The Annals of Mathematical Statistics
1968, Vol. 39, No. 3, 1030–1039

ON ESTIMATING MONOTONE PARAMETERS

BY TIM ROBERTSON[1] AND PAUL WALTMAN[2]

University of Iowa

1. Introduction and summary. Suppose for each $i = 1, 2, \cdots, k$ the random variable X_i has density function $f(x; \theta_i)$ where each of the parameters $\theta_1, \theta_2, \cdots, \theta_k$ is known to belong to some connected set Θ of real numbers. Independent random samples are taken from the distributions of $X_1, X_2, \cdots, X_k$ and we wish to find estimates $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$ of $\theta_1, \theta_2, \cdots, \theta_k$ which satisfy:

$$\hat{\theta}_1 \geqq \hat{\theta}_2 \geqq \cdots \geqq \hat{\theta}_k. \tag{1.1}$$

Brunk [3] considered such a problem when $f(x; \theta)$ belongs to a certain exponential family of distributions which includes the binomial, the normal with fixed mean and variable standard deviation, the normal with fixed standard deviation and variable mean, and the Poisson distributions. A discussion of the history of this type of problem is given by Brunk [4].

In this paper we assume that the density function $f(x; \theta)$ has certain properties and develop a procedure for finding the restricted estimates. The above mentioned densities have those properties as do certain others. One in particular, not covered by Brunk's formulation, the bilateral exponential distribution (i.e. $f(x; \theta) = \frac{1}{2}e^{-|x-\theta|}$), is considered in Section Four.

In Section Two we list those assumptions, describe our procedure for finding restricted estimates and prove that they are maximum likelihood estimates. In Section Three we give a representation theorem for our estimates and a theorem which implies that they are consistent in the special cases which we consider. In Section Five we describe an alternate method for obtaining restricted estimates and in Section Six we discuss another special case not considered by Brunk [3], which does not satisfy our regularity assumptions but for which the procedure described in Section Two clearly works.

2. A procedure for obtaining restricted maximum likelihood estimates. Suppose that n_i items are drawn from the distribution of X_i and that they are de-denoted by $X_{i1}, X_{i2}, \cdots, X_{in_i}$. We assume that the family of functions $\{f(x; \theta); \theta \varepsilon \Theta\}$ has the following properties:

(2.1) $f(x; \theta)$ has support S which is the same for all $\theta \varepsilon \Theta$,

(2.2) for each x in $S, f(x; \theta)$ is a continuous function of θ on Θ,

(2.3) if $x_1, x_2, \cdots, x_n$ are any members of S then there exists a number

Received 25 September 1967.

[1] The work of this author was partially supported by funds which were allocated to the Graduate College of the University of Iowa thorugh a USPHS Biomedical Grant.

[2] Research supported by NSF Grant GP-5599.

M in Θ such that if θ and θ' are in the closed interval with endpoints θ and M then $\prod_{i=1}^{n} f(x_i; \theta') \geqq \prod_{i=1}^{n} f(x_i; \theta)$ (i.e. the likelihood function $L(\theta; x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)$ is unimodal with (not necessarily unique) mode M),

(2.4) if $x_1, x_2, \cdots, x_n, y_1, y_2, \cdots, y_m$ are in S, if $M_x(M_y)$ is the mode of $L(\theta; x_1, x_2, \cdots, x_n)$ $(L(\theta; y_1, y_2, \cdots, y_m))$ and if M_{xy} is the mode of $L(\theta; x_1, x_2, \cdots, x_n, y_1, y_2, \cdots, y_m)$ then M_{xy} is between M_x and M_y.

It should be pointed out that although we are not assuming in (2.3) that the mode of $L(\theta; x_1, x_2, \cdots, x_n)$ is unique, we are assuming in (2.4) that we have a procedure, as in Section Four, for uniquely selecting one of these modes, M_x, and that this procedure yields a Borel measurable function of $x_1, x_2, \cdots, x_n$. On the other hand if we assume in (2.3) that the mode of $L(\theta; x_1, x_2, \cdots, x_n)$ is unique then this implies (2.4). As noted by Boswell [2], Van Eeden [6] obtained results similar to ours by assuming that the mode of $L(\theta; x_1, x_2, \cdots, x_n)$ is unique. By assuming (2.4) instead of the uniqueness of modes we include as a special case the bilateral exponential distribution discussed in Section Four. Furthermore, Van Eeden discusses convergence in probability and Theorem 3.2 gives convergence with probability one.

As an example of a situation where (2.1)–(2.4) are satisfied suppose, as in [1], that $\Theta = [0, 1]$, $S = \{0, 1\}$ and $f(x; \theta) = \theta^x(1 - \theta)^{1-x}$ for $x \in S$. Clearly properties (2.1)–(2.3) are satisfied where $M = n^{-1} \cdot \sum_{i=1}^{n} x_i = \bar{x}$ in (2.3). Also in (2.4) $M_{xy} = (n\bar{x} + m\bar{y}) \div (m + n)$ is between $\bar{x}$ and $\bar{y}$.

Let M_i be the mode of $L(\theta; x_{i1}, x_{i2}, \cdots, x_{in_i})$ and let S_k be that subset of Euclidean k-space defined by:

$$S_k = \{(\alpha_1, \alpha_2, \cdots, \alpha_k); \alpha_i \in \Theta, \alpha_1 \geqq \alpha_2 \geqq \cdots \geqq \alpha_k\}.$$

We wish to find a point $(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ in S_k which maximizes

$$L(\alpha_1, \alpha_2, \cdots, \alpha_k) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} f(x_{ij}; \alpha_i).$$

LEMMA 2.1. *There is a maximizing point* $[\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ *in* S_k.

PROOF. Let $L_i(\theta) = \prod_{j=1}^{n_i} f(x_{ij}; \theta)$ and

$$S_k^* = \{(\alpha_1, \alpha_2, \cdots, \alpha_k); \alpha_i \in \Theta, \max(M_1, M_2, \cdots, M_k) \geqq \alpha_1 \geqq \alpha_2 \geqq \cdots \geqq \alpha_k \geqq \min(M_1, M_2, \cdots, M_k)\}.$$

Using (2.3) and the assumption that Θ is connected it is easy to see that if $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_k) \in S_k - S_k^*$ then there is a point $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_k)$ in S_k^* such that $L(\boldsymbol{\beta}) \geqq L(\boldsymbol{\alpha})$. Hence we may restrict our attention to S_k^* and since this set is closed and bounded and L is continuous by (2.2) it follows that a maximizing point exists.

A maximum likelihood estimate of $(\theta_1, \theta_2, \cdots, \theta_k)$ in S_k may be obtained as follows: If $M_1 \geqq M_2 \geqq \cdots \geqq M_k$ then $\hat{\theta}_i = M_i$, $i = 1, 2, \cdots, k$. On the other hand if for some i we have $M_i < M_{i+1}$ then the ith and $i + 1$st samples are pooled and M_i^* is the mode of $\prod_{k=i}^{i+1} \prod_{j=1}^{n_k} f(x_{kj}; \theta)$. Furthermore $M_j^* = M_j$ for $j < i$ and $M_j^* = M_{j+1}$ for $j > i$. This procedure is repeated until a set of monotone non-increasing set of modes is obtained. Then, for each j, $\hat{\theta}_j$ is equal to that one of the final set of modes to which $x_{j1}, x_{j2}, \cdots, x_{jn_j}$ contributed. We note that this procedure is very similar to the one described in [1].

THEOREM 2.1. *If $\hat{\theta}_1 \geqq \hat{\theta}_2 \geqq \cdots \geqq \hat{\theta}_k$ are obtained from the sample items by the procedure described above then*

$$L(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k) \geqq L(\alpha_1, \alpha_2, \cdots, \alpha_k)$$

for all $(\alpha_1, \alpha_2, \cdots, \alpha_k)$ in S_k.

PROOF. (Induction on k). If $k = 1$ then by (2.3) M_1 is a maximum likelihood estimate of θ_1 and the result is clear. Suppose the result is true for $k = H$ and that we have $H + 1$ samples. If $M_1 \geqq M_2 \geqq \cdots \geqq M_{H+1}$ then $(M_1, M_2, \cdots, M_{H+1})$ provides a maximizing point as asserted since this is an unrestricted maximizing point by (2.3).

Suppose that for some i we have $M_i < M_{i+1}$. Let $\hat{\theta}_1 \geqq \hat{\theta}_2 \geqq \cdots \geqq \hat{\theta}_{H+1}$ be derived from the sample items by the procedure described above. It follows from the induction hypothesis that $L(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{H+1}) \geqq L(\alpha_1, \alpha_2, \cdots, \alpha_{H+1})$ for all $(\alpha_1, \alpha_2, \cdots, \alpha_{H+1}) \varepsilon S_{H+1}$ such that $\alpha_i = \alpha_{i+1}$. By Lemma 2.1 there exists a maximizing point $(\theta_1^*, \theta_2^*, \cdots, \theta_{H+1}^*) \varepsilon S_{H+1}$. Then using the assumptions that if θ' is between θ and M_j then $L_j(\theta') \geqq L_j(\theta)$ and (2.4), and by considering the three cases $\theta_{i+1}^* < \theta_i^* \leqq M_{i+1}$, $\theta_{i+1}^* \leqq M_i < M_{i+1} \leqq \theta_i^*$ and $M_i \leqq \theta_{i+1}^* < \theta_i^*$ it is easy to see that there is a maximizing point $(\bar{\theta}_1, \bar{\theta}_2, \cdots, \bar{\theta}_{H+1})$ such that $\bar{\theta}_i = \bar{\theta}_{i+1}$. Hence we conclude that

$$L(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{H+1}) \geqq L(\bar{\theta}_1, \bar{\theta}_2, \cdots, \bar{\theta}_{H+1}) \geqq L(\alpha_1, \alpha_2, \cdots, \alpha_{H+1})$$

for all $(\alpha_1, \alpha_2, \cdots, \alpha_{H+1}) \varepsilon S_{H+1}$. This completes the argument.

3. A representation theorem for the estimates and consistency. If $1 \leqq R \leqq S \leqq k$ then let $M(R, S)$ denote the mode of $\prod_{i=R}^{S} \prod_{j=1}^{n_i} f(x_{ij}; \theta)$. Let $(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ be the maximum likelihood estimates given by Theorem 2.1.

THEOREM 3.1. *For $j = 1, 2, \cdots, k$ we have*

$$\hat{\theta}_j = \min_{1 \leqq R \leqq j} \max_{R \leqq S \leqq k} M(R, S) = \min_{1 \leqq R \leqq j} \max_{j \leqq S \leqq k} M(R, S)$$
$$= \max_{j \leqq S \leqq k} \min_{1 \leqq R \leqq S} M(R, S) = \max_{j \leqq S \leqq k} \min_{1 \leqq R \leqq j} M(R, S).$$

PROOF. First note that if $M_1 \geqq M_2 \geqq \cdots \geqq M_k$, then for any R such that $1 \leqq R \leqq k$ we have $M_R \geqq M(R, R + 1) \geqq \cdots \geqq M(R, k)$ by (2.4). Hence for any j we have

$$\min_{1 \leqq R \leqq j} \max_{R \leqq S \leqq k} M(R, S) = \min_{1 \leqq R \leqq j} M_R = M_j = \hat{\theta}_j.$$

We prove the theorem by induction on k. The result is clear if $k = 1$. Suppose the theorem is true for $k = H$ and that we have $H + 1$ samples with corresponding modes $M_1, M_2, \cdots, M_{H+1}$.

Since we have already considered the case where $M_1 \geqq M_2 \geqq \cdots \geqq M_{H+1}$

suppose that $M_i < M_{i+1}$ for some i. Pool the ith and $i + 1$st samples obtaining H samples with modes $M_1^*, M_2^*, \cdots, M_H^*$ where $M_i^* = M(i, i + 1)$, $M_j^* = M_j$, $j < i$, and $M_j^* = M_{j+1}$, $j > i$.

Let

$$\theta_j^* = \min_{1 \leqq R \leqq j} \max_{R \leqq S \leqq H} M^*(R, S).$$

Then by the induction hypothesis

$$\hat{\theta}_j = \theta_j^*, \qquad j < i,$$
$$\hat{\theta}_i = \hat{\theta}_{i+1} = \theta_i^*,$$
$$\hat{\theta}_j = \theta_{j-1}^*, \qquad j > i + 1.$$

Let

$$\bar{\theta}_j = \min_{1 \leqq R \leqq j} \max_{R \leqq S \leqq H+1} M(R, S), \qquad j = 1, 2, \cdots, H + 1.$$

Note that if $R < i$ and if $\max_{R \leqq S \leqq H+1} M(R, S) = M(R, i)$ then $M(R, i) \geqq M(R, i + 1)$. It follows from (2.4) that

$$M(R, i) \geqq M(R, i + 1) \geqq M_{i+1} > M_i.$$

Using the fact that $M(R, i) > M_i$ and (2.4) we conclude that $M(R, i - 1) \geqq M(R, i) = \max_{R \leqq S \leqq H+1} M(R, S)$.

It follows that if $R < i$ then we can assume that $\max_{R \leqq S \leqq H+1} M(R, S) = M(R, S_R)$ where $S_R \neq i$. Clearly $M_i \leqq M(i, i + 1) \leqq M_{i+1}$ so that

$$\max_{R \leqq S \leqq H+1} M(R, S) = \max_{R \leqq S \leqq H} M^*(R, S), \qquad R \leqq i. \tag{3.1}$$

Hence if $j \leqq i$ then

$$\bar{\theta}_j = \min_{1 \leqq R \leqq j} \max_{R \leqq S \leqq H+1} M(R, S) = \min_{1 \leqq R \leqq j} \max_{R \leqq S \leqq H} M^*(R, S) = \hat{\theta}_j.$$

Now suppose $\max_{i+1 \leqq S \leqq H+1} M(i + 1, S) = M(i + 1, S_0)$. Then $M(i + 1, S_0) \geqq M_{i+1} > M_i$. Also $M(i + 1, S_0) \geqq M(i + 1, S)$ for all $S \geqq i + 1$. Combining those observations with (2.4) we conclude that $M(i + 1, S_0) \geqq M(i, S)$ for all $S \geqq i$.

$$\max_{i+1 \leqq S \leqq H+1} M(i + 1, S) \geqq \max_{i \leqq S \leqq H+1} M(i, S) \geqq \min_{1 \leqq R \leqq i} \max_{R \leqq S \leqq H+1} M(R, S). \tag{3.2}$$

Now using (3.1) and (3.2) we conclude that for $j \geqq i + 1$

$$\begin{aligned}\bar{\theta}_j &= \min [\min_{1 \leqq R \leqq i} \max_{R \leqq S \leqq H+1} M(R, S), \max_{i+1 \leqq S \leqq H+1} M(R, S), \\ &\qquad \min_{i+2 \leqq R \leqq j} \max_{R \leqq S \leqq H+1} M(R, S)] \\ &= \min [\min_{1 \leqq R \leqq i} \max_{R \leqq S \leqq H+1} M^*(R, S), \\ &\qquad \min_{i+2 \leqq R \leqq j} \max_{R \leqq S \leqq H+1} M^*(R - 1, S - 1)]\end{aligned}$$

$$= \min [\min_{1 \leqq R \leqq i} \max_{R \leqq S \leqq H+1} M^*(R, S),$$
$$\min_{i+1 \leqq R \leqq j-1} \max_{R \leqq S \leqq H} M^*(R, S)]$$
$$= \theta^*_{j-1} = \hat{\theta}_j .$$

The other parts of the theorem follow similarly.

The unrestricted maximum likelihood estimate M_i of θ_i is generally a consistent estimate of θ_i. In a sense (see Section Five) the restricted estimate $(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ is closer to $(\theta_1, \theta_2, \cdots, \theta_k)$ when $\theta_1 \geqq \theta_2 \geqq \cdots \geqq \theta_k$ than $(M_1, M_2, \cdots, M_k)$. Hence it would seem reasonable that if $(M_1, M_2, \cdots, M_k)$ is consistent then so is $(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ provided, of course, that $\theta_1 \geqq \theta_2 \geqq \cdots \geqq \theta_k$.

THEOREM 3.2. *Let* $m = \min (n_1, n_2, \cdots, n_k)$. *If* $\theta_1 \geqq \theta_2 \geqq \cdots \geqq \theta_k$ *and if*

$$\lim_{m\to\infty} \cdot \sum_{i=1}^k |M_i - \theta_i| = 0$$

with probability one then

$$\lim_{m\to\infty} \cdot \sum_{i=1}^k |\hat{\theta}_i - \theta_i| = 0$$

with probability one.

PROOF. The notation here is slightly inconsistent in that we are now thinking of M_i and $\hat{\theta}_i$ as random variables and prior to this point they were numbers obtained from the sample items. We will show that

$$[\lim_{m\to\infty} \cdot \sum_{i=1}^k |M_i - \theta_i| = 0] \subset [\lim_{m\to\infty} \cdot \sum_{i=1}^k |\hat{\theta}_i - \theta_i| = 0]$$

and the desired conclusion follows. Suppose $\lim_{m\to\infty} \cdot \sum_{i=1}^k |M_i - \theta_i| = 0$ and ϵ is arbitrary. If all of the θ_i's are equal then it follows from (2.4) and Theorem 3.1 that

$$\sum_{i=1}^k |\hat{\theta}_i - \theta_i|$$
$$\leqq k \cdot \max_{1 \leqq i \leqq k} |\hat{\theta}_i - \theta_i| \leqq k \cdot \max_{1 \leqq i \leqq k} |M_i - \theta_i| \leqq k \cdot \sum_{i=1}^k |M_i - \theta_i|$$

and the conclusion follows. If not all of the θ_i's are equal then we can assume, without loss of generality, that

$$\epsilon < \tfrac{1}{2} \min_{\theta_i > \theta_{i+1}} [\theta_i - \theta_{i+1}].$$

Suppose we are dealing with a point in $[\lim_{m\to\infty} \cdot \sum_{i=1}^k |M_i - \theta_i| = 0]$ and choose M such that $m \geqq M$ implies that

$$\sum_{i=1}^k |M_i - \theta_i| < \epsilon/K < \tfrac{1}{2} \min_{\theta_i > \theta_{i+1}} [\theta_i - \theta_{i+1}].$$

Suppose j is arbitrary, A is the smallest integer such that $\theta_A = \theta_j$ and B is the largest integer such that $\theta_B = \theta_j$. Then if $m \geqq M$ we have $M_{i_1} > M_{i_2} > M_{i_3}$ for $i_1 < A$, $A \leqq i_2 \leqq B$ and $i_3 > B$, since $\epsilon < \tfrac{1}{2} \min_{\theta_i > \theta_{i+1}} [\theta_i - \theta_{i+1}]$.

It follows from (2.4) that for $R < A$ we have

$$\max_{R \leqq S \leqq k} M(R, S) \geqq \max_{A \leqq S \leqq k} M(A, S)$$

so that

$$\hat{\theta}_j = \min_{1 \leqq R \leqq j} \max_{R \leqq S \leqq k} M(R, S) = \min_{A \leqq R \leqq j} \max_{R \leqq S \leqq k} M(R, S).$$

Furthermore if $A \leqq R \leqq j$ and $S > B$ then

$$M(R, B) \geqq M(R, S) \geqq M(B + 1, S)$$

by (2.4) so that

$$\max_{R \leqq S \leqq k} M(R, S) = \max_{R \leqq S \leqq B} M(R, S)$$

so that $\hat{\theta}_j = M(R_0, S_0)$ where $A \leqq R_0 \leqq S_0 \leqq B$. However, if $A \leqq i \leqq B$ then $\theta_j - \epsilon/k \leqq M_i \leqq \theta_j + \epsilon/k$ so that using (2.4), Theorem 3.1 and the above we conclude that for $m \geqq M$ we have

$$\sum_{j=1}^{k} |\hat{\theta}_j - \theta_j| < \epsilon$$

since j was arbitrary. This completes the argument.

4. The bilateral exponential distribution. Perhaps the most interesting case not covered by Brunk [3] or Van Eeden [6] is when $\Theta = S = (-\infty, \infty)$ and $f(x; \theta) = \frac{1}{2}e^{-|x-\theta|}$. In this case it is clear that properties (2.1)–(2.3) are satisfied where M in (2.3) is a median of the numbers $x_1, x_2, \cdots, x_n$. However if n is even then this mode is not unique and we must be careful in our selection of M if (2.4) is to be satisfied. When n is even we select M to be the average of the two middle items.

THEOREM 4.1. *If* $f(x; \theta) = \frac{1}{2}e^{-|x-\theta|}$ *and the mode of* $L(\theta; x_1, x_2, \cdots, x_n)$ *is defined to be* the *median of* $x_1, x_2, \cdots, x_n$, *as above, then* (2.4) *is satisfied.*

PROOF. This result seems intuitively obvious and in fact the only difficulty arises when both sample sizes are even. Suppose $x_1 \leqq x_2 \leqq \cdots \leqq x_{2n}$ ($y_1 \leqq y_2 \leqq \cdots \leqq y_{2m}$) has median $M_x(M_y)$. Let the pooled sample be denoted by $z_1 \leqq z_2 \leqq \cdots \leqq z_{2m+2n}$. Then

$$M_x = \tfrac{1}{2}(x_n + x_{n+1}), \qquad M_y = \tfrac{1}{2}(y_m + y_{m+1}), \qquad M_{xy} = \tfrac{1}{2}(z_{m+n} + z_{m+n+1}).$$

We can assume, without loss of generality, that $M_x \leqq M_y$. Suppose $M_{xy} < M_x$. Then $M_{xy} < x_{n+1}$ and $M_{xy} < y_{m+1}$ and since there are $m + n$ z's no bigger then M_{xy} we conclude that $x_n \leqq M_{xy}$ and $y_m \leqq M_{xy}$. Hence $z_{m+n} = \max(x_n, y_m)$ and $z_{m+n+1} = \min(x_{n+1}, y_{m+1})$. Now since $M_{xy} < M_x \leqq M_y$ we have either $z_{m+n} = x_n$ and $z_{m+n+1} = y_{m+1}$ or $z_{m+n} = y_m$ and $z_{m+n+1} = x_{n+1}$. In the first case

$$M_{xy} = \tfrac{1}{2}(x_n + y_{m+1}) \geqq \tfrac{1}{2}(y_m + y_{m+1}) = M_y$$

which is a contradiction. In the second case

$$M_{xy} = \tfrac{1}{2}(y_m + x_{n+1}) \geqq \tfrac{1}{2}(x_n + x_{n+1}) = M_x$$

which is also a contradiction. A similar contradiction can be concluded if $M_x \leqq M_y < M_{xy}$. If either sample size is odd a similar argument can be made. This completes the proof.

Using Theorems 2.1, 3.1 and 4.1 we infer the following result.

THEOREM 4.2. *If* $f(x;\theta) = \frac{1}{2}e^{-|x-\theta|}$ *and if* $M(R, S)$ *denotes the median of the items in the* Rth *through* Sth *samples pooled then* $\hat{\theta}_1 \geqq \hat{\theta}_2 \geqq \cdots \geqq \hat{\theta}_k$, *where*

$$\hat{\theta}_j = \min_{1 \leqq R \leqq j} \max_{R \leqq S \leqq k} M(R, S) = \max_{j \leqq S \leqq k} \min_{1 \leqq R \leqq S} M(R, S),$$

provide maximum likelihood estimates of $\theta_1, \theta_2, \cdots, \theta_k$ *subject to the restriction* (1.1).

If $f(x;\theta) = \frac{1}{2}e^{-|x-\theta|}$ then it follows from the Glivenko-Cantelli Theorem that $\lim_{n_i \to \infty} M_i = \theta_i$ with probability one. This together with Theorem 3.2 implies:

THEOREM 4.3. *Under the hypotheses of Theorem* 4.2 *we have*

$$\lim_{\min n_i \to \infty} \sum_{i=1}^{k} |\hat{\theta}_i - \theta_i| = 0$$

with probability one.

5. Conditional expectations given σ-lattices. Define the measurable space (Ω, φ) by $\Omega = \{1, 2, \cdots, k\}$ and φ is the collection of all subsets of Ω. We shall restrict our attention to totally finite measures on φ. Let $\mathcal{L}$ be the σ-lattice (cf. [4]) of left subintervals of Ω(i.e. $\mathcal{L} = \{\phi, \{1\}, \{1, 2\}, \cdots, \Omega\}$). Then any k-tuple $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_k)$ of real numbers can be thought of as a function on Ω and $\alpha_1 \geqq \alpha_2 \geqq \cdots \geqq \alpha_k$ if and only if α is $\mathcal{L}$-measurable (i.e. $[\alpha > r] \varepsilon \mathcal{L}$ for all real numbers r). Hence S_k in Section Two is a collection of $\mathcal{L}$-measurable functions on Ω.

Now let $M = (M_1, M_2, \cdots, M_k)$ be the unrestricted maximum likelihood estimate of $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$ which is given by (2.3) and let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ be the restricted estimate given by Theorem 2.1.

THEOREM 5.1. *There exists a measure* μ *on* φ *with the property that* $\hat{\theta} = E_\mu(M \mid \mathcal{L})$.

PROOF. (Induction on the number k of elements in Ω.) If $k = 1$ then $\hat{\theta} = M = E_\mu(M \mid \mathcal{L})$ for any measure μ. Suppose the theorem is true for $k = H$ and that we have $H + 1$ samples. If $M_1 \geqq M_2 \geqq \cdots \geqq M_{H+1}$ then $\hat{\theta} = M = E_\mu(M \mid \mathcal{L})$ for any μ.

Assume that for some i we have $M_i < M_{i+1}$. Let $\Omega^* = \{1, 2, \cdots, H\}$, φ^* be the collection of all subsets of Ω^* and $\mathcal{L}^*$ be the collection of left subintervals of Ω^*. Define the function $M^* = (M_1^*, M_2^*, \cdots, M_H^*)$ on Ω^* by:

$$\begin{aligned} M_j^* &= M_j, & j &< i,\\ &= M(i, i+1), & j &= i,\\ &= M_{j+1}, & j &> i. \end{aligned}$$

It follows from the induction hypothesis and the procedure by which $\hat{\theta}$ is derived from M that there exists a measure μ^* on φ^* such that $\theta^* = E_{\mu^*}(M^* \mid \mathcal{L}^*)$ where

$$\begin{aligned} \hat{\theta}_j &= \theta_j^*, & j &< i,\\ &= \theta_i^*, & j &= i, i+1,\\ &= \theta_{j-1}^*, & j &> i+1. \end{aligned}$$

Now by (2.4), $M(i, i+1)$ is between M_i and M_{i+1} so that there exists an α between 0 and 1 such that $M(i, i+1) = \alpha M_i + (1-\alpha)M_{i+1}$.

Define the measure μ on $\mathfrak{A}$ by:

$$\begin{aligned} \mu(\{j\}) &= \mu^*(\{j\}), & j &< i, \\ &= \alpha\mu^*(\{i\}), & j &= i, \\ &= (1-\alpha)\mu^*(\{i\}), & j &= i+1, \\ &= \mu^*(\{j-1\}), & j &> i+1. \end{aligned}$$

Certainly $\hat{\theta}$ is $\mathfrak{L}$-measurable. To complete the argument for $\hat{\theta} = E(M \mid \mathfrak{L})$ it is sufficient to show that $\hat{\theta}$ has the following properties:

$$\int_L (M - \hat{\theta})\, d\mu \leqq 0 \quad \text{for all} \quad L \varepsilon \mathfrak{L} \tag{5.1}$$

and

$$\int_B (M - \hat{\theta})\, d\mu = 0 \quad \text{for all} \quad B \varepsilon \hat{\theta}^{-1}(\beta) \tag{5.2}$$

where β denotes the collection of Borel subsets of the real line (cf. [5]). To verify (5.1) suppose $L = \{1, 2, \cdots, h\}$. If $h < i$ then

$$\int_L (M - \hat{\theta})\, d\mu = \int_L (M^* - \theta^*)\, d\mu^* \leqq 0$$

since $L \varepsilon \mathfrak{L}^*$. If $h \geqq i + 1$ then

$$\int_L (M - \hat{\theta})\, d\mu = \int_{\{1,2,\cdots,h-1\}} (M^* - \theta^*)\, d\mu^* \leqq 0.$$

If $h = i$ and if $\int_L (M - \hat{\theta})\, d\mu > 0$ then since

$$\begin{aligned} \int_L (M - \hat{\theta})\, d\mu &= \int_{\{1,2,\cdots,i-1\}} (M^* - \theta^*)\, d\mu^* + (M_i - \theta_i^*)\alpha\mu^*(\{i\}) \\ &= \int_{\{1,2,\cdots,i\}} (M^* - \theta^*)\, d\mu^* - (M_{i+1} - \theta_i^*)(1-\alpha)\mu^*(\{i\}) \end{aligned}$$

we must have

$$M_i - \theta_i^* > 0 \qquad \text{and} \qquad M_{i+1} - \theta_i^* < 0.$$

Hence $M_{i+1} < M_i$ which is a contradiction. (5.2) follows similarly and this completes the argument.

For example if we let

$$f(x; \theta) = (2\pi)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(x - \theta)^2],$$

then $M_i = \bar{X}_i = [n_i]^{-1} \sum_{j=1}^{n_i} X_{ij}$ and the choice of μ which will give $\hat{\theta}$ is $\mu(\{i\}) = n_i \cdot [\sum_{j=1}^{k} n_i]^{-1}$ (cf. [3]). (Note that $\mu(\{i\}) = n_i$ gives the same estimate.)

In the case of the bilateral exponential distribution it seems impossible to write the choice of μ which gives $\hat{\theta}$ as a "nice" function of the sample items as in the normal case. In view of the apparent difficulty in calculating the restricted estimates in this case, especially when the sample sizes are large, one might be tempted to use a simple μ such as $\mu(\{i\}) = 1, i = 1, 2, \cdots, k$, to obtain restricted estimates. Also $\mu(\{i\}) = n_i$ might be a good choice because if n_i is large in comparison to n_{i+1} then $M(i, i + 1)$ tends to be closer to M_i than to M_{i+1}.

In any case if μ is an arbitrary totally finite measure on φ and if $E_\mu(M \mid \mathfrak{L}) = \theta' = (\theta_1', \theta_2', \cdots, \theta_k')$ then it follows from Theorem 2.3 in [4] that if θ is $\mathfrak{L}$-measurable then

$$\sum_{i=1}^k (\theta_i' - \theta_i)^2\mu(\{i\}) = \int (\theta' - \theta)^2 \, d\mu \leqq \int (M - \theta)^2 \, d\mu = \sum_{i=1}^k (M_i - \theta_i)^2\mu(\{i\}).$$

It follows that if $\sum_{i=1}^k (M_i - \theta_i)^2\mu(\{i\})$ converges to zero with probability one then the same can be said for $\sum_{i=1}^k (\theta_i' - \theta_i)^2\mu(\{i\})$.

Furthermore, it follows from Corollary 2.1 in [4] that

$$\sum_{i=1}^k (M_i - \hat{\theta}_i)^2\mu(\{i\}) \geqq \sum_{i=1}^k (M_i - \theta_i')^2\mu(\{i\}) + \sum_{i=1}^k (\theta_i' - \hat{\theta}_i)^2\mu(\{i\}) \geqq \sum_{i=1}^k (M_i - \theta_i')^2\mu(\{i\}).$$

Hence θ' is closer, in a sense, to the unrestricted estimates than $\hat{\theta}$. The problem is in the choice of μ. If, for example, we choose $\mu(\{i\}) = 1$, $i = 1, 2, \cdots, k$, then θ' has the advantage that it is easy to compute. On the other hand $\mu(\{i\}) = n_i$, $i = 1, 2, \cdots, k$, gives $\hat{\theta}$ in several cases and does not seem to be unreasonable in the bilateral exponential case. This choice then seems to be rather "robust."

6. Non-regular cases. In certain non-regular cases (i.e., cases where $f(x; \theta)$ has support which depends on θ) the procedure described in Section Two for constructing restricted maximum likelihood estimates may still be valid even though conditions (2.1) and (2.2) are not satisfied. (Van Eeden's work also covers such nonregular cases.) For example, suppose $\Theta = (-\infty, \infty)$ and

$$f(x; \theta) = e^{-(x-\theta)}, \quad x \geqq \theta, \qquad = 0, \quad x < \theta.$$

In this case the unrestricted estimates of $\theta_1, \theta_2, \cdots, \theta_k$ are given by $M_1, M_2, \cdots, M_k$ where $M_i = \min_{1 \leqq j \leqq n_i} x_{ij}$. Even though for fixed x, $f(x; \theta)$ is not a continuous function of θ it is not difficult to see that the procedure described in this paper yields restricted maximum likelihood estimates. A representation theorem such as Theorem 3.1 is also valid since the minimum of the pooled samples is between the minima of the individual samples (it is equal to one of them). Furthermore, by the Borel strong law of large numbers $M_i \to \theta_i$ with probability one so that our estimates are consistent.

7. Acknowledgment. The authors wish to thank Professor H. D. Brunk for suggesting the investigation in the course of which the subject matter of this paper arose, and Mr. Gary Makowski for working out some forms of Theorem 3.1.

REFERENCES

[1] Ayer, Miriam, Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, Edward (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647.

[2] Boswell, M.T. (1966). Estimating and testing trend in a stochastic process of the Poisson type. *Ann. Math. Statist.* **37** 1564–1573.

[3] Brunk, H. D. (1955). Maximum likelihood estimation of monotone parameters. *Ann. Math. Statist.* **26** 607–616.

[4] Brunk, H. D. (1965). Conditional expectation given a σ-lattice and applications. *Ann. Math. Statist.* **36** 1339–1350.

[5] Robertson, Tim (1965). A note on the reciprocal of the conditional expectation of a positive random variable. *Ann. Math. Statist.* **36** 1302–1305.

[6] Van Eeden, Constance (1957). Maximum likelihood estimation of partially or completely ordered parameters, I and II. *Indag. Math.* **19** 128–136, 201–211.