SEQUENTIAL MINIMAX ESTIMATION FOR THE RECTANGULAR DISTRIBUTION WITH UNKNOWN RANGE[1]

By J. Kiefer

Cornell University

1. Summary. This paper is concerned with sequential minimax estimation of the parameter $\theta (0 < \theta < \infty)$ of the density function (3.1) when the observations are independently and identically distributed with this density, each observation costs the same amount $c > 0$, and the weight function is as given in Section 2. A procedure requiring a fixed sample size is shown to be a minimax solution for this problem.

2. Introduction. An important problem in the theory of statistical decision functions[2] is that of minimax sequential estimation of the parameter of an (unknown) member of a given family of distribution functions when the observations are taken on chance variables which are independently and identically distributed and when the cost of taking n observations is cn (with $c > 0$) regardless of the way in which they are taken. This problem was solved for the case of point estimation of the mean of the rectangular distribution from $\theta - \frac{1}{2}$ to $\theta + \frac{1}{2}$ $(-\infty < \theta < \infty)$, for weight function $W(\theta, d) = (\theta - d)^2$ by Wald [1]; the minimax sequential estimation problem for the normal distribution was solved for a variety of terminal decision spaces and weight functions by Wolfowitz [2] (see also [3]); certain extensions and modifications of the results of both of these cases were given by Blyth [4].

The present paper is devoted to a problem of sequential minimax estimation for the case where the family of possible distribution functions consists of all distributions for which the successive observations are independently and identically distributed with rectangular density function from 0 to θ (equation (3.1)) for $\theta \varepsilon \Omega = \{\theta \mid 0 < \theta < \infty\}$ and where the cost of taking n observations is $cn (c > 0)$ regardless of the way in which the observations are taken. The object is to estimate θ, the terminal decision space being $D = \{d \mid 0 \leqq d < \infty\}$. The weight function is $W(\theta, d) = [(\theta - d)/\theta]^2$; i.e., the loss incurred by making decision d when θ is the true parameter is the square of the fractional error in estimating θ. Thus, the minimax problem considered in this paper is that of finding a sequential estimation procedure which minimizes $\sup_\theta\{cE_\theta(n) + E_\theta[(\theta - d)/\theta]^2\}$. A word is in order concerning our choice of weight function. The reason we do not study the problem for such weight functions as $| \theta - d |$, $(\theta - d)^2$, or $[(\theta - d)^2/\theta]$ is that for such weight functions the supremum of the risk over all $\theta \varepsilon \Omega$ is infinite for every decision function, so that

[1] Research under a contract with the Office of Naval Research.

[2] See Wald [1] for an exposition of this theory and an explanation of the nomenclature used herein.

every decision function is minimax. In addition, weight functions which depend only on d/θ (such as $[(\theta - d)/\theta]^2$) have a structure which essentially simplifies matters when estimating a scale parameter. On the other hand, it does not seem convenient in the present case to consider simultaneously a large class of weight functions as was possible in the cases of symmetrical densities studied in [2] and [4]. We therefore treat only one typical weight function here, noting that the same method should be applicable to many others.

With Ω, D, and $W(\theta, d)$ as described above, we shall prove that there is a minimax solution for which a fixed number of observations is taken. Specifically, the function $r(m)$ of (3.20) (which is the constant risk corresponding to taking a sample of fixed sample size m and then estimating θ by the expression of (2.1) with m for m_0) has at most two minima (if there are two, they are for successive values of m; moreover, there is only one minimum for all but a denumerable set of values of c). A minimax decision function is given by taking m_0 observations $y_1, y_2, \cdots y_{m_0}$, where $r(m_0)$ is the minimum of $r(m)$ (if there are two minima, at m_0 and $m_0 + 1$, one may randomize in any way between the decisions to take m_0 or $m_0 + 1$ observations); and by then estimating θ by

$$\frac{m_0 + 2}{m_0 + 1} \max (y_1, \cdots, y_{m_0}) \tag{2.1}$$

if $m_0 > 0$ (we replace m_0 by $m_0 + 1$ throughout (2.1) if the latter number of observations is taken when there are two minima), and by 0 if $m_0 = 0$. The risk corresponding to this decision function is then $r(m_0)$ for all values of $\theta \varepsilon \Omega$. It follows, incidentally, that this decision function is uniformly best among all cogredient procedures (see [4]). It is also a minimax solution for some related problems discussed in Section 3 of [4].

The method of proof is to calculate a lower bound on the Bayes risk when the a priori density on Ω is given by (3.4). It follows from (3.24) that as the parameter a of (3.4) approaches zero, the corresponding Bayes risk approaches $r(m_0)$; hence, by an argument like that of [1], p. 167, the procedure described in the previous paragraph is a minimax solution. The lower bound (3.24) is calculated in detail, since the necessary steps in its calculation differ somewhat from those of [1], [2], and [4]. We also note that, in this case of estimating a scale parameter, the tool used in [1], [2], and [4] of attempting to attain a "uniform a priori distribution on the real line" in the location parameter case is replaced by trying to attain the "a priori density" $1/\theta$. The proof is somewhat shortened by restricting the positive range of $\lambda_a(\theta)$ to values $\theta < 1$. This asymmetry manifests itself in the fact that the estimator of (3.7) does not tend to a minimax solution as $a \rightarrow 0$.

The fact that $\lambda_a(\theta)$ is positive only for $\theta < 1$ also shows that the fixed sample procedure described above is minimax for the problem of estimating θ when the above setup is altered by making $\Omega = \{\theta \mid 0 < \theta < b\}$, where $0 < b < \infty$: the argument of Section 3 shows this for $b = 1$, and the result for general $b = b'$ follows immediately from the case $b = 1$ if one considers there the problem of estimating $b'\theta$ from the sequence $\{b'Y_i\}$ of chance variables. Similarly, by considering for each value of a in Section 3 the problem of estimating $ba\theta$ from the sequence $\{baY_i\}$, one sees that our fixed sample procedure is also minimax for the problem of estimating θ when our original setup is altered by making $\Omega = \{\theta \mid b < \theta < \infty\}$. However, the given procedure is obviously not admissible if $m_0 > 0$ (or $m_0 \geqq 0$ in the second case): for example, a trivially better procedure in the first case when $m_0 > 0$ is to estimate θ by b whenever the expression of (2.1) is $> b$.

Finally, we remark that the problem of estimating θ for the case where the $f(y; \theta)$ of (3.1) is replaced by $1/(2\theta)$ for $-\theta < y < \theta$, is obviously identical to the one we consider: one has only to note that after n observations a sufficient statistic is still given by (3.2) if only Y_i is replaced by $| Y_i |$ for $i = 1, \cdots, n$. It is also of interest to note that our problem may be translated (by considering $T_i = e^{-Y_i}$, $\phi = e^{-\theta}$) into that of sequential minimax estimation of the parameter ϕ of the density $e^{-(t-\phi)}$ for $t > \phi$, 0 otherwise $(-\infty < \phi < \infty)$, when the weight function is $W(\phi, d) = (1 - e^{-(d-\phi)})^2$.

3. Calculations. For brevity, we shall throughout this section state the values of density functions and discrete probability functions only over the domains where they are positive. Let $Y_1, Y_2, \cdots$ be a sequence of independently and identically distributed chance variables, each with density function

$$f(y; \theta) = 1/\theta \qquad 0 < y < \theta, \tag{3.1}$$

where $\theta \in \Omega = \{\theta \mid 0 < \theta < \infty\}$. Define

$$X_n = \max \{Y_1, \cdots, Y_n\}. \tag{3.2}$$

Clearly, if observations $y_1, \cdots, y_n$ on $Y_1, \cdots, Y_n$ are taken, then X_n is a sufficient statistic for θ; i.e., for any a priori probability distribution on Ω, the a posteriori distribution of θ depends on $y_1, \cdots, y_n$ only through the value x_n taken on by X_n. Thus, in constructing sequential Bayes solutions, we may restrict ourselves to decision functions for which the (perhaps randomized) rule for stopping and estimation depends, after n observations, only on x_n. The density function of X_n is given by

$$g_n(x; \theta) = \frac{nx^{n-1}}{\theta^n}, \qquad 0 < x < \theta. \tag{3.3}$$

For $0 < a < 1$, we define

$$\lambda_a(\theta) = \frac{1}{\log (1/a)} \frac{1}{\theta}, \qquad a < \theta < 1. \tag{3.4}$$

If $\lambda_a(\theta)$ is the a priori density function on Ω and $y_1, \cdots, y_n$ have been observed, the a posteriori density of θ given that $X_n = x$ is easily computed to be

$$h_n(\theta \mid X_n = x) = \frac{nz^n}{1 - z^n} \cdot \frac{1}{\theta^{n+1}}, \qquad z < \theta < 1, \tag{3.5}$$

where $z = \max (a, x)$ and we note that $P\{z < 1\} = 1$.

The a posteriori loss (excluding cost of experimentation) if one stops after n observations and uses d to estimate θ, is

$$
\begin{aligned}
W_n^*(d, z) &= \int_0^1 \left(\frac{d-\theta}{\theta}\right)^2 h_n(\theta \mid X_n = x)\, d\theta \\
&= 1 + \frac{n}{z^2(1-z^n)}\left[z^{n+2}\left(\frac{2d}{n+1} - \frac{d^2}{n+2}\right) - \left(\frac{2dz}{n+1} - \frac{d^2}{n+2}\right)\right].
\end{aligned}
\tag{3.6}
$$

The unique minimum of W_n^* with respect to d is easily seen to occur for

$$
d = \frac{n+2}{n+1} \cdot \frac{1-z^{n+1}}{1-z^{n+2}} \cdot z, \tag{3.7}
$$

the corresponding value of W_n^* being

$$
W_n^{**}(z) = 1 - \frac{n(n+2)}{(n+1)^2} \cdot \frac{(1-z^{n+1})^2}{(1-z^n)(1-z^{n+2})}. \tag{3.8}
$$

For $n = 0$, the integral in (3.6) must be altered by replacing h_n by λ_a; the final expression must be changed accordingly. Equation (3.7) then holds with $z = a$, and (3.8) becomes $1 - 2(1-a)/[(1+a)\log(1/a)]$.

Next we note that when $f(y; \theta)$ is the density of each Y_i, the conditional distribution function of X_n given that $X_{n-1} = u$ assigns probability mass u/θ at the point $x = u$ and density $1/\theta$ for $u < x < \theta$. For $n = 1$, the distribution of X_1 is of course given by the density $f(x; \theta)$. We conclude that if $\lambda_a(\theta)$ is the a priori density on Ω, the distribution of X_1 is given by the density

$$
p_1(x) = \int_0^1 f(x; \theta)\lambda_a(\theta)\, d\theta = \begin{cases} \dfrac{1}{\log(1/a)} \cdot \dfrac{(1-a)}{a} & \text{if } x \leqq a, \\[2ex] \dfrac{1}{\log(1/a)} \cdot \dfrac{(1-x)}{x} & \text{if } a < x < 1; \end{cases} \tag{3.9}
$$

and that (using (3.5) with n replaced by $n-1$), for $n > 1$, the conditional distribution of X_n given that $\lambda_a(\theta)$ is the a priori density and $X_{n-1} = u$, is given, if $u \leqq a$, by

$$
\begin{aligned}
P_n\{X = u\} &= \frac{n-1}{n} \cdot \frac{1-a^n}{1-a^{n-1}} \cdot \frac{u}{a}, \\
p_n(x \mid u) &= \begin{cases} \dfrac{n-1}{n} \cdot \dfrac{1-a^n}{1-a^{n-1}} \cdot \dfrac{1}{a}, & u < x \leqq a, \\[2ex] \dfrac{n-1}{n} \cdot \dfrac{1-x^n}{1-a^{n-1}} \cdot \dfrac{a^{n-1}}{x^n}, & a < x < 1; \end{cases}
\end{aligned}
\tag{3.10}
$$

and, if $u > a$, by

$$
\begin{aligned}
P_n\{X = u\} &= \frac{n-1}{n} \cdot \frac{1-u^n}{1-u^{n-1}}, \\
p_n(x \mid u) &= \frac{n-1}{n} \cdot \frac{1-x^n}{1-u^{n-1}} \cdot \frac{u^{n-1}}{x^n}, \qquad u < x < 1;
\end{aligned}
\tag{3.11}
$$

where in each case P_n is the probability mass at $x = u$ and $p_n(x \mid u)$ is the density elsewhere.

Equations (3.10) and (3.11) yield for the conditional distribution of $Z_n = \max(X_n, a)$ given that $\lambda_a(\theta)$ is the a priori density and that $Z_{n-1} = v$, for all $n > 1$,

$$Q_n\{Z = v\} = \frac{n-1}{n} \cdot \frac{1 - v^n}{1 - v^{n-1}},$$

(3.12)

$$q_n(z \mid v) = \frac{n-1}{n} \cdot \frac{1 - z^n}{1 - v^{n-1}} \cdot \frac{v^{n-1}}{z^n}, \qquad v < z < 1,$$

where again q_n is a density and Q_n is the probability mass at $z = v$.

Let $\overline{W}_{n-1}(v)$ be the conditional expected value of $W_n^{**}(Z_n)$ given that $\lambda_a(\theta)$ is the a priori density and that $Z_{n-1} = v$ (where we define $Z_0 = a$). Using (3.8) and (3.9), we have

$$\begin{aligned}
\overline{W}_0(a) &= E\{W_1^{**}(Z_1)\} \\
&= W_1^{**}(a) \int_0^a p_1(z)\, dz + \int_a^1 W_1^{**}(z) p_1(z)\, dz \\
&= 1 - \frac{3}{4 \log (1/a)} \left\{ \frac{(1 - a^2)^2}{(1 - a^3)} + \int_a^1 \frac{(1 - z^2)^2}{z(1 - z^3)}\, dz \right\} \\
&< 1 - \frac{3}{4 \log (1/a)} \int_a^1 \left[\frac{1}{z} - 1 + \frac{(1 - z)^2}{(1 - z^3)} \right] dz \\
&< 1 - \frac{3}{4 \log (1/a)} \left[\log \frac{1}{a} - (1 - a) \right] < \tfrac{1}{4} + \frac{1}{\log (1/a)}.
\end{aligned} \tag{3.13a}$$

For $n > 1$, we have from (3.8) and (3.12),

$$\begin{aligned}
\overline{W}_{n-1}(v) &= E\{W_n^{**}(Z_n) \mid v\} \\
&= W_n^{**}(v) Q_n(Z = v) + \int_v^1 W_n^{**}(z) q_n(z \mid v)\, dz \\
&= 1 - \frac{(n-1)(n+2)}{(n+1)^2(1 - v^{n-1})} \\
&\quad \cdot \left\{ \frac{(1 - v^{n+1})^2}{(1 - v^{n+2})} + v^{n-1} \int_v^1 \frac{(1 - z^{n+1})^2}{z^n(1 - z^{n+2})}\, dz \right\}.
\end{aligned} \tag{3.13b}$$

The term in the last set of braces in (3.13b) may be written as

$$\begin{aligned}
1 - v^{n-1} &+ \frac{v^{n-1}(1 - v)[1 + v - v^2 - v^{n+2}]}{(1 - v^{n+2})} \\
&+ v^{n-1} \int_v^1 \left[\frac{1}{z^n} - \frac{z(2 - z - z^{n+1})}{(1 - z^{n+2})} \right] dz \\
&> (1 - v^{n-1}) + \frac{1}{n-1}(1 - v^{n-1}) - v^{n-1} \int_v^1 2z\, dz \\
&= \frac{n}{n-1}(1 - v^{n-1}) - v^{n-1}(1 - v^2).
\end{aligned} \tag{3.14}$$

We conclude that, whenever $n > 1$,

$$\overline{W}_{n-1}(v) < 1 - \frac{n(n+2)}{(n+1)^2} + \frac{(n-1)(n+2)}{(n+1)^2} \cdot \frac{v^{n-1}(1-v^2)}{(1-v)^{n-1}} < \frac{1}{(n+1)^2} + \frac{1}{\log(1/v)}, \tag{3.15}$$

where in the last step we have used the fact that $(n-1)(n+2)(n+1)^{-2} < \frac{1}{2}$ if $n = 2$ and < 1 otherwise, that $(1-v^2)(1-v^{n-1})^{-1} < 2$ if $n = 2$ and $\leqq 1$ otherwise, and that if $n > 1$ we have $v^{n-1} \leqq v < (\log 1/v)^{-1}$. From (3.13a) and (3.15), we have for all $n > 0$,

$$\overline{W}_{n-1}(v) < \frac{1}{(n+1)^2} + \frac{1}{\log(1/v)}. \tag{3.16}$$

Similarly, we have from (3.8) for $n > 1$,

$$\begin{aligned} W^{**}_{n-1}(v) &= 1 - \frac{(n-1)(n+1)}{n^2} \cdot \frac{(1-v^n)^2}{(1-v^{n-1})(1-v^{n+1})} \\ &= 1 - \frac{(n-1)(n+1)}{n^2}\left[1 + \frac{v^{n-1}(1-v)^2}{(1-v^{n-1})(1-N^{n+1})}\right] \\ &> \frac{1}{n^2} - v^{n-1} \geqq \frac{1}{n^2} - v > \frac{1}{n^2} - \frac{1}{\log(1/v)}; \end{aligned} \tag{3.17}$$

and, for $n = 1$ (putting $v = a$),

$$W^{**}_0(v) = 1 - \frac{2(1-a)}{(1+a)\log(1/a)} > 1 - \frac{2}{\log(1/v)}. \tag{3.18}$$

Combining (3.16), (3.17), and (3.18), we have for all $m \geqq 0$,

$$W^{**}_m(v) - \overline{W}_m(v) > \frac{2m+3}{(m+1)^2(m+2)^2} - \frac{3}{\log(1/v)}. \tag{3.19}$$

We now define, for all integers $m \geqq 0$,

$$r(m) = cm + \frac{1}{(m+1)^2}. \tag{3.20}$$

We note that $r(m+1) - r(m) = c - (2m+3)/((m+1)^2(m+2)^2)$. The function $r(m)$ evidently has at most two minima (if there are two, they are for consecutive values of m). Denote by m_0 the first integer for which $r(m_0)$ is a minimum. Let ϵ $(0 < \epsilon < 1)$ be such that $3\epsilon < r(m_0 - 1) - r(m_0)$ (if $m_0 = 0$, the last restriction is omitted). Let $d = e^{-1/\epsilon}$ and $a = e^{-1/\epsilon^2}$. Let m_1 be the smallest integer not less than $1/c$.

For any integer $K > 0$, if $\lambda_a(\theta)$ is the a priori density we have (noting that $d > a$)

$$(3.21)\qquad P\{X_K \geqq d\} = \int_d^1 \int_x^1 g_K(x;\theta)\lambda_a(\theta)\,d\theta\,dx = \int_d^1 \frac{1-x^K}{x\log(1/a)}\,dx$$
$$= \frac{\log(1/d)}{\log(1/a)} - \frac{1-d^K}{K\log(1/a)} < \epsilon.$$

We note that, after m observations ($m = 0, 1, \cdots$, ad inf. and putting $v = a$ if $m = 0$), any Bayes solution will certainly prescribe taking another observation if $W_m^{**}(v) - \overline{W}_m(v) - c > 0$, since this quantity is the a posteriori expected saving over stopping after m observations if instead one takes one additional observation and then stops and makes the best terminal decision.

We also note that, since $(\log 1/a)^{-1} = \epsilon^2 < \epsilon$, it follows from (3.21) that, when $\lambda_a(\theta)$ is the a priori density,

$$P\left\{\frac{1}{\log(1/Z_i)} < \epsilon \text{ for } i = 1, 2, \cdots, m_0 + m_1\right\} = P\left\{\frac{1}{\log(1/Z_{m_0+m_1})} < \epsilon\right\}$$
$$= P\left\{\frac{1}{\log(1/X_{m_0+m_1})} < \epsilon\right\} = P\{X_{m_0+m_1} < d\} > 1 - \epsilon.$$

Since $r(m-1) - r(m)$ is a decreasing function of $m(m > 0)$ and since $3\epsilon < r(m_0 - 1) - r(m_0)$, we conclude that, if $m_0 > 0$, the event

$$(3.23)\qquad \frac{1}{\log(1/Z_{m_0+m_1})} < \epsilon$$

entails the event $(\log(1/Z_{m_0-1}))^{-1} < \epsilon$, which entails $3(\log(1/Z_{m_0-1}))^{-1} < r(m_0 - 1) - r(m_0)$; or, equivalently, $-3(\log(1/Z_i))^{-1} + r(i) - r(i+1) > 0$ for $i = 0, 1, \cdots, m_0 - 1$. Finally, it follows from (3.19) that this entails the event $W_i^{**}(v) - \overline{W}_i(v) - c > 0$ for $i = 0, 1, \cdots, m_0 - 1$; and, for any Bayes solution relative to $\lambda_a(\theta)$, this entails the event that at least m_0 observations will be taken. Furthermore, the last statement is always true for $m_0 = 0$.

Similarly, we note from (3.17) and (3.18) that the event (3.23) certainly entails the event $W_i^*(v) > (1/(1+i)^2) - 2\epsilon$ for $i = m_0, m_0 + 1, \cdots, m_0 + m_1$. That is, if a terminal decision is made after exactly i observations ($i = m_0, \cdots, m_0 + m_1$), the total a posteriori loss plus cost of experimentation will be $> ci + (1/(1+i)^2) - 2\epsilon \geqq cm_0 + (1/(1+m_0)^2) - 2\epsilon$. Moreover, it follows from the definition of m_1 that this last expression is less than the cost of experimentation alone if more than $m_0 + m_1$ observations are taken.

To summarize, then, the event (3.23) implies for any Bayes solution relative to $\lambda_a(\theta)$ that the experiment will terminate with a total a posteriori loss plus cost of experimentation exceeding $cm_0 + (1/(1+m_0)^2) - 2\epsilon$. But it follows from (3.22) that (3.23) occurs with probability $>1 - \epsilon$. Since $m_0c + (1/(m_0+1)^2) \leqq 1$, it follows that the Bayes risk relative to $\lambda_a(\theta)$ exceeds

$$(3.24)\qquad (1-\epsilon)\left(m_0c + \frac{1}{(m_0+1)^2} - 2\epsilon\right) > m_0c + \frac{1}{(m_0+1)^2} - 3\epsilon.$$

Since ϵ may be taken to be arbitrarily small in magnitude, we conclude (see Section 2) that the fixed sample procedure described in Section 2 is indeed minimax.

REFERENCES

[1] ABRAHAM WALD, *Statistical Decision Functions*, John Wiley and Sons, 1950.

[2] J. WOLFOWITZ, "Minimax estimates of the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 218–230.

[3] C. STEIN AND A. WALD, "Sequential confidence intervals for the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 427–433.

[4] C. R. BLYTH, "On minimax statistical procedures and their admissibility," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 22–42.