NOTE ON ESTIMATING ORDERED PARAMETERS

By Ester Samuel

Hebrew University

1. Introduction. We consider the problem of estimating a set of k real valued parameters, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$ where $\theta_i \varepsilon S$, $i = 1, \cdots, k$. Let $\mathbf{X}$ be the (usually vector valued) random variable with values $\mathbf{x}$, the distribution of which depends upon $\boldsymbol{\theta}$ and let $\boldsymbol{\delta} = \boldsymbol{\delta}(\mathbf{X}) = (\delta_1(\mathbf{X}), \cdots, \delta_k(\mathbf{X}))$ be an estimator of $\boldsymbol{\theta}$. Since $\boldsymbol{\theta}$ is known to belong to S^k, the k-fold Cartesian product of S, we shall restrict $\boldsymbol{\delta}$ to belong to S^k with probability one.

We assume that the loss incurred by saying $\boldsymbol{\delta}$ when the parameter is $\boldsymbol{\theta}$ is

$$(1) \qquad L(\boldsymbol{\delta}, \boldsymbol{\theta}) = \textstyle\sum_{i=1}^{k} \phi(|\delta_i - \theta_i|)$$

where $\phi(t)$, $t \geqq 0$, is a monotone increasing function.

The problem described above is usually called an estimation problem only if S is an interval. We shall however not put any restrictions on S except (to avoid trivialities) that it contains at least two elements. Thus, e.g., when S is finite we consider what is usually called a multidecision problem. We shall also allow randomized procedures, but in order not to complicate the notation we shall not introduce a special notation when $\boldsymbol{\delta}$ is randomized. Thus, in what follows, $\boldsymbol{\delta}$ should be interpreted to be the value of the estimator after the randomization experiment has been carried out.

Suppose now that $\boldsymbol{\theta}$ is known to belong to Ω, a subset of S^k. Is it then necessary for $\boldsymbol{\delta}$ to belong to Ω in order for $\boldsymbol{\delta}$ to be admissible? That is, must

$$(2) \qquad P(\boldsymbol{\delta} \,\varepsilon\, \Omega; \boldsymbol{\theta}) = 1 \qquad \text{for every } \boldsymbol{\theta} \,\varepsilon\, \Omega$$

in order for $\boldsymbol{\delta}$ to be admissible?

In this generality, the answer is known to be in the negative. Robbins in [2] considers the (nonsequential) compound decision problem where for $i = 1, \cdots, k$ one has observations X_i from a normal population with variance 1 and mean, $\theta_i \varepsilon \{-1, 1\}$, and the X_i's are independent. Thus here $\mathbf{X} = (X_1, \cdots, X_k)$, and S^k contains 2^k points. The only values of $\phi(t)$ of interest here are $\phi(0)$ and $\phi(2)$, which are taken to be 0 and 1 respectively. Suppose it is known that exactly one of the parameters θ_i equals 1 and the $k - 1$ others equal -1. Thus Ω contains the k points having one coordinate $+1$ and the others -1. In [2], p. 138, it is shown that for $k > 2$ the Bayes rule $\boldsymbol{\delta}$ with respect to the *a priori* distribution which assigns equal probability $1/k$ to each element of Ω takes the value $\boldsymbol{\delta} = (-1, \cdots, -1)$ with positive probability under every $\boldsymbol{\theta} \,\varepsilon\, \Omega$, and hence clearly fails to satisfy (2). Since this Bayes rule is essentially unique the rule obtained certainly is admissible for the restricted problem of deciding on $\boldsymbol{\theta} \,\varepsilon\, \Omega$. (This result is actually not too surprising. $\boldsymbol{\delta}(\mathbf{x})$ takes the value $(-1, \cdots, -1)$ when all x_i's are nearly equal. In that case assigning the value $+1$ to some

Received 31 December 1963; revised 2 November 1964.

δ_i has high (posterior) probability of causing a loss of 2, rather than the (certain) loss of 1 incurred by $\boldsymbol{\delta}(\mathbf{x}) = (-1, \cdots, -1)$.)

In this note we consider the restriction to the set

$$\Omega^* = \{\boldsymbol{\theta} : \boldsymbol{\theta} \varepsilon S^k, \theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_k\}.$$

This restriction is of interest e.g. when θ_i is the p_ith percentile point of some distribution function F and $0 \leqq p_1 \leqq \cdots \leqq p_k \leqq 1$, or when $\theta_i = F(t_i)$ and $-\infty < t_1 < t_2 < \cdots < t_k < \infty$, or whenever it is known *a priori* that the parameters under consideration are ordered as in Ω^*.

2. A theorem.

THEOREM 1. *Let the loss function be given in* (1) *with* $\phi(t)$, $t \geqq 0$ *a monotone increasing strictly convex function. Then the class of estimators* $\boldsymbol{\delta}$ *satisfying*

$$P(\boldsymbol{\delta} \varepsilon \Omega^*; \boldsymbol{\theta}) = 1 \qquad \textit{for every } \boldsymbol{\theta} \varepsilon \Omega^* \tag{3}$$

is essentially complete. If S *is an interval, the class is complete.*

PROOF. Suppose first $k = 2$. Let $\boldsymbol{\delta} = (\delta_1, \delta_2)$ be such that for some $\boldsymbol{\theta}^* \varepsilon \Omega^*$

$$P(\delta_1 > \delta_2; \boldsymbol{\theta}^*) > 0. \tag{4}$$

Define $\boldsymbol{\delta}^* = (\delta_1^*, \delta_2^*)$ by

$$\delta_1^* = \min(\delta_1, \alpha\delta_1 + (1 - \alpha)\delta_2), \qquad \delta_2^* = \max(\delta_2, (1 - \alpha)\delta_1 + \alpha\delta_2)$$

for some fixed $0 \leqq \alpha \leqq \frac{1}{2}$. Notice that $\boldsymbol{\delta} = \boldsymbol{\delta}^*$ whenever $\boldsymbol{\delta} \varepsilon \Omega^*$. Since for $\alpha = 0$ (3) holds for $\boldsymbol{\delta}^*$, the first part of the theorem follows for $k = 2$ if we show that for every $\boldsymbol{\theta} \varepsilon \Omega^*$

$$L(\boldsymbol{\delta}, \boldsymbol{\theta}) \geqq L(\boldsymbol{\delta}^*, \boldsymbol{\theta}). \tag{5}$$

When S is an interval (3) holds for $\boldsymbol{\delta}^*$ for *all* values of α satisfying $0 \leqq \alpha \leqq \frac{1}{2}$. Actually we shall show that whenever $\boldsymbol{\delta} \not\varepsilon \Omega^*$ (5) holds with strict inequality unless $\theta_1 = \theta_2$ and $\alpha = 0$. Hence, if (4) holds, $\boldsymbol{\delta}^*$ is a true improvement over $\boldsymbol{\delta}$ if we may chose some $\alpha > 0$, and the second part of the theorem follows for $k = 2$.

Let $\phi^*(t) = \phi(|t|)$, $-\infty < t < \infty$. Then by our assumptions ϕ^* is strictly convex and (5) is equivalent to

$$\phi^*(\delta_1 - \theta_1) + \phi^*(\delta_2 - \theta_2) \geqq \phi^*(\alpha\delta_1 + (1 - \alpha)\delta_2 - \theta_1) + \phi^*((1 - \alpha)\delta_1 + \alpha\delta_2 - \theta_2). \tag{6}$$

A well-known inequality for convex functions states

$$\phi^*(t) + \phi^*(s) \geqq \phi^*(t - u) + \phi^*(s + u) \qquad \text{when } t - s \geqq u > 0, \tag{7}$$

where (7) holds with strict inequality when $t - s > u$. Now (6) is obtained from (7) upon substituting $t = \delta_1 - \theta_1$, $s = \delta_2 - \theta_2$ and $u = (\delta_1 - \delta_2)(1 - \alpha)$, and clearly the strict inequality holds unless $\theta_1 = \theta_2$ and $\alpha = 0$.

Suppose now that $k > 2$. If $\boldsymbol{\delta}$ violates (3) for some $\boldsymbol{\theta}^*$, then there exist i, j, $1 \leqq i < j \leqq k$, such that $P(\delta_i > \delta_j; \boldsymbol{\theta}^*) > 0$. Then by the preceding argument (δ_i, δ_j) can be replaced by (δ_i', δ_j') satisfying $P(\delta_i' \leqq \delta_j'; \boldsymbol{\theta}) = 1$ for all $\boldsymbol{\theta} \varepsilon \Omega^*$, thereby either strictly decreasing the sum of the losses for the ith and jth component, or leaving it unchanged. After a finite number of steps we obtain a rule $\boldsymbol{\delta}^*$ satisfying (3) with losses satisfying (5). (5) is satisfied with strict inequality unless both (I): $\boldsymbol{\delta}$ violates (3) by satisfying $P(\boldsymbol{\delta} \varepsilon \Omega^*; \boldsymbol{\theta}) < 1$ only for vectors $\boldsymbol{\theta}$ having some coordinates equal, and (II): one is forced to choose $\alpha = 0$.

It should be noticed that whenever $\boldsymbol{\delta}^*$ dominates $\boldsymbol{\delta}$, the domination is in the strong sense; viz., for every $\mathbf{x}$ the *loss* of $\boldsymbol{\delta}^*$ is smaller than (or equal to) the *loss* of $\boldsymbol{\delta}$. Usual domination of decision functions considers *risks* only.

If the parametric family is such that for every measurable set A, $P(\mathbf{X} \varepsilon A; \boldsymbol{\theta}) > 0$ for some $\boldsymbol{\theta} \varepsilon \Omega^*$ implies $P(\mathbf{X} \varepsilon A; \boldsymbol{\theta}) > 0$ for every $\boldsymbol{\theta} \varepsilon \Omega^*$, then the above proof shows that, irrespective of the structure of S, the class of $\boldsymbol{\delta}$'s satisfying (3) is complete.

We remark that Katz in [1] considers the above situation with $k = 2$, $\Omega^* = \{0 \leqq \theta_1 \leqq \theta_2 \leqq 1\}$ where $\mathbf{X}$ is a vector of $2n$ independent Bernoulli random variables, n of which having parameter θ_1 and the other n having parameter θ_2. Our Theorem 1 and its proof are generalizations of Theorem 1 in [1].

An immediate question arising is whether the theorem remains true when $\phi(t)$, $t \geqq 0$, is assumed to be monotone increasing but not necessarily strictly convex. The general answer is in the negative, as we shall now show.

If S contains only 2 elements and ϕ is monotone increasing, the class satisfying (3) is always essentially complete, or complete, respectively. This follows since then only two values of ϕ, viz. at 0 and at some point $\lambda > 0$ are of interest, and for any two such given values with $\phi(0) < \phi(\lambda)$ there exists a convex function ϕ' such that $\phi'(0) = \phi(0)$ and $\phi'(\lambda) = \phi(\lambda)$ and Theorem 1 is applicable.

Suppose now that S contains 3 elements $\lambda_1 < \lambda_2 < \lambda_3$, and for definiteness let $\lambda_2 - \lambda_1 \leqq \lambda_3 - \lambda_2$. We may without loss of generality assume $\phi(0) = 0$. Thus let

$$(8) \quad \phi(0) = 0, \quad \phi(\lambda_2 - \lambda_1) = a, \quad \phi(\lambda_3 - \lambda_2) = b, \quad \phi(\lambda_3 - \lambda_1) = c,$$

where $0 < a \leqq b \leqq c$. Suppose $k = 2$. We shall show that in certain cases $(\delta_1, \delta_2) = (\lambda_2, \lambda_1)$ can be admissible even when $\boldsymbol{\theta} \varepsilon \Omega^*$. Table 1 indicates the losses suffered by saying $(\delta_1, \delta_2) = (\lambda_1, \lambda_2)$ or (λ_2, λ_1) respectively, when the true value is (θ_1, θ_2). It is immediately seen that unless $c < a + b$ the first row in Table 1 dominates the second.

TABLE 1

(δ_1, δ_2)	(θ_1, θ_2)					
	(λ_1, λ_1)	(λ_1, λ_2)	(λ_1, λ_3)	(λ_2, λ_2)	(λ_2, λ_3)	(λ_3, λ_3)
(λ_1, λ_2)	a	0	b	a	$a + b$	$b + c$
(λ_2, λ_1)	a	$2a$	$a + c$	a	c	$b + c$

Let S be countable and denote its elements by $\lambda_1, \lambda_2, \cdots$. In order to show that a rule is admissible we shall show that it is a Bayes rule with respect to some *a priori* distribution.

$$P(\theta_1 = \lambda_i, \theta_2 = \lambda_j) = p(\lambda_i, \lambda_j) = 0 \qquad \text{if } \lambda_i > \lambda_j$$
$$= \text{positive} \qquad \text{otherwise.}$$

Let $\mathbf{x}$ be a value such that $P(\mathbf{X} = \mathbf{x}; \boldsymbol{\theta}) = p(\mathbf{x} \mid \theta_1, \theta_2) > 0$ for some $\boldsymbol{\theta} \varepsilon \Omega^*$. In order that the Bayes rule with respect to the given *a priori* distribution be $(\delta_1(\mathbf{x}), \delta_2(\mathbf{x})) = (\lambda_2, \lambda_1)$ it is sufficient that

$$(9) \qquad \sum_i \phi(|\lambda_2 - \lambda_i|)g_1(\lambda_i \mid \mathbf{x}) < \sum_i \phi(|\lambda_j - \lambda_i|)g_1(\lambda_i \mid \mathbf{x}) \qquad \text{for all } j \neq 2$$

and

$$(10) \qquad \sum_i \phi(|\lambda_1 - \lambda_i|)g_2(\lambda_i \mid \mathbf{x}) < \sum_i \phi(|\lambda_j - \lambda_i|)g_2(\lambda_i \mid \mathbf{x}) \quad \text{for all } j \neq 1,$$

where $g_r(\lambda_i \mid \mathbf{x})$ $r = 1, 2$ is proportional to the posterior distribution of θ_r given $\mathbf{x}$.

Thus in the example where S consists of the 3 elements $\lambda_1 < \lambda_2 < \lambda_3$ we find that if $p(\mathbf{x} \mid \lambda_i, \lambda_j)$ is a constant for $1 \leqq i \leqq j \leqq 3$ and $p(\theta_i, \theta_j)$ is given by Table 2 where the margins are proportional to g_1 and g_2, respectively, then (9) and (10) hold, i.e. $(\delta_1(\mathbf{x}), \delta_2(\mathbf{x})) = (\lambda_2, \lambda_1)$, both when $a = b = c = 1$ and when $a = 99$, $b = 100$, $c = 101$. This shows that the convexity condition in Theorem 1 cannot generally be weakened. (Obviously $p(x \mid \lambda_i, \lambda_j)$ was chosen to be a constant only to simplify the computations.)

TABLE 2

θ_1	θ_2			
	λ_1	λ_2	λ_3	g_1
λ_1	.35	.01	.01	.37
λ_2	0	.30	.10	.40
λ_3	0	0	.23	.23
g_2	.35	.31	.34	1

From the above counterexample one can easily obtain counterexamples for S of any finite or denumerable cardinality. The writer has no counterexamples when S is an interval, but she believes that such counterexamples can be obtained.

We remark that whenever the class of extended Bayes rules for our problem is complete then Theorem 1 remains valid if (1) is replaced by

$$(11) \qquad L^*(\boldsymbol{\delta}, \boldsymbol{\theta}) = \sum_{i=1}^{k} [d_i\phi(|\delta_i - \theta_i|) + e_i]$$

with $d_i > 0$, where $\phi(t)$ for $t \geqq 0$ is a monotone increasing strictly convex function. (This is e.g. the case when S is finite, or when S is a closed bounded interval and $\mathbf{X}$ has a density which is continuous in $\boldsymbol{\theta}$.) This follows since $\boldsymbol{\delta}$ is a Bayes rule with respect to some prior distribution for loss function (1) if and only if it is a Bayes rule with respect to the same prior distribution for loss function (11).

It thus follows that *whenever S contains only two elements and the loss on every component i is greater for a wrong decision than for a correct one, the class of decision functions $\boldsymbol{\delta}$ satisfying* (3) *is essentially complete*, since in that case the loss can always be written as (11).

(11) is a particular case of

$$L(\boldsymbol{\delta}, \boldsymbol{\theta}) = \sum_{i=1}^{k} \phi_i(|\delta_i - \theta_i|) \tag{12}$$

where $\phi_i(t)$ for $t \geqq 0$ are monotone increasing strictly convex functions' $i = 1, \cdots, k$. One may be tempted to believe that for finite S the class of rules satisfying (3) is essentially complete also when the loss has structure (12). That this conclusion is false follows if we consider S containing 3 elements, and let $a_1 = 10$, $b_1 = 40$, $c_1 = 100$, $a_2 = 10$, $b_2 = 20$, $c_2 = 31$ where these values are defined for ϕ_1 and ϕ_2 by (8). Then for $p(\mathbf{x} \mid \boldsymbol{\theta})$ equal a constant and the prior distribution of Table 3, it follows that (9) and (10) hold, with ϕ replaced by ϕ_1 and ϕ_2, respectively.

TABLE 3

θ_1	θ_2			
	λ_1	λ_2	λ_3	g_1
λ_1	.55	.01	.01	.57
λ_2	0	.01	.01	.02
λ_3	0	0	.41	.41
g_2	.55	.02	.43	1

Acknowledgment. The author is indebted to the referee for pointing out some errors in an earlier version of this note.

REFERENCES

[1] KATZ, MORRIS W. (1963). Estimating ordered probabilities. *Ann. Math. Statist.* **34** 967–972.

[2] ROBBINS, HERBERT (1951). Asymptotic subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 131–148. Univ. of California Press.