau \rightarrow \infty$, and by the same argument $b(\tau) \rightarrow 0$ as $\tau \rightarrow -\infty$. Now if $b'(\tau) \leq 0$, and $b(\tau) \rightarrow 0$ as $\tau \rightarrow \pm\infty$, $b(\tau) = 0$ for all τ . This in turn implies that

$$(4.26) \quad \sigma^2(\tau) \leq E[g(u) - \theta(\tau)]^2,$$

as can be seen from equation (4.23). Consequently no uniformly better estimate can exist.

We remark that if T is not the entire real line, u need not be an admissible estimate. This is illustrated by the distribution of $\sigma^2\chi^2$ (an incomplete Γ family with a scale parameter) where, as is shown by Hodges and Lehmann [5] the admissible minimax estimate is $u^* = nu/(n+2)$. It is to be hoped that some simple treatment of the general case will be found.

Theorem 4.4 is somewhat more general than may be apparent at first sight. In the first place, many of the most important families of distributions met in statistics such as the normal, binomial, Poisson, negative binomial, the scale and index parameters of the Γ families, are of the exponential type and many have for T the whole line. Secondly, if u_1, \dots, u_n are different observations from exponential families $dF'_i(u_i|\tau)$, $u = \sum u_i$ is according to theorem 4.1 sufficient for τ . Accordingly it is to be expected that admissible and minimax estimates for virtually any function of τ depending only on u , will exist. In particular it is known by a theorem of Blackwell [2], [3, p. 240] that if the loss is quadratic in any function of τ , only functions of u can be admissible, and if any estimate is minimax there is a minimax function of u . Specializing still further, if the loss function is of the form

$$(4.27) \quad W(g, \tau) = \frac{\left[g - \sum E(u_i|\tau) \right]^2}{\sum V(u_i|\tau)} \\ = \frac{[g - E(u|\tau)]^2}{V(u|\tau)}$$

and T is the whole line, then theorem 4.4 implies that u itself is minimax with constant risk 1. As an example consider the familiar problem of estimating a success ratio from several independent binomial samples of different sizes. Finally, if all

the samples have the same distribution with mean and variance $\theta(\tau)$ and $\sigma^2(\tau)$, (4.27) may be expressed essentially thus:

$$(4.28) \quad W_0(g, \tau) = \frac{[g - \theta(\tau)]^2}{\sigma^2(\tau)}.$$

Here $\bar{u} = \sum u_i/n$ is an admissible minimax estimate with constant risk $1/n$. Since \bar{u} is unbiased, in the light of section 2, it is either not Bayes for any *a priori* distribution or only for one for which $E[\sigma^{-2}(\tau)] = \infty$. The first case is exemplified by the sample mean from a normal family and a class of examples of the second case will be found in the following section.

5. A class of minimax estimating procedures which are fixed sample size procedures

Let u be a random variable distributed according to an exponential family for which the measure $\psi(u)$ is assumed to have no rise outside a finite interval and only a finite rise in this interval. This interval will, without loss of generality, be taken to be $(0, \beta)$ with $\beta > 0$, and such that for all u interior to $(0, \beta)$, $Q(u) = \int_0^u d\psi(u') > 0$ and $Q(\beta) - Q(u) > 0$.

From the definition of $\psi(u)$, it is easily seen that the range of τ is from $-\infty$ to ∞ .

Thus, by theorem 4.4, $\bar{u} = \sum_{i=1}^n u_i/n$ is an admissible minimax estimate for the risk function (4.28) and for a fixed sample size n of observations u_1, \dots, u_n .

If in addition to the loss function (4.28), we assume that the cost of taking m observations is given by some real and positive function $c(m)$, the problem of finding Bayes or minimax estimating procedures for $\theta(\tau)$ involves introducing the general class of measurable sequential procedures as possible strategies for the statistician ([1], [16]). For the particular subclass of distributions under consideration, however, the minimax estimating procedure turns out to be a fixed sample size procedure. This result follows from several lemmas and theorems which we shall now prove.

LEMMA 1. $0 \leq \theta(\tau) \leq \beta$.

PROOF. Follows from theorem 4.3, which states that

$$(5.1) \quad \theta(\tau) = \frac{\int_0^\beta u e^{\tau u} d\psi(u)}{\int_0^\beta e^{\tau u} d\psi(u)}.$$

LEMMA 2. (a) $\lim_{\tau \rightarrow \infty} \theta(\tau) = \beta$; (b) $\lim_{\tau \rightarrow -\infty} \theta(\tau) = 0$.

PROOF (a). Let u_0 be any point of continuity of $\psi(u)$ interior to $(0, \beta)$. Then

$$(5.2) \quad \theta(\tau) = \frac{\int_0^\beta u e^{\tau u} d\psi(u)}{\omega(\tau)} \geq \frac{\int_{u_0}^\beta u e^{\tau u} d\psi(u)}{\omega(\tau)} \geq \frac{u_0}{1 + a(\tau) b^{-1}(\tau)},$$

where

$$(5.3) \quad a(\tau) = \int_0^{u_0} e^{\tau u} d\psi(u); \quad b(\tau) = \int_{u_0}^{\beta} e^{\tau u} d\psi(u).$$

Now if $\tau > 0$

$$(5.4) \quad a(\tau) \leq e^{\tau u_0} Q(u_0); \quad b(\tau) \geq e^{\tau(\beta+u_0/2)} \left[Q(\beta) - Q\left(\frac{\beta+u_0}{2}\right) \right],$$

therefore

$$(5.5) \quad \frac{a(\tau)}{b(\tau)} \leq e^{\tau(\beta-u_0/2)} \frac{Q(u_0)}{Q(\beta) - Q\left(\frac{\beta+u_0}{2}\right)}.$$

Comparing this conclusion with (5.2) and lemma 1, $\beta \geq \lim_{\tau \rightarrow \infty} \theta(\tau) \geq u_0$, which completes the proof of part (a). The proof of part (b) is similar.

In fact, an analogous lemma holds for any positive moment of u as can be seen from the proof of lemmas 1 and 2.

LEMMA 3. $\lim_{\tau \rightarrow \infty} \sigma^2(\tau) = \lim_{\tau \rightarrow -\infty} \sigma^2(\tau) = 0$.

PROOF. $\sigma^2(\tau) = E(u^2|\tau) - \theta^2(\tau)$.

LEMMA 4. Let u be an interior point of $(0, \beta)$. Then

$$(5.6) \quad \lim_{\tau \rightarrow \infty} \frac{e^{u\tau}}{\omega(\tau)} = \lim_{\tau \rightarrow -\infty} \frac{e^{u\tau}}{\omega(\tau)} = 0.$$

PROOF. Let $\tau > 0$ and $u_0 > u$ be a point of continuity of $\psi(u)$. Then

$$(5.7) \quad \omega(\tau) = \int_0^{\beta} e^{\tau u} d\psi(u) > \int_{u_0}^{\beta} e^{\tau u} d\psi(u) > e^{\tau u_0} \int_{u_0}^{\beta} d\psi(u).$$

Hence

$$(5.8) \quad \frac{e^{\tau u}}{\omega(\tau)} < e^{\tau(u-u_0)} [Q(\beta) - Q(u_0)]^{-1} \rightarrow 0 \quad \text{as} \quad \tau \rightarrow \infty.$$

Similarly for the second half of the lemma.

The conclusion of lemma 4 no longer necessarily holds when $u = 0$ and $\tau \rightarrow -\infty$ and $u = \beta$ and $\tau \rightarrow \infty$. However, we have the following

LEMMA 5. If $\psi(u)$ has a jump at $u = 0$ then $\lim_{\tau \rightarrow \pm\infty} \frac{1}{\omega(\tau)} < \infty$ and if $\psi(u)$ has a jump at $u = \beta$, then $\lim_{\tau \rightarrow \pm\infty} \frac{e^{\tau\beta}}{\omega(\tau)} < \infty$.

PROOF. Let $\psi(u)$ have measure γ at $u = 0$ and measure δ at $u = \beta$. Then $\omega(\tau) > \gamma$ and $\omega(\tau) > e^{\tau\beta}\delta$, so that

$$(5.9) \quad \frac{1}{\omega(\tau)} < \frac{1}{\gamma}; \quad \frac{e^{\tau\beta}}{\omega(\tau)} < \frac{1}{\delta} \quad \text{for all } \tau$$

which proves the lemma.

THEOREM 5.1. The estimate u for θ is Bayes for θ uniformly distributed from 0 to β .

PROOF. If θ is uniformly distributed from 0 to β , the distribution of τ is in view of lemmas 1 and 2 and equation (4.17) given by

$$(5.10) \quad dH(\tau) = \frac{d\theta(\tau)}{\beta} = \frac{\sigma^2(\tau)}{\beta}.$$

Under this restriction as under any other, the expected risk of the estimate u is 1, according to theorem 4.4. Therefore, according to theorem 2.3, a Bayes estimate of finite expected risk exists and is given by

$$(5.11) \quad g(u) = \frac{E(\theta\sigma^{-2}|u)}{E(\sigma^{-2}|u)},$$

whenever the numerator and denominator both exist, and $g(u) = u$ elsewhere.

Now

$$(5.12) \quad Ef(u, \tau) = \int \frac{\sigma^2(\tau)}{\beta} \int \frac{f(u, \tau) e^{u\tau}}{\omega(\tau)} d\psi(u),$$

therefore for any h for which the left member is finite

$$(5.13) \quad E[h(u)\theta(\tau)\sigma^{-2}(\tau)] = \frac{1}{\beta} \int d\psi(u) h(u) \int \frac{\theta(\tau)}{\omega(\tau)} d\tau.$$

If $h(u) \neq 0$, the inner integral exists and is given by

$$(5.14) \quad \begin{aligned} \int \frac{\omega'(\tau)}{\omega^2(\tau)} e^{u\tau} d\tau &= \left. \frac{-e^{u\tau}}{\omega(\tau)} \right|_{-\infty}^{\infty} + u \int \frac{e^{u\tau}}{\omega(\tau)} d\tau \\ &= u \int \frac{e^{u\tau}}{\omega(\tau)} d\tau, \end{aligned}$$

for almost all u , according to lemmas 4 and 5. Therefore

$$(5.15) \quad \begin{aligned} E[h(u)\theta(\tau)\sigma^{-2}(\tau)] &= E[h(u)u\sigma^{-2}(\tau)] \\ &= E[h(u)uE\sigma^{-2}(\tau)|\tau], \end{aligned}$$

so that

$$(5.16) \quad E[\theta(\tau)\sigma^{-2}(\tau)|\tau] = uE[\sigma^{-2}(\tau)|\tau],$$

where both exist, which completes the proof.

LEMMA 6. Let $r(\bar{u})$ be the conditional risk of $W_0(\bar{u}, \tau) = [\bar{u} - \theta(\tau)]^2/\sigma^2(\tau)$ given $\bar{u} = \sum_{i=1}^n u_i/n$, with respect to the a priori distribution (5.10). Then $r(\bar{u}) = 1/n$ for almost all \bar{u} .

PROOF. By the definition of conditional probability, $r(\bar{u})$ is given by

$$(5.17) \quad r(\bar{u}) = \frac{\int [\bar{u} - \theta(\tau)]^2 \frac{e^{\bar{u}\tau}}{\omega(\tau)} d\tau}{\int \sigma^2(\tau) \frac{e^{\bar{u}\tau}}{\omega(\tau)} d\tau},$$

where the denominator can be easily shown to exist for almost all \bar{u} .

Now

$$(5.18) \quad [\bar{u} - \theta(\tau)] \frac{e^{\bar{u}\tau}}{\omega(\tau)} = \frac{d}{d\tau} \left[\frac{e^{\bar{u}\tau}}{\omega(\tau)} \right],$$

so that

$$(5.19) \quad \int [\bar{u} - \theta(\tau)]^2 \frac{e^{\bar{u}\tau}}{\omega(\tau)} d\tau = \int [\bar{u} - \theta(\tau)] \frac{d}{d\tau} \left[\frac{e^{\bar{u}\tau}}{\omega(\tau)} \right] d\tau,$$

which when integrated by parts yields

$$(5.20) \int [\bar{u} - \theta(\tau)]^2 \frac{e^{-\bar{u}\tau}}{\omega(\tau)} d\tau = [\bar{u} - \theta(\tau)] \frac{e^{-\bar{u}\tau}}{\omega(\tau)} \Big|_{-\infty}^{\infty} + \frac{1}{n} \int \sigma^2(\tau) \frac{e^{-\bar{u}\tau}}{\omega(\tau)} d\tau \\ = \frac{1}{n} \int \sigma^2(\tau) \frac{e^{-\bar{u}\tau}}{\omega(\tau)} d\tau.$$

by lemmas 4 and 5, which proves the lemma.

The following lemma¹² is stated in terms of an estimation problem although it obviously holds for any decision problem. The proof of this lemma is trivial and is omitted. The implication of this lemma is that if the Bayes conditional risk for any point x_1, \dots, x_m depends only on m , then the optimum sequential decision procedure is a fixed sample size procedure.

LEMMA 7. *Hypothesis:*

1. $(\theta, x) = \theta, x_1, x_2, \dots$, an infinite sequence of random variables.
2. $W(g, \theta)$, the loss function, measurable and defined for g in the range of θ .
3. $g_m^*(x_1, \dots, x_m)$ a measurable sequence of functions (defined on the domains x_1, \dots, x_m as indicated), such that

$$h(m) = E\{W[g_m^*(x_1, \dots, x_m), \theta] | x_1, \dots, x_m\}$$

is independent of x_1, \dots, x_m , and

$$h(m) \leq E[W(g, \theta) | x_1, \dots, x_m]$$

for all g and all x_1, \dots, x_m .

4. $c(m)$, the cost of making m observations, a real valued function on the integers.
5. S a sequential sampling scheme and $g(x)$ an estimate defined for the x 's permitted by S .

Conclusion:

The total expected risk

$$E\{W[g(x), \theta] + c(m) | S\} \geq \inf_m [h(m) + c(m)],$$

and if S is of the single sample size m and $g(x) = g_m^*(x_1, \dots, x_m)$, the total expected risk is $h(m) + c(m)$.

THEOREM 5.2. Let the cost of taking m observations be given by a nonnegative function $c(m)$ and let the loss be given by 4.28. Then (a) the minimax sequential strategy for estimating θ is to use the estimate \bar{u} based on a fixed sample size N which is determined by $c(m)$, and (b) Nature's least favorable distribution for θ is uniform over the interval $(0, \beta)$.

PROOF. In view of theorem 5.1, it is only necessary to show that if Nature's strategy is uniform over the interval $(0, \beta)$ the statistician cannot improve his total average risk by using a sequential procedure provided the proper sample size is used. This conclusion follows from lemma 7 in conjunction with theorem 5.2 and

¹² This lemma was independently discovered by C. R. Blyth of the Statistical Laboratory, University of California, Berkeley, California.

lemma 6, which implies that for any stopping point x_1, x_2, \dots, x_m , prescribed by a sequential procedure S

$$(5.21) \quad E [W_0(\bar{u}, \tau) \mid x_1, \dots, x_m] = \inf_g [W_0(g, \tau) \mid x_1, \dots, x_m] = \frac{1}{m}.$$

We conjecture that part (a) of theorem 5.2 holds even if the range of $\psi(u)$ is not restricted to a finite interval provided T is the whole real line.

REFERENCES

- [1] K. J. ARROW, D. BLACKWELL and M. A. GIRSHICK, "Bayes and minimax solutions of sequential decision problems," *Econometrica*, Vol. 17 (1949), pp. 213-244.
- [2] D. BLACKWELL, "Conditional expectation and unbiased sequential estimation," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 105-110.
- [3] P. R. HALMOS and L. J. SAVAGE, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 225-241.
- [4] P. R. HALMOS, *Measure Theory*, D. Van Nostrand, New York, 1950.
- [5] J. L. HODGES, JR., and E. L. LEHMANN, "Some problems in minimax point estimation," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 182-197.
- [6] B. O. KOOPMAN, "On distributions admitting a sufficient statistic," *Trans. Amer. Math. Soc.*, Vol. 39 (1936), pp. 399-409.
- [7] J. VON NEUMANN, "Zur Theorie der Gesellschaftspiele," *Math. Annalen*, Vol. 100 (1928), pp. 295-320.
- [8] J. VON NEUMANN and O. MORGANSTERN, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, 1944.
- [9] J. NEYMAN, "Su un teorema concernente le cosiddette statistiche sufficienti," *Giorn. Ist. Ital. Atti.*, Vol. 6 (1935), pp. 320-334.
- [10] E. J. G. PITMAN, "Sufficient statistics and intrinsic accuracy," *Proc. Camb. Phil. Soc.*, Vol. 32 (1936), pp. 567-579.
- [11] ———, "The estimation of location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1939), pp. 391-421.
- [12] S. SAKS, *Theory of the Integral*, 2nd rev. ed., Eng. trans. by L. C. YOUNG, Monografie Matematyczne, Warszawa-Lwow, 1937.
- [13] A. WALD, "Contributions to the theory of statistical estimation and testing hypotheses," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 299-325.
- [14] ———, "Statistical decision functions which minimize the maximum risk," *Annals of Math.*, Vol. 46 (1945), pp. 265-280.
- [15] ———, "An essentially complete class of admissible decision functions," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 549-555.
- [16] ———, "Statistical decision functions," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 165-205.
- [17] D. V. WIDDER, *The Laplace Transform*, Princeton University Press, Princeton, 1941.