THE ESTIMATION OF THE LOCATION OF A DISCONTINUITY IN DENSITY

HERMAN CHERNOFF
STANFORD UNIVERSITY

AND

HERMAN RUBIN
STANFORD UNIVERSITY AND UNIVERSITY OF OREGON

1. Introduction

It is well known that the maximum likelihood estimates of the parameters of a rectangular distribution are extremely efficient. In fact, these estimates differ from the parameter by an amount which is of the order of magnitude of $1/n$ where n is the sample size. In this paper a similar result is obtained for the maximum likelihood estimate of α where α is one of a set of parameters determining the family of distributions and α is the location of a point of discontinuity of the density of the random variable observed. The estimate $\hat{\alpha}$ is approximately one of the observations in the neighborhood of α.

The limiting distribution of $n(\hat{\alpha} - \alpha)$ is related to a random walk problem. One part of this random walk problem involves the distribution of $\sum_{i=1}^{R} z_i$ where R is the value of r which minimizes $\sum_{i=1}^{r} (z_i - \omega)$, the z_i are independent and have the exponential distribution with mean 1 and $0 < \omega < 1$. The limiting distribution depends only on β and γ where these are the one-sided limits of the density at α.

This type of problem arises naturally whenever one considers a population which has two subpopulations and one of these is truncated. For example, a teacher may have students who can afford to pay tuition or those who cannot pay but have obtained a scholarship on the basis of a mark in a competitive examination. If he does not know the passing mark or which students have scholarships, the problem of estimating the passing mark is of the above type.

Consider the family of distributions given by the density

$$(1)\qquad \begin{aligned} f(x, \alpha, \beta, \gamma) &= \beta, & 0 \leqq x \leqq \alpha, \\ f(x, \alpha, \beta, \gamma) &= \gamma, & \alpha < 1 \leqq 1 \end{aligned}$$

where

$$(2)\qquad \beta\alpha + \gamma(1 - \alpha) = 1,$$

and $0 < \alpha < 1$.

This paper was written with the partial support of the Office of Naval Research, and the Office of Ordnance Research, U.S. Army, under Contract DA-04-200-ORD-355.

The likelihood based on n independent observations is

$$[\beta^{F_n(\alpha)}\gamma^{1-F_n(\alpha)}]^n \tag{3}$$

where $F_n(x)$ is the sample cumulative distribution function. Maximizing with respect to β and γ, we find that the maximum likelihood estimate $\hat{\alpha}$ of α maximizes

$$F_n(\alpha)\log\frac{F_n(\alpha)}{\alpha}+[1-F_n(\alpha)]\log\frac{1-F_n(\alpha)}{1-\alpha}. \tag{4}$$

It is easy to see that the above expression attains its maximum at $\alpha = 0$, $\alpha = 1$ or at one of the n observations. In this section we shall proceed heuristically, replacing the above expression by a first-order expansion in α and $F_n(\alpha)$ about α_0 and $F_n(\alpha_0) \approx \beta_0\alpha_0$ where α_0, β_0 and γ_0 are the true values of the parameters. We obtain

$$\begin{aligned} F_n(\alpha_0)\log\frac{F_n(\alpha_0)}{\alpha_0}&+[1-F_n(\alpha_0)]\log\frac{1-F_n(\alpha_0)}{1-\alpha_0} \\ &+(\alpha-\alpha_0)(\gamma_0-\beta_0)+[F_n(\alpha)-F_n(\alpha_0)]\log\frac{\beta_0}{\gamma_0}. \end{aligned} \tag{5}$$

Now let us assume that $\beta_0 > \gamma_0 > 0$. One might then expect that the maximum likelihood estimate would maximize

$$M(\alpha) = F_n(\alpha) - F_n(\alpha_0) - c_0(\alpha-\alpha_0) \tag{6}$$

where

$$c_0 = \frac{\beta_0-\gamma_0}{\log\beta_0-\log\gamma_0} \tag{7}$$

and it is easy to see that[1]

$$\beta_0 > c_0 > \gamma_0 . \tag{8}$$

This expression decreases between observations but has a positive jump at each observation. Let us order the observations and label them

$$\cdots,\ x_{-2}^{(n)},\ x_{-1}^{(n)},\ x_1^{(n)},\ x_2^{(n)},\cdots \tag{9}$$

so that[2]

$$x_{-2}^{(n)} < x_{-1}^{(n)} < x_0^{(n)} = \alpha_0 < x_1^{(n)} < x_2^{(n)} < \cdots . \tag{10}$$

Then

$$\begin{aligned} M(x_r^{(n)}) &= \frac{r}{n} - c_0(x_r^{(n)}-x_0^{(n)}), && r>0, \\ M(x_r^{(n)}) &= \frac{r+1}{n} - c_0(x_r^{(n)}-x_0^{(n)}), && r<0. \end{aligned} \tag{11}$$

We shall study the behavior of $\hat{\alpha}$ by studying that of $x_R^{(n)} - x_0^{(n)}$ where R is the value of r which maximizes $M(x_r^{(n)})$. It is convenient to introduce

$$y_r^{(n)} = n(x_r^{(n)} - x_{r-1}^{(n)}) \tag{12}$$

[1] It is of interest to note that an estimate proposed by Karlin has a similar behavior except that c_0 is replaced by one. In fact Karlin's estimate is the simple one of finding that α which maximizes $|F_n(\alpha) - \alpha|$.

[2] We may use strict inequalities since the observation will differ with probability 1. Furthermore, no observation will correspond to $x_0 = \alpha_0$.

since the $y_r^{(n)}$ have asymptotically independent exponential distributions. In fact, if we define $E(\omega)$ as the exponential distribution with density $e^{-y/\omega}/\omega$ for $y > 0$

$$(13)\quad \begin{aligned} &\mathcal{L}(y_r^{(n)}) \to \mathcal{L}(y_r) = E\left(\frac{1}{\gamma_0}\right) \qquad \text{for } r > 0 \\ &\mathcal{L}(y_r^{(n)}) \to \mathcal{L}(y_r) = E\left(\frac{1}{\beta_0}\right) \qquad \text{for } r \leqq 0 . \end{aligned}$$

Then, asymptotically, the distribution of $n(\hat{a} - a_0)$ behaves like that of A_R where R minimizes B_r and

$$(14)\quad \begin{aligned} &A_r = y_1 + y_2 + \cdots + y_r \qquad &&\text{for } r > 0 \\ &A_r = -\ [y_{r+1} + y_{r+2} + \cdots + y_0] \qquad &&\text{for } r < 0 . \end{aligned}$$

$$(15)\quad \begin{aligned} &B_r = \left(y_1 - \frac{1}{c_0}\right) + \left(y_2 - \frac{1}{c_0}\right) + \cdots + \left(y_r - \frac{1}{c_0}\right) \qquad &&\text{for } r > 0 \\ &B_r = -\left[y_0 + \left(y_{-1} - \frac{1}{c_0}\right) + \left(y_{-2} - \frac{1}{c_0}\right) + \cdots + \left(y_{(r+1)} - \frac{1}{c_0}\right)\right] \qquad &&\text{for } r < 0 . \end{aligned}$$

Detailed proofs will be presented for the above assertions which reduce the asymptotic distribution of $n(\hat{a} - a_0)$ to the solution of the above random walk problem. The random walk problem will also be studied.

Now suppose that our problems were modified so that the densities in the neighborhood of a_0 were allowed to vary but had a right-hand limit $\gamma_0 > 0$ and a left-hand limit $\beta_0 > \gamma_0 > 0$. If there were available a consistent estimate of a_0, one should be able to obtain an interval whose length approaches zero but which contains a_0 and many observations with large probability. The density in this interval is very similar to that in the original problem. We may transform our small interval to (0, 1), forget all observations not in this interval and apply the estimate $\hat{a}$ which is maximum likelihood for the original problem. It is reasonable to expect this "quasi-maximum likelihood estimate" to have the asymptotic distribution discussed above.

One may now ask whether the "true" maximum likelihood estimate would have a better distribution. We shall show that under certain regularity conditions, this is not so. In fact, the two estimates differ by a quantity which is small compared to $1/n$.

In the general problem the case where $\gamma = 0$ introduces technical difficulties and has not been treated.

2. Consistency for the special problem

We shall later treat the generalized version of the problem introduced above under certain regularity conditions. One of these conditions will be the consistency of the maximum likelihood estimate. In this section we shall prove that the maximum likelihood estimate is consistent for the special problem of the introduction. In doing so we shall make use of the O_p and o_p notation of Mann and Wald [1] and a paraphrase due to Pratt of a section of one of their theorems [2]. Roughly speaking, this theorem states that the calculus of O_p and o_p is the same as that of O and o. In later sections we shall make use of a slight extension of another of their theorems which states that $\mathcal{L}[g(y_n)] \to \mathcal{L}[g(y)]$ if $\mathcal{L}(y^{(n)}) \to \mathcal{L}(y)$ and the set of discontinuities of g is of probability measure zero with respect to the limiting distribution $\mathcal{L}(y)$.

Since $\hat{a}$ maximizes (2) we may treat $\hat{a}$ as a function of the sample c.d.f. F_n.

$$\hat{a} = \Phi(F_n) . \tag{16}$$

Let F_0 denote the "true" c.d.f. That is,

$$\begin{aligned} F_0(x) &= \beta_0 x && \text{for } 0 \leqq x \leqq a_0 \\ F_0(x) &= 1 - \gamma_0 (1 - x) && \text{for } a_0 \leqq x \leqq 1 . \end{aligned} \tag{17}$$

The function

$$H_0(x) = F_0(x) \log \frac{F_0(x)}{x} + [1 - F_0(x)] \log \frac{1 - F_0(x)}{1 - x}, \qquad 0 \leqq x \leqq 1 , \tag{18}$$

reaches a positive peak at $x = a_0$. Both right-hand and left-hand derivatives exist at $x = a_0$, but they are not in general equal to zero. It can be shown that for a sequence of nonrandom c.d.f.'s $G_n(x)$ such that

$$\sup_{0<x<1} \left| \frac{G_n(x)}{x} \right| = O(1) , \tag{19}$$

$$\sup_{0<x<1} \left| \frac{1 - G_n(x)}{1 - x} \right| = O(1) , \tag{20}$$

and

$$\sup_{0<x<1} |G_n(x) - F_0(x)| = o(1) , \tag{21}$$

it follows that

$$\Phi(G_n) = a_0 + o(1) .$$

Furthermore, we shall show that for the sample c.d.f.'s (which are random)

$$\sup_{0<x<1} \left| \frac{F_n(x)}{x} \right| = O_p(1) , \tag{22}$$

$$\sup_{0<x<1} \left| \frac{1 - F_n(x)}{1 - x} \right| = O_p(1) , \tag{23}$$

and

$$\sup_{0<x<1} |F_n(x) - F_0(x)| = o_p(1) . \tag{24}$$

Hence, it follows by the above-mentioned result of Mann and Wald that

$$\hat{a} = \Phi(F_n) = a_0 + o_p(1) , \tag{25}$$

that is, $\hat{a}$ is a consistent estimate.

The left-hand side of (24) is the Kolmogorov-Smirnov statistic. It is well known that this statistic is $O_p(1/\sqrt{n})$ see [3]. To prove (22) and (23), it suffices to establish (22) for the uniform distribution. In fact we shall find it slightly more convenient to show that $\sup_{0<x<1} |F_n(x)/x - 1| = O_p(1)$ for the uniform distribution. Select an $\epsilon > 0$. Since the reciprocal of the smallest observation is $O_p(n)$, there is an a such that

$$P\left\{F_n\left(\frac{a}{n}\right) = 0\right\} = P\left\{\frac{F_n(x)}{x} - 1 = -1 \quad \text{for } x \leqq \frac{a}{n}\right\} > 1 - \frac{\epsilon}{2} . \tag{26}$$

By Chebyshev's inequality

$$P\left\{\left|\frac{F_n(x)}{x}-1\right|>M\right\}\leq\frac{1}{M^2nx},$$

$$\text{(27)}\qquad P\left\{\sup_{i\geq 0}\left|\frac{F_n(2^ia/n)}{2^ia/n}-1\right|>M\right\}\leq\frac{1}{M^2a}\sum_{i=0}^{\infty}\frac{1}{2^i}=\frac{2}{M^2a}<\frac{\epsilon}{2}$$

for M large enough. But if

$$\text{(28)}\qquad \left|\frac{F_n(x)}{x}-1\right|<M \quad\text{and}\quad \left|\frac{F_n(2x)}{2x}-1\right|<M,$$

then for $x<y<2x$,

$$\text{(29)}\qquad \left|\frac{F_n(y)}{y}-1\right|<2M+1.$$

Hence,

$$\text{(30)}\qquad P\left\{\sup_{0\leq x\leq 1}\left|\frac{F_n(x)}{x}-1\right|>(2M+1)\right\}<\epsilon,$$

which is the desired result.

3. The general problem

In section 1, we discussed the special problem and heuristically indicated that we expected our maximum likelihood estimate to be close to the value of a which maximizes $M(a)$. In this section we shall obtain a similar result for the general problem.

Several regularity conditions will be used. A brief discussion of them is presented in the appendix.

Let a family of distributions be specified by the parameters θ and a where θ may be multidimensional and the density $f(x,\theta,a)$ satisfies

$$\text{(31)}\qquad \lim_{x\to a-} f(x,\theta,a)=\beta(\theta,a)$$

$$\text{(32)}\qquad \lim_{x\to a+} f(x,\theta,a)=\gamma(\theta,a).$$

We shall assume that the regularity conditions of the appendix hold. We also assume that if (θ_0, a_0) is the "true value" of the parameter, then the maximum likelihood estimate $(\hat{\theta}, \hat{a})$ based on n independent observations converges in probability to (θ_0, a_0). Let $X=(X_1, X_2,\cdots, X_n)$ represent n independent observations with density given by $f(x,\theta_0,a_0)$. The logarithm of the likelihood function is given by

$$\text{(33)}\qquad L(X,\theta,a)=\frac{1}{n}\sum_{i=1}^{n}\log f(X_i,\theta,a_0)+\frac{1}{n}\sum_{i=1}^{n}[\log f(X_i,\theta,a) - \log f(X_i,\theta,a_0)].$$

Let

$$\text{(34)}\qquad S=S_1+S_2$$

where

$$\text{(35)}\qquad S_1=\frac{1}{n}\sum_{X_i\in B(a,a_0)}[\log f(X_i,\theta,a)-\log f(X_i,\theta,a_0)]$$

$$\text{(36)}\qquad S_2=\frac{1}{n}\sum_{X_i\notin B(a,a_0)}[\log f(X_i,\theta,a)-\log f(X_i,\theta,a_0)]$$

and $B(a, a_0)$ is the set of points between a and a_0. We now apply condition R_1 to S_1 (see appendix)

$$(37) \qquad S_1 = [\log \beta(\theta, a_0) - \log \gamma(\theta, a_0) + o(1)][F_n(a) - F_n(a_0)].$$

[Hereafter and in equation (37), the o and O will be understood to hold uniformly in θ in some interval about θ_0. In almost all cases o and O refer to $a \to a_0$ while o_p and O_p refer to $n \to \infty$. We shall point out any nontrivial situations where this is not the case.] Condition R_2 gives us

$$(38) \quad S_2 = \frac{1}{n} \sum_{X_i \notin B(a, a_0)} (a - a_0) \frac{\partial \log f(X_i, \theta, a_0)}{\partial a} + O_p(1)\, o(1)\, O(|a - a_0|)$$

$$= \frac{1}{n} \sum_{i=1}^{n} (a - a_0) \frac{\partial \log f(X_i, \theta, a_0)}{\partial a} + o(|a - a_0|) O_p(1) + O(|a - a_0|)[F_n(a) - F_n(a_0)].$$

Let

$$(39) \qquad \delta(\theta) = E\left\{\frac{\partial \log f(X, \theta, a_0)}{\partial a}\right\}$$

where the above expectation is relative to the "true" distribution. Applying R_3

$$(40) \quad S_2 = \delta(\theta)(a - a_0) + (a - a_0)\, o_p(1) + o(|a - a_0|) O_p(1) + O(|a - a_0|)[F_n(a) - F_n(a_0)]$$

$$(41) \quad S(\theta, a) = S_1 + S_2 = [\log \beta(\theta, a_0) - \log \gamma(\theta, a_0) + o(1)] M(\theta, a) + (a - a_0)[o_p(1) + o(1)] + o(|a - a_0|) O_p(1)$$

where

$$(42) \qquad M(\theta, a) = [F_n(a) - F_n(a_0)] - c(\theta)(a - a_0),$$

and

$$(43) \qquad c(\theta) = \frac{-\delta(\theta)}{\log \beta(\theta, a_0) - \log \gamma(\theta, a_0)}.$$

We remark that

$$(44) \qquad c(\theta_0) = \frac{\beta(\theta_0, a_0) - \gamma(\theta_0, a_0)}{\log \beta(\theta_0, a_0) - \log \gamma(\theta_0, a_0)},$$

and

$$(45) \qquad \beta(\theta_0, a_0) > c(\theta_0) > \gamma(\theta_0, a_0) > 0.$$

Since $\hat{a}$ is consistent, there is a sequence $\{a_n\}$ of positive real numbers which converges to zero such that $\hat{a} - a_0 = o_p(a_n)$. Let $\tilde{a}$ be that value of a in the closed interval $[a_0 - a_n, a_0 + a_n]$ which maximizes $M(\theta_0, a)$. This function is the analogue of $M(a)$ in the introduction and is maximized at an observation or at an end point of the interval. This section will be concluded when we prove that $\hat{a} = \tilde{a} + o_p(1/n)$ so that our study of $\hat{a}$ is essentially reduced to that of $\tilde{a}$.

LEMMA 1.

$$(46) \qquad \tilde{a} - a_0 = O_p\left(\frac{1}{n}\right).$$

PROOF. Since $M(\theta_0, a_0) = 0$, we need only show that $M(\theta_0, a)$ is negative for $n(a - a_0)$

large enough. More specifically, we shall show that for each $\epsilon > 0$ there is a K and $\eta > 0$ such that

$$(47) \qquad P\left\{\sup_{K/n<|a-a_0|<a_n} \frac{M(\theta_0, a)}{a-a_0} < -\eta\right\} > 1-\epsilon.$$

For $a > a_0$,

$$(48) \qquad F(a) - F(a_0) = [\gamma(\theta_0, a_0) + o(1)](a-a_0).$$

If we can show that with large probability $[F_n(a) - F_n(a_0)]/[F(a) - F(a_0)]$ is arbitrarily close to one for $n(a - a_0)$ large enough, we would then have $M(\theta_0, a)/|a - a_0|$ is arbitrarily close to $\gamma(\theta_0, a_0) - c(\theta_0) + o(1)$. Since $\gamma(\theta_0, a_0) - c(\theta_0) < 0$, there would be an $\eta > 0$ such that

$$(49) \qquad \frac{M(\theta_0, a)}{|a-a_0|} < -\eta,$$

for $n(a - a_0)$ large enough but also less than na_n. A similar argument applies for $a < a_0$ since $\beta(\theta_0, a_0) - c(\theta_0) > 0$. Hence it suffices to prove

LEMMA 2. *For the uniform distribution, for each* $\epsilon > 0$ *and* $\eta_1 > 0$, *there is a* K *such that*

$$(50) \qquad P\left\{\sup_{x>K/n}\left|\frac{F_n(x)}{x}-1\right| < \eta_1\right\} > 1-\epsilon.$$

PROOF. The proof of this lemma is similar to that of (22). For $b > 1$,

$$(51) \quad P\left\{\sup_{0\le i\le\infty}\left|\frac{F_n(b^iK/n)}{(b^iK/n)}-1\right| > \eta_2\right\} \le \frac{1}{\eta_2^2K}\sum_{i=0}^{\infty} b^{-i} = \frac{1}{\eta_2^2K(1-1/b)}.$$

If

$$(52) \qquad \left|\frac{F_n(x)}{x}-1\right| < \eta_2 \quad \text{and} \quad \left|\frac{F_n(bx)}{bx}-1\right| < \eta_2$$

then for $x \le y \le bx$

$$(53) \quad -\frac{1}{b}\eta_2 + \left(\frac{1}{b}-1\right) < \frac{F_n(x)}{bx} - 1 < \frac{F_n(y)}{y} - 1 < \frac{F_n(bx)}{x} - 1 < b\eta_2 + (b-1).$$

We may select η_2 and $b > 1$ such that $b\eta_2 + (b - 1) < \eta_1$ and $-\eta_2/b + (1/b - 1) > -\eta_1$. Then select K such that $\eta_2^2K(1 - 1/b) > 1/\epsilon$ and the lemma follows.

If $(\hat{\theta}, \hat{a})$ is the maximum likelihood estimate of (θ, a) we have upon substituting in (41) and using $M(\hat{\theta}, \tilde{a}) = O_p(1/n)$

$$(54) \quad 0 \le S(\hat{\theta}, \hat{a}) - S(\hat{\theta}, \tilde{a})$$
$$\le [\log\beta(\hat{\theta}, a_0) - \log\gamma(\hat{\theta}, a_0) + o_p(1)][M(\hat{\theta}, \hat{a}) - M(\hat{\theta}, \hat{a})]$$
$$+ (\hat{a} - \tilde{a})\, o_p(1) + o_p\left(\frac{1}{n}\right).$$

If we can show that for $a \ne \hat{a}$, $M(\hat{\theta}, a) - M(\hat{\theta}, \tilde{a})$ must be negative and not of smaller order of magnitude than $|a - \tilde{a}|$, it would follow that $\hat{a} - \tilde{a} = o_p(1/n)$. To do this we first treat $M(\theta_0, a) - M(\theta_0, \tilde{a})$.

LEMMA 3.

$$(55) \qquad Y_n = \sup_{\substack{a\ne\tilde{a}\\ |a-a_0|\le a_n}} \left\{\frac{|\tilde{a}-a|}{M(\theta_0, \tilde{a}) - M(\theta_0, a)}\right\} = O_p(1).$$

We divide the interval $|a - a_0| \leqq a_n$ into $|a - a_0| \leqq K/n$ and $K/n \leqq |a - a_0| \leqq a_n$ where K is selected so that (47) holds. Then, since $M(\theta_0, \tilde{a}) > 0$, the supremum over the second set is less than $1/\eta$ with probability greater than $1 - \epsilon$. We now treat the supremum over $|a - a_0| \leqq K/n$ while $\tilde{a}$ coincides with one of the observations in this interval. There is an R such that with probability greater than $1 - \epsilon$, $x_{-R}^{(n)} < a_0 - K/n$ and $a_0 + K/n < x_R^{(n)}$. Then the supremum over $|a - a_0| \leqq K/n$ is dominated by

$$Z_n = \sup_{-R \leqq j_1 < j_2 \leqq R} \left| \frac{x_{j_2}^{(n)} - x_{j_1}^{(n)}}{M(\theta_0, x_{j_2}^{(n)}) - M(\theta_0, x_{j_1}^{(n)})} \right| \tag{56}$$

$$= \sup_{-R \leqq j_1 < j_2 \leqq R} \left| \frac{\sum_{i=j_1+1}^{j_2} y_i^{(n)}}{(j_2^* - j_1^*) + c(\theta_0) \sum_{i=j_1+1}^{j_2} y_i^{(n)}} \right|$$

where $y_i^{(n)} = n(x_i^{(n)} - x_{i-1}^{(n)})$, $j^* = j$ if $j > 0$ and $j^* = j + 1$ if $j \leqq 0$. It suffices to prove that $Z_n = O_p(1)$. Now Z_n can be expressed as a function Ψ of $y_i^{(n)}$, $-R \leqq i \leqq R$, where Ψ is continuous except on a set of measure zero. The $y_i^{(n)}$ have a joint continuous limiting distribution. This distribution is that of the independent exponentially distributed variables y_i, $-R \leqq i \leqq R$. With respect to the limiting distribution the set of discontinuities have probability zero and $Z = \Psi(y_i, -R \leqq i \leqq R)$ is a random variable. A slight extension of a theorem due to Mann and Wald tells us that $\mathcal{L}(Z_n) \to \mathcal{L}(Z)$ and hence $Z_n = O_p(1)$.

We shall now extend lemma 3 to obtain the same property uniformly in θ for some interval about θ_0.

LEMMA 4. *There is a* $\delta > 0$ *such that*

$$Y_n^* = \sup_{\substack{a \neq \tilde{a} \\ |a - a_0| \leqq a_n \\ |\theta - \theta_0| \leqq \delta}} \left\{ \frac{|\tilde{a} - a|}{M(\theta, \tilde{a}) - M(\theta, a)} \right\} = O_p(1) \tag{57}$$

and

$$P\{M(\theta, \tilde{a}) - M(\theta, a) > 0 \text{ for } a \neq \tilde{a}, |a - a_0| \leqq a_n, |(\theta - \theta_0)| \leqq \delta\} \to 1 \tag{58}$$

as $n \to \infty$.

PROOF.

$$M(\theta, \tilde{a}) - M(\theta, a) = [M(\theta_0, \tilde{a}) - M(\theta_0, a)] + [c(\theta_0) - c(\theta)](\tilde{a} - a). \tag{59}$$

For $\epsilon > 0$, there is a $\delta > 0$ such that

$$P\{2Y_n |c(\theta_0) - c(\theta)| \leqq 1\} > 1 - \epsilon \quad \text{for} \quad |\theta_0 - \theta| \leqq \delta. \tag{60}$$

But then

$$Y_n^* \leqq 2Y_n \tag{61}$$

and

$$M(\theta, \tilde{a}) - M(\theta, a) > \frac{|\tilde{a} - a|}{2Y_n} \tag{62}$$

and lemma 4 follows.

The remark following equation (54) now furnishes a proof of the desired result.

LEMMA 5.

$$\hat{a} - \tilde{a} = o_p\left(\frac{1}{n}\right). \tag{63}$$

It might be remarked that lemma 4 practically implies that with large probability $\tilde{a}$ simultaneously maximizes $M(\theta, a)$ for all θ in some interval about θ_0. [In fact a complete proof would require a slight extension of the proof of lemma 1 to show that there is a K independent of θ such that (47) holds with θ_0 replaced by θ.]

4. Reduction to the random walk problem

In this section we shall prove that the asymptotic distribution of $n(\hat{a} - a_0)$ is related to a random walk problem. Because of lemma 5 it will suffice to treat the distribution of $n(\tilde{a} - a_0)$.

Let

$$\begin{aligned} A_r(y) &= y_1 + y_2 + \cdots + y_r && \text{for} \quad r > 0\,, \\ A_r(y) &= -\,[y_{r+1} + y_{r+2} + \cdots + y_0] && \text{for} \quad r < 0\,, \end{aligned} \tag{64}$$

$$\begin{aligned} B_r(y) &= \left(y_1 - \frac{1}{c_0}\right) + \cdots + \left(y_r - \frac{1}{c_0}\right) = A_r(y) - \frac{r}{c_0} \quad \text{for} \quad r > 0 \\ B_r(y) &= -\left[y_0 + \left(y_{-1} - \frac{1}{c_0}\right) + \cdots + \left(y_{r+1} - \frac{1}{c_0}\right)\right] = A_r(y) - \frac{r+1}{c_0} \end{aligned} \tag{65}$$

for $r < 0$, where

$$c_0 = c(\theta_0) = \frac{\beta(\theta_0, a_0) - \gamma(\theta_0, a_0)}{\log \beta(\theta_0, a_0) - \log \gamma(\theta_0, a_0)}. \tag{66}$$

Let $R(y, b)$ be that value of r which minimizes $B_r(y)$ subject to $|r| \leqq b$,

$$S(y, b) = B_{R(y,b)}(y) \tag{67}$$

and

$$T(y, b) = A_{R(y,b)}(y)\,. \tag{68}$$

It should be remarked that $R(y, b)$ and $T(y, b)$ may fail to be uniquely defined. This may be remedied by a suitable convention but we shall see that it is of little importance to do so.

Suppose that the components of $y = (\cdots, y_{-1}, y_0, y_1, \cdots)$ are independently distributed random variables where

$$\begin{aligned} \mathcal{L}(y_i) &= E\left[\frac{1}{\gamma(\theta_0, a_0)}\right] && \text{for } i > 0 \\ \mathcal{L}(y_i) &= E\left[\frac{1}{\beta(\theta_0, a_0)}\right] && \text{for } i \leqq 0 \end{aligned} \tag{69}$$

and $E(\mu)$ is the exponential distribution with mean μ and density $\exp(-x/\mu)/\mu$ for $x > 0$.

In section 7 we shall discuss the distribution of $R(y, \infty)$, $S(y, \infty)$ and $T(y, \infty)$. For the time being we shall merely use the fact that these are well defined random variables, that is, they are measurable functions and except on a set of probability measure zero they are uniquely defined and finite. The object of this section is to prove that

$$\mathcal{L}[n(\hat{a}-a_0)]\rightarrow\mathcal{L}[T(y,\infty)]. \tag{70}$$

Let $y_i^{(n)} = n(x_i^{(n)} - x_{i-1}^{(n)})$ wherever this definition makes sense, that is, for $-nF_n(a_0) \leqq i \leqq n[1 - F_n(a_0)]$. Let $y_i^{(n)} = 1/c_0$ otherwise. We first prove

LEMMA 6. *Given any* $\epsilon > 0$, *there is a* K' *such that*

$$P\{n(\tilde{a}-a_0)=T(y^{(n)},b)\}>1-2\epsilon \qquad \text{for } b>K'. \tag{71}$$

PROOF. A comparison of $M(\theta_0, a)$ with $B_r(y^{(n)})$ clearly establishes the desired result once it is shown that there is a K' such that

$$P\left\{F_n\left(a_0+\frac{K}{n}\right)-F_n\left(a_0-\frac{K}{n}\right)\leqq\frac{K'}{n}\right\}>1-\epsilon \tag{72}$$

where K is the constant in the proof of lemma 1. [There we proved that $n(\tilde{a} - a_0) \leqq K$ with large probability.] The above inequality follows immediately upon applying Chebyshev's inequality.

LEMMA 7. *As* $n \rightarrow \infty$

$$\mathcal{L}[R(y^{(n)},b),S(y^{(n)},b),T(y^{(n)},b)]\rightarrow\mathcal{L}[R(y,b),S(y,b),T(y,b)]. \tag{73}$$

PROOF. It is easily shown that

$$\mathcal{L}(y_i^{(n)},|i|\leqq b)\rightarrow\mathcal{L}(y_i,|i|\leqq b), \qquad b<\infty. \tag{74}$$

$R(y, b)$, $S(y, b)$ and $T(y, b)$ are continuous functions of y_i, $|i| \leqq b$, except possibly at those y where $R(y, b)$ is not uniquely defined. This set clearly has probability zero with respect to the limiting distribution. The lemma follows.

LEMMA 8.

$$\mathcal{L}[n(\tilde{a}-a_0)]\rightarrow\mathcal{L}[T(y,\infty)]. \tag{75}$$

PROOF. In view of lemmas 6 and 7 it suffices to prove that

$$\mathcal{L}[T(y,b)]\rightarrow\mathcal{L}[T(y,\infty)] \qquad \text{as } b\rightarrow\infty. \tag{76}$$

Since $R(y, \infty)$, $S(y, \infty)$ and $T(y, \infty)$ are random variables, there is for each $\epsilon > 0$ a K'' such that

$$P\{|R(y,\infty)|\leqq K''\}>1-\epsilon. \tag{77}$$

But then for $b \geqq K''$

$$P\{T(y,b)=T(y,\infty)\}>1-\epsilon \tag{78}$$

which implies the desired result. [In fact, $T(y, b) \rightarrow T(y, \infty)$ with probability one as $b \rightarrow \infty$.]

Applying lemmas 5 and 8 we have

THEOREM 1. *If the maximum likelihood estimates* $(\hat{\theta}, \hat{a})$ *converge to* (θ_0, a_0) *when these are the true values of the parameters and conditions* R_1, R_2 *and* R_3 *are satisfied and* $\beta(\theta_0, a_0) > \gamma(\theta_0, a_0) > 0$, *then*

$$\mathcal{L}[n(\hat{a}-a_0)]\rightarrow\mathcal{L}[T(y,\infty)]. \tag{79}$$

Remarks.

1. Essentially we have also proved that $\hat{a}$ is close to $x_r^{(n)}$ with probability almost equal to $P\{R(y, \infty) = r\}$.

2. The case where $\gamma(\theta_0, a_0) > \beta(\theta_0, a_0) > 0$ is trivially related to the one we treated.

3. Except for a scale factor, the distribution of $T(y, \infty)$ depends only on $\beta(\theta_0, a_0)/\gamma(\theta_0, a_0)$.

4. It is clear that the special problem discussed in section 2 satisfies the regularity condition with $\theta = \beta$. Since consistency was established in section 3, it follows that theorem 1 applies to this example.

5. Because of technical difficulties, it was decided not to treat the case where $\gamma(\theta_0, a_0) = 0$. One would expect that in this case $\hat{a}$ tends to be very close to the largest observation to the left of a_0.

6. A variation of this problem arises when several related points are discontinuity points of the density. For example, we may be interested in the parameters of a rectangular distribution of known range. In general suppose that $\psi_1(a), \psi_2(a), \cdots, \psi_m(a)$ are the location of m discontinuities. Then a reasonable modification of condition R_3 would require that

$$\int_{-\infty}^{\infty} \frac{\partial f(x, \theta, a)}{\partial a} dx = \sum_{i=1}^{m} \psi_i'(a_0)[\gamma_i(\theta, a_0) - \beta_i(\theta, a_0)] \tag{80}$$

where β_i and γ_i represent the left-hand and right-hand limits at the ith discontinuity point. Then $M(\theta, a_0)$ would be naturally replaced by

$$\sum_{i=1}^{m} [\log \beta_i(\theta, a_0) - \log \gamma_i(\theta, a_0)]\{F_n[\psi_i(a_0) + (a - a_0)\psi_i'(a_0)] - F_n[\psi_i(a_0)]\} - \psi_i'(a_0)[\beta_i(\theta, a_0) - \gamma_i(\theta, a_0)](a - a_0). \tag{81}$$

The value of a which is close to a_0 and maximizes this expression has an asymptotic distribution which is determined by the natural extension of the random walk problem we treated in this section.

5. Quasi-maximum likelihood estimates

In this section we shall define the quasi-maximum likelihood estimate and examine its properties. This estimate tends to be insensitive to irregularities in the distribution function and its properties can be shown to be the expected ones without the application of involved regularity conditions.

Suppose that a^* is a consistent estimate of a. Then there is a sequence $\{a_n\} \to 0$ such that $na_n \to \infty$ and $a^* - a_0 = o_p(a_n^2)$. Map the interval $[a^* - a_n, a^* + a_n]$ into $[0, 1]$ by a linear transformation. If m observations fall in $[a^* - a_n, a^* + a_n]$, they give rise to a sample c.d.f. F_m^* on $[0, 1]$ under the above transformation. It would seem natural to define the quasi-maximum likelihood estimate with respect to a^* and $\{a_n\}$ by

$$a^* + a_n[2\Phi(F_m^*) - 1] \tag{82}$$

where Φ is defined in equation (16). [$\Phi(F_n)$ is the true maximum likelihood estimate of a for the special problem of section 2.] However, this estimate has the following shortcoming. Suppose α^* is selected so that there is some observation within $\alpha^* - a_n$ and $\alpha^* - a_n + n^{-n}$. Then $\Phi(F_m^*)$ is likely to be close to 0 instead of 1/2. To avoid this possibility we define $\Phi^*(F_m^*)$ to be that value of τ in $|\tau - 1/2| \leqq a_n$ which maximizes

$$S(\tau) = F_m^*(\tau) \log \frac{F_m^*(\tau)}{\tau} + [1 - F_m^*(\tau)] \log \frac{1 - F_m^*(\tau)}{1 - \tau}. \tag{83}$$

Let our quasi-maximum likelihood estimate be

$$\alpha^{**} = \alpha^* + a_n [2 \Phi^* (F_m^*) - 1]. \tag{84}$$

THEOREM 2. *If α^* converges in probability to α_0 and*

$$\begin{aligned} \lim_{x \to \alpha_0 -} f(x, \theta_0, \alpha_0) &= \beta(\theta_0, \alpha_0) \\ \lim_{x \to \alpha_0 +} f(x, \theta_0, \alpha_0) &= \gamma(\theta_0, \alpha_0) \end{aligned} \tag{85}$$

where

$$\beta(\theta_0, \alpha_0) > \gamma(\theta_0, \alpha_0) > 0, \tag{86}$$

then

$$\mathcal{L}[n(\alpha^{**} - \alpha_0)] \to \mathcal{L}[T(y, \infty)]. \tag{87}$$

PROOF. Let τ_0 be that value in (0, 1] which corresponds to α_0, that is, $\alpha_0 - \alpha^* + a_n = 2a_n\tau_0$. Then $\tau_0 - 1/2 = o_p(a_n)$. Expanding $S(\tau)$ about $\tau = \tau_0$ and $F_m^*(\tau) = F_m^*(\tau_0)$ we have for $|\tau - 1/2| \leqq a_n$

$$S(\tau) = S(\tau_0) + [\log \xi_0 - \log \eta_0 + o_p(1)][F_m^*(\tau) - F_m^*(\tau_0)] - [\xi_0 - \eta_0 + o_p(1)](\tau - \tau_0) \tag{88}$$

where

$$\xi_0 = \frac{2\beta(\theta_0, \alpha_0)}{\beta(\theta_0, \alpha_0) + \gamma(\theta_0, \alpha_0)} \tag{89}$$

$$\eta_0 = \frac{2\gamma(\theta_0, \alpha_0)}{\beta(\theta_0, \alpha_0) + \gamma(\theta_0, \alpha_0)} \tag{90}$$

$$F_m^*(\tau) - F_m^*(\tau_0) = \frac{n}{m}\{F_n[\alpha_0 + 2a_n(\tau - \tau_0)] - F_n(\alpha_0)\}. \tag{91}$$

Let

$$\alpha = \alpha_0 + 2a_n(\tau - \tau_0). \tag{92}$$

Then

$$\frac{m}{n} = a_n[\beta(\theta_0, \alpha_0) + \gamma(\theta_0, \alpha_0)][1 + o_p(1)], \tag{93}$$

and

$$S(\tau) = S(\tau_0) + \frac{1}{a_n}\left\{\frac{\log \xi_0 - \log \eta_0}{\beta(\theta_0, \alpha_0) + \gamma(\theta_0, \alpha_0)} + [1 + o_p(1)] M^*(\tau) + (\alpha - \alpha_0) o_p(1)\right\}, \tag{94}$$

where

$$M^*(\tau) = [F_n(a) - F_n(a_0)] - c_0(a - a_0) \tag{95}$$

$$c_0 = \frac{\beta(\theta_0, a_0) - \gamma(\theta_0, a_0)}{\log \beta(\theta_0, a_0) - \log \gamma(\theta_0, a_0)}. \tag{96}$$

Now a^{**} corresponds to that value of a which maximizes $S(\tau)$ subject to $|\tau - 1/2| \leqq a_n$. Let $\tilde{a}$ be that value of a which maximizes $M^*(\tau)$ subject to $|\tau - 1/2| \leqq a_n$. The proofs of lemmas 1, 3, 5 and 8 (lemma 4 can be bypassed in this case) apply giving the desired result.

6. The random walk problem

In this section we shall discuss the distribution corresponding to the random walk problem. In doing so, we shall consider two associated problems. These problems may be attacked from a general point of view which does not essentially involve the fact that the distributions are exponential. However, the process of getting reasonable expressions for the solution involves a certain amount of ingenuity and makes heavy use of the exponential distribution. In particular, we may, at one point, apply the fundamental identity of sequential analysis because of the special properties of the exponential distribution.

The results in this section are closely related to some as yet unpublished work by J. V. Breakwell and Lionel Weiss who independently of us and each other treated the O.C. curve for the sequential probability ratio test with exponential variables.

Let $X_1, X_2, \cdots, X_n, \cdots, Z_1, Z_2, \cdots, Z_n, \cdots$ be independently and exponentially distributed with mean 1, that is, $\mathcal{L}(X_i) = \mathcal{L}(Z_i) = E(1)$. Let R_1 be that value of r for which $\sum_{i=1}^{r} (X_i - \omega_1)$ attains its minimum where $0 < \omega_1 < 1$. We shall later show that R_1 is uniquely defined with probability one. Let $S_1 = \sum_{i=1}^{R_1} (X_i - \omega_1)$. Now let R_2 be that value of r for which $\sum_{i=1}^{r} (Z_i - \omega_2)$ attains its maximum where $\omega_2 > 1$. Let $S_2 = \sum_{i=1}^{R_2} (Z_i - \omega_2)$.

LEMMA 9. *The vectors (R_1, S_1) and (R_2, S_2) are random variables and the conditional distributions of S_1 given R_1 and S_2 given R_2 are continuous.*

PROOF. Let R_{1M} minimize $\sum_{i=1}^{r} (X_i - \omega_1)$ for $1 \leqq r \leqq M$ and let $S_{1M} = \sum_{i=1}^{R_{1M}} (X_i - \omega_1)$.

Since the joint distribution of $\left\{ \sum_{i=1}^{r} (X_i - \omega_1), 1 \leqq r \leqq M \right\}$ is continuous and non-degenerate, R_{1M} is uniquely determined with probability one, (R_{1M}, S_{1M}) is a random variable and the conditional distribution of S_{1M} given R_{1M} is continuous.

Given any $\epsilon > 0$, there is, by the law of large numbers, an M such that $P\left\{ \sum_{i=2}^{r} (X_i - \omega_1) > 0 \text{ for all } r > M \right\} > 1 - \epsilon$. Hence

$$1 - \epsilon < P\{R_1 = R_{1M} \text{ and } R_1 \text{ is unique}\} \leqq P\{R_1 \text{ is unique}\}. \tag{97}$$

Since ϵ is arbitrary, $P\{R_1 \text{ is unique}\} = 1$. Furthermore, for $M \geqq r_1$, $P\{S_1 \leqq s_1 | R_1 = r_1\} = P\{S_{1M} \leqq s_1 | R_{1M} = r_1\}$. The proof for (R_2, S_2) is similar.

Examining equation (65), it becomes evident that we are interested in (R, S, T) where $\omega_1 = \gamma_0/c_0$, $\omega_2 = \beta_0/c_0$

$$R = R_1\,, \quad S = \frac{1}{\gamma_0} S_1\,, \quad T = S + \frac{R}{c_0} \quad \text{if} \quad \frac{S_1}{\gamma_0} \leqq \frac{-S_2}{\beta_0} - \frac{1}{c_0} \tag{98}$$

$$R = -R_2\,, \quad S = \frac{-S_2}{\beta_0} - \frac{1}{c_0}\,, \quad T = S + \frac{R+1}{c_0} \quad \text{if} \quad \frac{S_1}{\gamma_0} > \frac{-S_2}{\beta_0} - \frac{1}{c_0}\,. \tag{99}$$

Suppose that the densities of (R_1, S_1), (R_2, S_2) and (R, S) are given by $f_r(s)$, $g_r(s)$ and $h_r(s)$, respectively. Suppose also that the cumulative distribution functions of S_1 and S_2 are $F(s)$ and $G(s)$. We express $h_r(s)$ in terms of the other quantities.

LEMMA 10. *For $r > 0$*

$$h_r(s) = \gamma_0 f_r(\gamma_0 s)\, G(-\beta_0 s - \omega_2) \tag{100}$$

and

$$h_r(s) = \beta_0 g_{-r}[-(\beta_0 s - \omega_2)][1 - F(\gamma_0 s)] \quad \text{for} \quad r < 0\,. \tag{101}$$

The density of (R, T) is given by

$$\begin{aligned} k_r(t) &= h_r\left(t - \frac{r}{c_0}\right), & r > 0\,, \\ k_r(t) &= h_r\left(t - \frac{r+1}{c_0}\right), & r < 0\,, \end{aligned} \tag{102}$$

and the density of T is

$$\begin{aligned} k(t) &= \sum_{r=1}^{\infty} h_r\left(t - \frac{r}{c_0}\right), & t > 0\,, \\ k(t) &= \sum_{r=-\infty}^{-1} h_r\left(t - \frac{r+1}{c_0}\right), & t < 0\,. \end{aligned} \tag{103}$$

To obtain expressions for $f_r(s)$ and $g_r(s)$ we shall establish recursion relations. Suppose X_0 and Z_0 are also exponentially distributed with mean one. Let

$$\begin{aligned} R_1^* &= 1 \text{ and } S_1^* = (X_0 - \omega_1) & \text{if} \quad S_1 \geqq 0 \\ R_1^* &= R_1 + 1 \text{ and } S_1^* = (X_0 - \omega_1) + S_1 & \text{if} \quad S_1 < 0 \end{aligned} \tag{104}$$

$$\begin{aligned} R_2^* &= 1 \text{ and } S_2^* = (X_0 - \omega_2) & \text{if} \quad S_2 \leqq 0 \\ R_2^* &= R_2 + 1 \text{ and } S_2^* = (X_0 - \omega_2) + S_2 & \text{if} \quad S_2 > 0\,. \end{aligned} \tag{105}$$

It is clear that (R_1^*, S_1^*) and (R_2^*, S_2^*) have the same distributions as (R_1, S_1) and (R_2, S_2).

Hence we have

LEMMA 11.

$$f_1(s^*) = \pi_1 e^{-s^*-\omega_1} \qquad \text{for} \qquad s^* > -\omega_1 \tag{106}$$
$$f_{r+1}(s^*) = \int_{-r\omega_1}^{\min(0,\, s^*+\omega_1)} e^{-(s^*-s+\omega_1)} f_r(s)\, ds \qquad \text{for} \qquad s^* > -(r+1)\,\omega_1$$

$$\pi_1 = P\{R_1 = 1\} = P\{S_1 \geqq 0\} \tag{107}$$

$$g_1(s^*) = \pi_2 e^{-s^*-\omega_2} \qquad \text{for} \qquad s^* > -\omega_2 \tag{108}$$

$$g_{r+1}(s^*) = \int_0^{s^*+\omega_2} e^{-(s^*-s+\omega_2)} g_r(s)\, ds \qquad \text{for} \qquad s^* > -\omega_2 \tag{109}$$

$$\pi_2 = P\{R_2 = 1\} = P\{S_2 \leqq 0\}. \tag{110}$$

We shall transform the above recursion relationships to a slightly more suitable form. Let

$$\eta_r(u) = \frac{f_{r+1}[\omega_1(u - r - 1)]\, e^{u\omega_1}}{\omega_1^r \pi_1}; \tag{111}$$

then

$$\eta_0(u) = 1 \qquad \text{for} \qquad u > 0 \tag{112}$$
$$\eta_{r+1}(u^*) = \int_0^{\min(r+1,\, u^*)} \eta_r(u)\, du \qquad \text{for} \qquad u^* > 0.$$

Since $\eta_r(u)$ is an $r - i$ fold integral for $0 < u < i + 1$, $i \leqq r$, it must have $r - i$ continuous derivatives on this interval. On the other hand, $\eta_r(u)$ is clearly a polynomial of degree $r - i$ for $i < u < i + 1$, $i \leqq r$. We use the fact that

$$\eta_r(u) = \frac{u^r}{r!} \qquad \text{for} \qquad 0 < u < 1 \tag{113}$$

together with the relatively easily established identity (a special case of an identity due to Abel [4])

$$\frac{u^r}{r!} = \frac{(u-1)^r}{r!} + a_1 \frac{(u-2)^{r-1}}{(r-1)!} + \cdots + a_{r-1} \frac{(u-r)}{1!} + a_r \tag{114}$$

where

$$a_i = \frac{(i+1)^i}{(i+1)!}. \tag{115}$$

It is clear then that for $i < u < i + 1$, $i \leqq r$,

$$\eta_r(u) = \frac{u^r}{r!} - \frac{(u-1)^r}{r!} - a_1 \frac{(u-2)^{r-1}}{(r-1)!} - \cdots - a_{i-1} \frac{(u-i)^{r-i+1}}{(r-i+1)!} \tag{116}$$
$$= a_r + a_{r-1} \frac{(u-r)}{1!} + \cdots + a_i \frac{(u-i+1)^{r-i}}{(r-i)!}$$

and

$$\eta_r(u) = a_r \qquad \text{for} \qquad u > r. \tag{117}$$

Furthermore,

$$(118)\quad \pi_1 = P\{S_1 > 0\} = \sum_{r=1}^{\infty} \int_0^{\infty} f_r(s)\, ds = \pi_1 \sum_{r=0}^{\infty} \int_{r+1}^{\infty} \eta_r(u)\, e^{-u\omega_1}\, \omega_1^{r+1} du$$

$$(119)\quad 1 = \frac{1}{\omega_1} \sum_{r=0}^{\infty} a_r (e^{-\omega_1} \omega_1)^{r+1} = \frac{1}{\omega_1} \sum_{r=0}^{\infty} \frac{r^{r-1}}{r!} (e^{-\omega_1} \omega_1)^r$$

and

$$(120)\quad 1 - \pi_1 = P\{S_1 < 0\} = \pi_1 \sum_{r=0}^{\infty} \int_0^{r+1} \eta_r(u)\, e^{-u\omega_1}\, \omega_1^{r+1} du$$

$$(121)\quad \frac{1-\pi_1}{\pi_1} = \sum_{i=0}^{\infty} \sum_{r=i}^{\infty} \frac{a_i}{(r-i)!} \int_0^{i+1} (u-i-1)^{r-i} e^{-u\omega_1}\, \omega_1^{r+1} du$$

$$= \sum_{i=0}^{\infty} a_i (\omega_1 e^{-\omega_1})^{i+1} (i+1)$$

$$(122)\quad \frac{1-\pi_1}{\pi_1} = \sum_{r=1}^{\infty} \frac{r^r}{r!} (\omega_1 e^{-\omega_1})^r = (\omega_1 e^{-\omega_1}) \frac{d}{d(\omega_1 e^{-\omega_1})} \left\{ \sum_{r=1}^{\infty} \frac{r^{r-1}}{r!} (\omega_1 e^{-\omega_1})^r \right\}$$

$$(123)\quad \frac{1-\pi_1}{\pi_1} = \omega_1 e^{-\omega_1} \frac{d(\omega_1)}{d(\omega_1 e^{-\omega_1})} = \frac{\omega_1}{1-\omega_1}$$

$$(124)\quad \pi_1 = 1 - \omega_1 .$$

Now we observe that the density of S_1 is given by

$$(125)\quad f(s) = \sum_{r=0}^{\infty} f_{r+1}(s) = \frac{\pi_1}{\omega_1} \sum_{r=0}^{\infty} \eta_r \left[\frac{s}{\omega_1} + (r+1) \right] e^{-s} (\omega_1 e^{-\omega_1})^{r+1}$$

$$(126)\quad f(s) = \pi_1 e^{-s} \qquad \text{for} \qquad s > -\omega_1$$

$$(127)\quad f(s) = \frac{\pi_1 e^{-s}}{\omega_1} \left\{ \sum_{r=0}^{\infty} a_r (\omega_1 e^{-\omega_1})^{r+1} \right.$$

$$\left. + \sum_{r=1}^{\infty} a_{r-1} (\omega_1 e^{-\omega_1})^{r+1} \frac{\left(\frac{s}{\omega_1}+1\right)}{1!} + \cdots + \sum_{r=\infty} a_{r-j} (\omega_1 e^{-\omega_1})^{r+1} \frac{\left(\frac{s}{\omega_1}+j\right)^j}{j!} \right\}$$

$$\text{for} \quad -(j+1)\omega_1 < s < -j\omega_1$$

$$(128)\quad f(s) = \pi_1 e^{-s} \left\{ 1 + (\omega_1 e^{-\omega_1}) \frac{\left(\frac{s}{\omega_1}+1\right)}{1!} + \cdots + (\omega_1 e^{-\omega_1})^j \frac{\left(\frac{s}{\omega_1}+j\right)^j}{j!} \right\}$$

$$\text{for} \quad -(j+1)\omega_1 < s < -j\omega_1 .$$

The densities g_r are easily obtained from the recursion relationships. However, to obtain the over-all density $g = \sum_{r=1}^{\infty} g_r$ from g_r seems to involve considerable work. On the other hand, we shall find that the fundamental identity of sequential analysis is applicable here and furnishes a simple derivation that S_2 has an exponential distribution.

Consider the truncated process where

$$W = \sum_{i=1}^{n} (Z_i - \omega_2) \tag{129}$$

if n is the first index r less than N for which $\sum_{i=1}^{r} (Z_i - \omega_2) \geqq s$. If $\sum_{i=1}^{r} (Z_i - \omega_2) \leqq s$ for $1 \leqq r \leqq N - 1$, let $n = N$. Then the fundamental identity is

$$E\{ e^{tW} \varphi(t)^{-n}\} = 1 \tag{130}$$

where

$$\varphi(t) = E(e^{t(Z-\omega_2)}) = [e^{t\omega_2}(1-t)]^{-1}. \tag{131}$$

Note that because Z_i has an exponential distribution, the conditional distribution of Z_i given $Z_i > a$ is exponential. Hence

$$E(e^{tW} \mid n = r < N) = e^{ts}(1-t)^{-1}, \qquad s > -\omega_2, \tag{132}$$

$$E(e^{tW} \mid n = N) = E\left\{ e^{t\sum_{i=1}^{N-1}(Z_i-\omega_2)} \mid n = N\right\}(1-t)^{-1}. \tag{133}$$

But under the condition $n = N$, $0 \leqq \sum_{i=1}^{N-1} Z_i < (N-1)\omega_2 + s$. Hence $E[\exp(tW)\varphi(t)^{-n} \mid n = N] \to 0$ as $N \to \infty$ for $t > 0$ and $\exp(t\omega_2)(1 - t) < 1$. Then

$$\sum_{r=1}^{\infty} p_r(s)[e^{t\omega_2}(1-t)]^r e^{ts}(1-t)^{-1} = 1 \qquad \text{for} \qquad t > 0,\ e^{t\omega_2}(1-t) < 1, \tag{134}$$

where

$$p_r(s) = P\left\{ \sum_{i=1}^{j} (Z_i - \omega) \text{ exceeds } s \text{ for the first time at } j = r \right\}. \tag{135}$$

Let $t \to t_2+$ where t_2 is uniquely defined by

$$e^{t_2\omega_2}(1 - t_2) = 1, \qquad 0 < t_2 < 1. \tag{136}$$

Then

$$1 - G(s) = P\{S_2 \geqq s\} = \sum_{r=1}^{\infty} p_r(s) = (1 - t_2)\, e^{-t_2 s}, \qquad s > -\omega_2, \tag{137}$$

which means that S_2 has an exponential distribution. In particular

$$\pi_2 = t_2. \tag{138}$$

To obtain the densities $g_r(s)$, one may use equation (134) but we shall do so by the recursion formulas.

Let

$$\zeta_r(u) = \frac{g_{r+1}(u\omega_2)\, e^{(u+r+1)\omega_2}}{\omega_2^r \pi_2}. \tag{139}$$

Then

$$\zeta_0(u) = 1 \quad \text{for} \quad u > -1 \tag{140}$$

and

$$\zeta_{r+1}(u^*) = \int_0^{u^*+1} \zeta_r(u)\, du \quad \text{for} \quad u^* > -1. \tag{141}$$

It is easy to check that

$$\zeta_r(u) = \frac{(u+1)(u+r+1)^{r-1}}{r!} = \frac{(u+r+1)^r}{r!} - \frac{(u+r+1)^{r-1}}{(r-1)!}, u > -1, \tag{142}$$

$$g_r(s) = \frac{t_2}{\omega_2}(\omega_2 e^{-\omega_2})^r e^{-s} \frac{\left(\frac{s}{\omega_2}+1\right)\left(\frac{s}{\omega_2}+r\right)^{r-2}}{(r-1)!}, \qquad s > -\omega_2. \tag{143}$$

The closed form expression for $k(t)$, $t < 0$, seems to be difficult to obtain. The authors hope that the difficulty may be resolved in the near future so that the moments of the asymptotic distribution can be conveniently computed. It may be remarked, however, that replacing c_0 by one (as Karlin's method does for the special problem) may lead to a relatively poor asymptotic distribution in the case where β_0/γ_0 is large.

APPENDIX. REGULARITY CONDITIONS

CONDITION R_1. The following limits

$$\lim_{\substack{x\to a_0 \\ a\to a_0 \\ x<a}} f(x, \theta, a) = \beta(\theta, a_0) > 0 \tag{144}$$

and

$$\lim_{\substack{x\to a_0 \\ a\to a_0 \\ x>a}} f(x, \theta, a) = \gamma(\theta, a_0) > 0 \tag{145}$$

hold uniformly in θ for θ in some neighborhood about θ_0. Also $\beta(\theta, a_0)$ and $\gamma(\theta, a_0)$ are continuous in θ at θ_0. We may assume without loss of generality that $\beta(\theta_0, a_0) > \gamma(\theta_0, a_0)$.

CONDITION R_2. For $x \notin B(a, a_0)$ (x not between a and a_0)

$$\frac{\log f(x, \theta, a) - \log f(x, \theta, a_0)}{a - a_0} = \frac{\partial \log f(x, \theta, a_0)}{\partial a} + H(x)\, o(1) \tag{146}$$

where $E\{|H(X)|\,|\theta_0, a_0\} < \infty$.

For x and θ in some intervals about a_0 and θ_0, $\partial \log f(x, \theta, a_0)/\partial a$ is bounded.

CONDITION R_3. For some interval about θ_0,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\log f(X_i, \theta, a_0)}{\partial a} \tag{147}$$

converges uniformly in probability to

$$\delta(\theta) = E\left\{\frac{\partial \log f(X, \theta, a_0)}{\partial a} \,\middle|\, \theta_0, a_0\right\} \tag{148}$$

where $\delta(\theta)$ is continuous at θ_0 and $\delta(\theta_0) = \gamma(\theta_0, a_0) - \beta(\theta_0, a_0)$.

The last part of R_3 is related to some simple interchange of derivative and integral conditions which are apparent from the following argument.

$$\int_{-\infty}^{\infty} \frac{f(x, \theta, \alpha) - f(x, \theta, \alpha_0)}{\alpha - \alpha_0}\, dx = 0. \tag{149}$$

Suppose $\alpha > \alpha_0$. Consider the intervals $(-\infty, \alpha_0)$, (α_0, α), (α, ∞). Then

$$\int_{-\infty}^{\alpha_0} \frac{\partial f}{\partial \alpha}(x, \theta, \alpha_0)\, dx + o(1) + \int_{\alpha}^{\infty} \frac{\partial f}{\partial \alpha}(x, \theta, \alpha_0)\, dx + o(1) + [\beta(\theta, \alpha_0) - \gamma(\theta, \alpha_0)] + o(1) = 0 \tag{150}$$

$$\gamma(\theta, \alpha_0) - \beta(\theta, \alpha_0) = \int_{-\infty}^{\infty} \frac{\partial f}{\partial \alpha}(x, \theta, \alpha_0)\, dx = E\left\{\frac{\partial \log f(x, \theta, \alpha_0)}{\partial \alpha} \,\middle|\, \theta, \alpha_0\right\}. \tag{151}$$

It is clear that the special problem treated in section 2 satisfies the regularity conditions.

REFERENCES

[1] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 217–226.

[2] H. Chernoff, "Large sample theory: parametric case," *Annals of Math. Stat.*, Vol. 27 (1956), pp. 1–22.

[3] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Annals of Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[4] N. H. Abel, *Oeuvres Complètes*, Vol. 1, Christiania, C. Gröndahl, 1839, p. 102.