MAXIMUM LIKELIHOOD ESTIMATION IN TRUNCATED SAMPLES[1]

By Max Halperin[2]

USAF School of Aviation Medicine

1. Summary. In this paper we consider the problem of estimation of parameters from a sample in which only the first r (of n) ordered observations are known. If $r = [qn]$, $0 < q < 1$, it is shown under mild regularity conditions, for the case of one parameter, that estimation of θ by maximum likelihood is best in the sense that $\hat{\theta}$, the maximum likelihood estimate of θ, is

(a) consistent,
(b) asymptotically normally distributed,
(c) of minimum variance for large samples.

A general expression for the variance of the asymptotic distribution of $\hat{\theta}$ is obtained and small sample estimation is considered for some special choices of frequency function. Results for two or more parameters and their proofs are indicated and a possible extension of these results to more general truncation is suggested.

2. Introduction. We suppose we are sampling from a univariate population governed by a probability law, $f(x, \theta)$, $-\infty < x < \infty$, where θ is a single parameter. Our sampling process is assumed to be such that for any sample size, n, we have as sample observations only $x_1, x_2, \cdots, x_r$, the r smallest observations in the sample where r is defined for every n by $r = [qn]$. The notation $[a]$ has the usual meaning of the largest integer contained in a. It is assumed that q is known and $0 < q < 1$. Such a sampling process as defined above could easily arise in an experiment of the life-testing variety.

As a case in point, consider the testing of airplane propeller assemblies in a wind tunnel. The assemblies are quite expensive, costing several thousand dollars each. Furthermore, the test, which consists of increasing the wind velocity in the tunnel and observing the velocity at which each assembly is ruptured is of the destructive type. That is, if an assembly fails, it is not repairable, while if it does not fail, its function is not impaired. Thus, on the basis of budget limitations for testing purposes, it may be desirable to limit the number of assemblies that fail. An obvious solution to this problem is to terminate the testing procedure after a fixed percentage of the propellers in the sample fail. The percentage would be fixed in advance so as to keep the total monetary loss within budgetary restrictions. Supposing that the velocity required to rupture a propeller is a random variable following a continuous probability law, we have a simple example of the type of truncated sampling process described above.

[1] Part of a doctoral thesis presented to the Department of Mathematical Statistics, University of North Carolina. Work on this problem was begun while the author was with the Rand Corporation.

[2] Now at National Heart Institute, Bethesda, Maryland.

The sampling process we have described may be generalized to the case of several parameters and further to the case of several points of truncation, each of the latter being defined as a particular sample percentage point. We do not consider these generalizations in detail in this discussion.

To obtain our results we shall need the following assumptions on $f(x, \theta)$. Not all the assumptions are needed in some of our further discussions, but are listed here for brevity and easy reference.

Assumption A. *For almost all* x, *the derivatives*

$$\frac{\partial \log f(x,\theta)}{\partial \theta}, \quad \frac{\partial^2 \log f(x,\theta)}{\partial \theta^2}, \quad \frac{\partial^3 \log f(x,\theta)}{\partial \theta^3}, \tag{2.1}$$

exist for every θ *belonging to a nondegenerate interval* R.

Assumption B. *For every* θ *in* R *we have,*

$$\left|\frac{\partial f(x,\theta)}{\partial \theta}\right| < F_1(x), \qquad \left|\frac{\partial^2 f(x,\theta)}{\partial \theta^2}\right| < F_2(x),$$

$$\left|\frac{\partial^3 f(x,\theta)}{\partial \theta^3}\right| < F_3(x), \qquad \left|\frac{\partial^3 \log f(x,\theta)}{\partial \theta^3}\right| < H(x),$$

where $F_1(x)$, $F_2(x)$, $F_3(x)$ *are integrable over* $(-\infty, \infty)$, *while* $\int_{-\infty}^{\infty} H(x)f(x, \theta)\, dx < M$, *where* M *is independent of* θ.

Assumption C. *For every* θ *in* R

$$K^2 = \int_{-\infty}^{\lambda} \left(\frac{\partial \log f(x,\theta)}{\partial \theta}\right)^2 f(x,\theta)\, dx + \frac{1}{p}\left(\int_{-\infty}^{\lambda} \frac{\partial f(x,\theta)}{\partial \theta}\, dx\right)^2$$

is greater than zero. Here, if θ_0 is the true value of θ, λ is defined by $q = \int_{-\infty}^{\lambda} f(x, \theta_0)\, dx$. That is, λ is the population $100q$ percentage point.

Assumption D. $f(x, \theta)$ *is continuous in the neighborhood of* $x = \lambda$ *and has a continuous derivative in* x, $f'(x, \theta)$, *while*

$$\frac{\partial \log f(x,\theta)}{\partial \theta}, \quad \frac{\partial^2 \log f(x,\theta)}{\partial \theta^2}, \quad \frac{\partial^3 \log f(x,\theta)}{\partial \theta^3},$$

are continuous in the neighborhood of $x = \lambda$.

Finally, we define regular estimation from a joint frequency function, say $h(x_1, \cdots, x_r, \theta)$, in a manner completely analogous to that of Cramér, ([1], p. 479). That is, we suppose we can transform $x_1, \cdots, x_r$ to new variables θ^*, $\lambda_1, \cdots, \lambda_{r-1}$, (where θ^* estimates θ), in a one-to-one manner so that

$$h(x_1, \cdots, x_r; \theta) \prod_{i=1}^{r} dx_i = g(\theta^*; \theta) m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta) \prod_{i=1}^{r-1} d\lambda_i\, d\theta^*, \tag{2.2}$$

where $g(\theta^*; \theta)$ is the density of the estimate θ^*, while $m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta)$ is the conditional density of $\lambda_1, \cdots, \lambda_{r-1}$, given θ^*. Then, if $\partial h/\partial\theta$, $\partial g/\partial\theta$, $\partial m/\partial\theta$

exist for every θ in R and if

$$\left|\frac{\partial h}{\partial \theta}\right| < H_0(x_1, \cdots, x_r), \quad \left|\frac{\partial g}{\partial \theta}\right| < G_0(\theta^*), \quad \left|\frac{\partial m}{\partial \theta}\right| < M_0(\lambda_1, \cdots, \lambda_{r-1}; \theta^*),$$

where H_0, G_0, $\theta^* G_0$, and M_0 are integrable over the whole space of $(x_1, \cdots, x_r)$, θ^*, θ^*, and $\lambda_1, \cdots, \lambda_{r-1}$, respectively, we shall say we are in a regular estimation case of the continuous type and θ^* will be called a regular estimate of θ.

3. Derivation of results. Since our problem is of prominence in the field of life-testing, it is convenient to use a terminology which stems from this connection. It may be remarked that though it is then implied that our random variable is nonnegative the latter point is in no way critical to our proofs.

Thus, let $f(x, \theta)$, $0 \leqq x < \infty$, be a probability density satisfying Assumptions A–D. We suppose that n individuals, each subject to $f(x, \theta)$ as a death law, have been observed from age zero until $r(=[qn])$ of the group have died at times $x_1, x_2, \cdots, x_r$, $0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_r < \infty$. If we denote the sampling density of $x_1, \cdots, x_r$ by $h(x_1, \cdots, x_r)$, we clearly have

$$h(x_1, \cdots, x_r) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(x_i, \theta)[p(x_r, \theta)]^{n-r}, \tag{3.1}$$

where $p(x_r, \theta) = 1 - q(x_r, \theta) = \int_{x_r}^{\infty} f(x, \theta)\, dx$. If we further denote the conditional joint density of $x_1, \cdots, x_{r-1}$, given x_r, by $h(x_1, \cdots, x_{r-1}; x_r)$ and denote the density of x_r in a sample of n by $S_n(x_r)$, we have

$$\begin{aligned} h(x_1, \cdots, x_r) &= h(x_1, \cdots, x_{r-1}; x_r) S_n(x_r) \\ &= (r-1)! \prod_{i=1}^{r-1} \left[\frac{f(x_i, \theta)}{q(x_r, \theta)}\right] r \binom{n}{r} f(x_r, \theta)[p(x_r, \theta)]^{n-r}[q(x_r, \theta)]^{r-1}. \end{aligned} \tag{3.2}$$

We have now, denoting by the symbol E the operation of taking an expected value, the following lemma

LEMMA 1. *If Assumptions A and B hold,*

$$E\left[\frac{\partial \log h(x_1, \cdots, x_r)}{\partial \theta}\right]^2 = -E\left[\frac{\partial^2 \log h(x_1, \cdots, x_r)}{\partial \theta^2}\right]. \tag{3.3}$$

PROOF. The proof consists of verifying that under Assumptions A–D

$$\int_{E_r} \frac{\partial h(x_1, \cdots, x_r)}{\partial \theta} \prod_{i=1}^{r} dx_i = \int_{E_r} \frac{\partial^2 h(x_1, \cdots, x_r)}{\partial \theta^2} \prod_{i=1}^{r} dx_i = 0, \tag{3.3.1}$$

and then proceeding exactly as in Cramér ([1], p. 502). Here E_r is the domain of $x_1, \cdots, x_r$.

In order for (3.3.1) to hold we must have $|\partial h/\partial \theta| < H_0(x_1, \cdots, x_r)$, $|\partial^2 h/\partial \theta^2| < H_1(x_1, \cdots, x_r)$, where H_0 and H_1 are integrable over E_r. We have

$$\frac{\partial h}{\partial \theta} = \frac{n!}{(n-r)!} \sum_{i=1}^{r} \prod_{j \neq i} f(x_j, \theta) \frac{\partial f(x_i, \theta)}{\partial \theta} [p(x_r, \theta)]^{n-r} + \frac{n!}{(n-r-1)!} \prod_{i=1}^{r} f(x_i, \theta)[p(x_r, \theta)]^{n-r-1} \left[\frac{\partial p(x_r, \theta)}{\partial \theta}\right],$$

and

$$\left|\frac{\partial h}{\partial \theta}\right| < \frac{n!}{(n-r)!} \sum_{i=1}^{r} \prod_{j \neq i} f(x_j, \theta) F_1(x_i) + \frac{n!}{(n-r-1)!} \prod_{i=1}^{r} f(x_i, \theta) \int_0^{\infty} F_1(x)\, dx$$

from Assumption B. We also know that $\partial f(x, \theta)/\partial \theta$ exists for all θ in some interval. Thus we may choose a θ_0 in that interval and assert for all θ in the interval

$$f(x, \theta) < f(x, \theta_0) + F_1(x)\, d = F_0(x),$$

where d is the length of the continuity interval on θ. We then have

$$\left|\frac{\partial h}{\partial \theta}\right| < \frac{n!}{(n-r)!} \sum_{i=1}^{r} \prod_{j \neq i} F_0(x_j) F_1(x_i) + \frac{n!}{(n-r-1)!} \prod_{i=1}^{r} F_0(x_i) \int_0^{\infty} F_1(x)\, dx = H_0(x_1, \cdots, x_r).$$

It is clear that $H_0(x_1, \cdots, x_r)$ as just defined is integrable over E_r. A similar discussion holds for $\partial^2 h/\partial \theta^2$ and the lemma follows.

Next we prove the following lemma.

LEMMA 2. *Let* $\theta^* = \theta^*(x_1, \cdots, x_r)$ *be an unbiased estimate of* θ, θ^* *being continuous and possessing partial derivatives* $\partial \theta^*/\partial x_j (j = 1, 2, \cdots, r)$ *in almost all points* $(x_1, \cdots, x_r)$. *If estimation from* $h(x_1, \cdots, x_r)$ *is regular, we have asymptotically*

$$nE(\theta^* - \theta)^2 \geqq \frac{1}{K^2}, \tag{3.4}$$

where

$$K^2 = \int_0^{\lambda} \left[\frac{\partial \log f(x, \theta)}{\partial \theta}\right]^2 f(x, \theta)\, dx + \frac{1}{p} \left[\int_0^{\lambda} \frac{\partial f(x, \theta)}{\partial \theta}\, dx\right]^2.$$

PROOF. Consider

$$\int_{-\infty}^{\infty} g(\theta^*; \theta)\, d\theta^* = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta) \prod_{i=1}^{r-1} d\lambda_i = 1,$$

where $g(\theta^*; \theta)$ and $m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta)$ are as defined in Section 2, in the definition of a regular estimation case. Under our regularity assumptions on $m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta)$ and $g(\theta^*; \theta)$, these integrals may be differentiated with respect to θ under the integral signs. Thus we have

$$\int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial \theta}\right) g(\theta^*; \theta)\, d\theta^* = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{\partial \log m}{\partial \theta}\right) m(\lambda_1, \cdots, \lambda_{r-1}; \theta^*, \theta) \prod_{i=1}^{r-1} d\lambda_i = 0. \tag{3.4.1}$$

Now, referring back to (2.2), we take logarithms of both sides of that relationship, neglecting differentials, differentiate and have

$$\frac{\partial \log h(x_1, \cdots, x_r)}{\partial \theta} = \frac{\partial \log g(\theta^*; \theta)}{\partial \theta} + \frac{\partial \log m}{\partial \theta}. \tag{3.4.2}$$

Squaring both sides of (3.4.2) and taking expected values of each side, we get

$$\begin{aligned} \int_{E_r} \left(\frac{\partial \log h}{\partial \theta}\right)^2 h(x_1, \cdots, x_r) \prod_{i=1}^{r} dx_i &= \int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial \theta}\right)^2 g(\theta^*; \theta)\, d\theta^* \\ &+ \int_{-\infty}^{\infty} g(\theta^*; \theta) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{\partial m}{\partial \theta}\right)^2 m \prod_{i=1}^{r-1} d\lambda_i\, d\theta^* \\ &\geqq \int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial \theta}\right)^2 g(\theta^*; \theta)\, d\theta^*. \end{aligned} \tag{3.4.3}$$

The cross-product terms, resulting from squaring, all vanish by (3.4.1). Since, under our assumptions, we can write (cf. [1], p. 475), for θ^* unbiased,

$$E(\theta^* - \theta)^2 \geqq \left[\int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial \theta}\right)^2 g(\theta^*; \theta)\, d\theta^*\right]^{-1}, \tag{3.4.4}$$

it follows that we have exactly

$$nE(\theta^* - \theta)^2 \geqq \left\{\frac{1}{n} E\left[\frac{\partial \log h(x_1, \cdots, x_r)}{\partial \theta}\right]^2\right\}^{-1}, \tag{3.4.5}$$

or by (3.3)

$$\geqq -\left\{\frac{1}{n} E \frac{\partial^2 \log h(x_1, \cdots, x_r)}{\partial \theta^2}\right\}^{-1}.$$

From (3.1) we can calculate $(\partial^2 \log h(x_1, \cdots, x_r))/(\partial\theta^2)$ in detail, and integrating out $x_1, \cdots, x_{r-1}$, we get, after some manipulation,

$$\begin{aligned} -\frac{1}{n} E \frac{\partial^2 \log h(x_1, \cdots, x_r)}{\partial \theta^2} &= -\int_0^{\infty} \frac{\partial^2 q(x_{r-1}, \theta)}{\partial \theta^2} S_{n-1}(x_{r-1})\, dx_{r-1} \\ &+ \int_0^{\infty} \frac{\partial^2 q(x_r, \theta)}{\partial \theta^2} S_{n-1}(x_r)\, dx_r \\ &+ \frac{n-1}{(n-r-1)} \int_0^{\infty} \left[\frac{\partial q(x_r, \theta)}{\partial \theta}\right]^2 S_{n-2}(x_r)\, dx_r \\ &+ \int_0^{\infty} \left[\int_0^{x_{r-1}} \left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2 f(x, \theta)\, dx\right] S_{n-1}(x_{r-1})\, dx_{r-1} \\ &- \frac{1}{n} \int_0^{\infty} \frac{\partial^2 \log f(x_r, \theta)}{\partial \theta^2} S_n(x_r)\, dx_r. \end{aligned} \tag{3.4.6}$$

In (3.4.6)

$S_{n-c}(x_{r-d})$

$$= \frac{(n-c)!}{(r-d-1)!\,(n-c-r+d)!} [q(x_{r-d}, \theta)]^{r-d-1} [p(x_{r-d}, \theta)]^{n-c-r+d} f(x_{r-d}, \theta).$$

That is, $S_{n-c}(x_{r-d})$ is simply the sampling likelihood of the $(r - d)$th smallest order statistic in a sample of size $(n - c)$. With this understanding the basis for the right-hand side of (3.4.6) is readily apparent. If we consider the last term on the right-hand side of (3.4.6), we see that

$$\frac{1}{n}\left|\int_0^\infty \frac{\partial^2 \log f(x_r, \theta)}{\partial\theta^2} S_n(x_r)\, dx_r\right| \leqq \frac{1}{\sqrt{n}} \int_0^\infty \left|\frac{\partial^2 \log f(x, \theta)}{\partial\theta^2}\right| f(x, \theta)\, dx,$$

which is $O(1/\sqrt{n})$, since from our assumptions the integral exists. Hence, asymptotically we can disregard such a term. Now we consider the integrands of the remaining terms of (3.4.6) and on the integrand containing $S_{n-c}(x_{r-d})$, we perform the transformation

$$y = \frac{\sqrt{n-c}}{a}(x_{r-d} - \lambda), \tag{3.4.7}$$

where $a = (\sqrt{qp}/f(\lambda, \theta))$. It follows from Cramér [1], pp. 367–9, that the functions $(a/\sqrt{n-c})S_{n-c}(\lambda + ay/\sqrt{n-c})$ each converge to $1/\sqrt{2\pi}\exp(-\frac{1}{2}y^2)$ in any finite y interval and are each uniformly bounded in any such interval. Denoting the function associated with $S_{n-c}(\lambda + (a/\sqrt{n-c})y)$ by $g_{n-c}(\lambda + (a/\sqrt{n-c})y)$, it is apparent that we can expand g_{n-c} in series about $y = 0$ to zero-order terms plus a remainder, and consequently that

$$\lim_{n\to\infty} g_{n-c}\left(\lambda + \frac{a}{\sqrt{(n-c)}}y\right) = g(\lambda), \text{ say,} \tag{3.4.8}$$

for any fixed y. Furthermore, it is clear that each g_{n-c} is uniformly bounded in every y interval, finite or infinite. Thus the general relation desired is that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} S_n(y)g_n(y)\, dy = \frac{g}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}y^2)\, dy, \tag{3.4.9}$$

where we know that $g_n(y)$ and $S_n(y)$ converge to $g(\lambda)$ and $1/\sqrt{2\pi}\exp(-\frac{1}{2}y^2)$ respectively, for any fixed y, and that $g_n(y)$ is absolutely bounded by a constant, G, while $S_n(y)$ is uniformly bounded in any finite y interval. To establish (3.4.9) we choose a $y_0 > 0$ such that, for any preassigned $\epsilon > 0$

$$\frac{1}{\sqrt{2\pi}} \int_{|y| \leqq y_0} \exp(-\tfrac{1}{2}y^2)\, dy = 1 - \frac{\epsilon}{6(G + |g|)}.$$

We can also write

$$\left|\int_{-\infty}^{\infty} g_n(y)S_n(y)dy - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g \exp(-\tfrac{1}{2}y^2)\, dy,\right|$$

$$\leqq \left|\int_{|y| \leqq y_0} g_n(y)S_n(y)\, dy - \frac{g}{\sqrt{2\pi}}\int_{|y| \leqq y_0} g \exp(-\tfrac{1}{2}y^2)\, dy\right|$$

$$+ G\int_{|y|>y_0} S_n(y)\, dy + \frac{|g|}{\sqrt{2\pi}}\int_{|y|>y_0} \exp(-\tfrac{1}{2}y^2)\, dy.$$

Since for $| y | \leqq y_0$, $g_n(y)S_n(y)$ converges everywhere to $g/\sqrt{2\pi} \exp(-\frac{1}{2}y^2)$ and since $g_n(y)S_n(y)$ is uniformly bounded in this interval, it follows that we can choose an n_0' such that for $n > n_0'$,

$$\left| \int_{|y| \leqq y_0} g_n(y)S_n(y)\, dy - \frac{g}{\sqrt{2\pi}} \int_{|y| \leqq y_0} \exp(-\tfrac{1}{2}y^2)\, dy \right| < \tfrac{1}{2}\epsilon.$$

We also have by construction

$$\frac{|g|}{\sqrt{2\pi}} \int_{|y| > y_0} \exp(-\tfrac{1}{2}y^2)\, dy = \frac{|g|\,\epsilon}{6(G + |g|)} < \frac{\epsilon}{6}.$$

Finally, we have

$$G \int_{|y| \geqq y_0} S_n(y)\, dy = G\left[1 - \int_{|y| \leqq y_0} S_n(y)\, dy\right] > 0,$$

and for $| y | \leqq y_0$, we can choose an n_0'' such that for $n > n_0''$

$$\int_{|y| \leqq y_0} S_n(y)\, dy > \frac{1}{\sqrt{2\pi}} \int_{|y| \leqq y_0} \exp(-\tfrac{1}{2}y^2)\, dy - \frac{\epsilon}{6G},$$

so that

$$G \int_{|y| \leqq y_0} S_n(y)\, dy < G\left[1 - 1 + \frac{\epsilon}{6(G + |g|)} + \frac{\epsilon}{6G}\right] < \frac{\epsilon}{3}.$$

Thus, choosing $n_0 = \max(n_0', n_0'')$, we can assert that for any preassigned $\epsilon > 0$, we can find an n_0 such that for $n > n_0$,

$$\left| \int_{-\infty}^{\infty} g_n(y)S_n(y)\, dy - \frac{g}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}y^2)\, dy \right| < \epsilon.$$

Thus (3.4.9) holds. Then taking the limit of (3.4.6) and simplifying, (3.4) follows.

We are now ready to prove the following theorem.

THEOREM 1. *The likelihood equation*

$$(3.5) \quad \frac{\partial \log h(x_1, \cdots, x_r)}{\partial \theta} = \sum_{i=1}^{r} \frac{\partial \log f(x_i, \theta)}{\partial \theta} + (n - r) \frac{\partial \log p(x_r, \theta)}{\partial \theta} = 0,$$

corresponding to (3.1) *has a root* $\hat{\theta}$ *which*

(1) *converges in probability* (*i.p.*) *to the true value of* θ;

(2) *is asymptotically normally distributed*;

(3) *is asymptotically efficient.*

PROOF. First we show $\hat{\theta}$ converges i.p. to the true value of θ, θ_0, say. We can write

$$(3.5.1) \quad \begin{aligned} \frac{1}{n} \frac{\partial \log h(x_1, \cdots, x_r)}{\partial \theta} &= \frac{1}{n}\left(\frac{\partial \log h}{\partial \theta}\right)_{\theta_0} + \frac{(\theta - \theta_0)}{n}\left(\frac{\partial^2 \log h}{\partial \theta^2}\right)_{\theta_0} \\ &+ \tfrac{1}{2}\Delta(\theta - \theta_0)^2 T(x_1, \cdots, x_r) = B_0 + (\theta - \theta_0)B_1 + \tfrac{1}{2}\Delta(\theta - \theta_0)^2 B_2. \end{aligned}$$

Here $| \Delta | < 1$, the subscript θ_0 denotes evaluation at θ_0, and B_0, B_1, B_2 are functions of the random variables $x_1, \cdots, x_r$. We note that for the method of proof used here we must have

$$\frac{1}{n}\left|\frac{\partial^3 \log h}{\partial\theta^3}\right| < T(x_1, \cdots, x_r), \tag{3.5.2}$$

where

$$ET(x_1, \cdots, x_r) < M, \tag{3.5.3}$$

where M is a positive constant independent of θ. If we assume, in addition to Assumptions A–D, that $1/p(x_r, \theta)$ is bounded independent of θ, say by $I(x_r)$, where $E[I(x_r)] < I$, where I is independent of θ, (3.5.2) and (3.5.3) are easily seen to hold. The calculations are simple and are omitted.

Now we consider the characteristic function, $\phi_0(t)$ of B_0. We have

$$\begin{aligned}\phi_0(t) = \int_{E_r} \exp\left[\frac{it}{n}\left\{\sum_{i=1}^{r}\frac{\partial \log f(x_i, \theta)}{\partial\theta} - \frac{(n-r)}{p(x_r, \theta)}\int_0^{x_r}\frac{\partial f(x, \theta)}{\partial\theta}\,dx\right\}\right] \\ \cdot\, h(x_1, \cdots, x_r)\prod_{i=1}^{r} dx_i,\end{aligned} \tag{3.5.4}$$

or integrating on $x_1, \cdots, x_{r-1}$,

$$= \int_0^\infty \left[U\left(\frac{t}{n}, x_r\right)\right]^{r-1} V\left(\frac{t}{n}, x_r\right)\left[W\left(\frac{t}{n}, x_r\right)\right]^{n-r} S_n(x_r)\,dx_r,$$

where

$$U\left(\frac{t}{n}, x_r\right) = \frac{\int_0^{x_r}\exp\left\{\frac{it}{n}\frac{\partial \log f(x, \theta)}{\partial\theta}\right\}f(x, \theta)\,dx}{q(x_r, \theta)},$$

$$V\left(\frac{t}{n}, x_r\right) = \exp\left\{\frac{it}{n}\frac{\partial \log f(x_r, \theta)}{\partial\theta}\right\},$$

$$W\left(\frac{t}{n}, x_r\right) = \exp\left\{-\frac{it}{n}\frac{\partial q(x_r, \theta)}{\partial\theta}\frac{1}{p(x_r, \theta)}\right\}.$$

Now on the integrand of (3.5.4), we perform the transformation (3.4.7) and have

$$\begin{aligned}\frac{a}{\sqrt{n}}\left[U\left(\frac{t}{n}, \lambda + \frac{ay}{\sqrt{n}}\right)\right]^{r-1} V\left(\frac{t}{n}, \lambda + \frac{ay}{\sqrt{n}}\right) \\ \cdot\left[W\left(\frac{t}{n}, \lambda + \frac{ay}{\sqrt{n}}\right)\right]^{n-r} S_n\left(\lambda + \frac{ay}{\sqrt{n}}\right).\end{aligned} \tag{3.5.5}$$

We want the limit for fixed y and t of (3.5.5). From [1], pp. 367–69, we have

$$\lim_{n\to\infty}\frac{a}{\sqrt{n}} S_n\left(\lambda + \frac{ay}{\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}y^2\right)$$

for every fixed y. Also

$$\lim_{n\to\infty} \log V\left(\frac{t}{n}, \lambda + \frac{ay}{\sqrt{n}}\right) = \lim_{n\to\infty} \frac{it}{n} \frac{\partial \log f\left(\lambda + \frac{ay}{\sqrt{n}}\right)}{\partial\theta} = 0,$$

for every fixed y and t, from the continuity of $\partial \log f(x, \theta)/\partial\theta$ about $x = \lambda$. If we now consider the function $U(t, \lambda + ay)$, we see that for every fixed y it is a characteristic function. Further $(\partial U/\partial t)_{t=0}$ exists for every fixed y. Thus we can expand $U(t, \lambda + ay)$ in a series in t and have for values of t near zero

$$U(t, \lambda + ay) = 1 + \int_0^{\lambda+ay} \frac{\partial f(x, \theta)}{\partial\theta} dx \frac{it}{q(\lambda + ay, \theta)}$$

$$(3.5.6) \qquad + \frac{\left[\int_0^{\lambda+ay} \frac{\partial f(x, \theta)}{\partial\theta} \exp\left\{i\Delta_1 t \frac{\partial \log f(x, \theta)}{\partial\theta}\right\} dx - \int_0^{\lambda+ay} \frac{\partial f(x, \theta)}{\partial\theta} dx\right] it}{q(\lambda + ay, \theta)},$$

where $|\Delta_1| < 1$ and the term in square brackets goes to zero with t. We can also write, for example

$$\frac{\int_0^{\lambda+ay} \frac{\partial f(x, \theta)}{\partial\theta} dx}{q(\lambda + ay)} = \frac{\int_0^{\lambda} \frac{\partial f(x, \theta)}{\partial\theta} dx}{q}$$

$$+ \left[\frac{ay}{q(\lambda + a\Delta_2 y, \theta)} \frac{\partial f(\lambda + a\Delta_2 y, \theta)}{\partial\theta} - \frac{af(\lambda + a\Delta_2 y, \theta) \int_0^{\lambda+a\Delta_2 y} \frac{\partial f(x, \theta)}{\partial\theta} dx}{q(\lambda + a\Delta_2 y, \theta)}\right],$$

where $|\Delta_2| < 1$. Then putting t/n and $y/\sqrt{n}$ for t and y respectively, we get

$$(3.5.7) \qquad U\left(\frac{t}{n} + \frac{ay}{\sqrt{n}}\right) = 1 + \frac{\int_0^{\lambda} \frac{\partial f(x, \theta)}{\partial\theta} dx}{q} \frac{it}{n} + \frac{\rho_1(n, t, y)}{n},$$

where $\rho_1(n, t, y)$ approaches zero for any fixed y and t, as $n \to \infty$. Similar considerations lead us to

$$(3.5.8) \qquad W\left(\frac{t}{n}, \lambda + \frac{ay}{\sqrt{n}}\right) = 1 - \frac{\int_0^{\lambda} \frac{\partial f(x, \theta)}{\partial\theta} dx}{p} \frac{it}{n} + \frac{\rho_2(n, t, y)}{n},$$

where $\rho_2(n, t, y)$ approaches zero for any fixed y and t as $n \to \infty$. It follows that (3.5.5) becomes, asymptotically,

$$(3.5.9) \qquad \exp\left[it\left\{\int_0^{\lambda} \frac{\partial f(x, \theta)}{\partial\theta} dx - \int_0^{\lambda} \frac{\partial f(x, \theta)}{\partial\theta} dx\right\}\right] \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}y^2\right).$$

Since (3.5.5) meets the conditions indicated for validity of (3.4.9), we can apply a convergence argument as indicated in Lemma 2 and conclude

$$\lim_{n\to\infty} \phi_0(t) = \exp\left[it\left\{\int_0^\lambda \frac{\partial f(x, \theta)}{\partial \theta}\, dx - \int_0^\lambda \frac{\partial f(x, \theta)}{\partial \theta}\, dx\right\}\right], \tag{3.5.10}$$

so that B_0 converges i.p. to zero.

Similar arguments lead us to

$$\lim_{n\to\infty} E \exp(itB_1) = \exp(-K^2 it), \tag{3.5.11}$$

and

$$\lim_{n\to\infty} E \exp(itB_2) = \exp(M'it), \tag{3.5.12}$$

so that B_1 converges i.p. to $-K^2$ while B_2 converges i.p. to $M' < M$, a positive constant independent of θ. The precise argument given in [1], pp. 502–3, may then be employed to show that (3.5) has a solution, $\hat{\theta}$, which converges in probability to θ_0. We omit these arguments.

Now from (3.5.1), we have

$$K\sqrt{n}(\theta - \theta_0) = \frac{\dfrac{1}{K\sqrt{n}}\left(\dfrac{\partial \log h}{\partial \theta}\right)_{\theta_0}}{-\dfrac{B_1}{K^2} - \dfrac{\Delta}{2K^2} B_2(\hat{\theta} - \theta_0)}. \tag{3.5.13}$$

The denominator of the right-hand side of (3.5.13) converges i.p. to 1, so that we may infer by well known theorems that the asymptotic distribution of the ratio is simply the asymptotic distribution of the numerator. Thus we need only show that $(1/K\sqrt{n})(\partial \log h/\partial\theta)_{\theta_0}$ is asymptotically normal with zero mean and unit variance in order to complete the proof of our theorem. Denoting the characteristic function of $(1/K\sqrt{n})(\partial \log h/\partial\theta)_{\theta_0}$ by $\phi(t)$, we have, by virtue of (3.5.4)

$$\phi(t) = \phi_0\left(\frac{\sqrt{n}\, t}{K}\right). \tag{3.5.14}$$

Applying the transformation (3.4.7) to (3.5.14) we can exactly as before show that $V(t/K\sqrt{n}, \lambda + ay/\sqrt{n})$ converges to 1 for every fixed y and t, while $(a/\sqrt{n})S_n(\lambda + ay/\sqrt{n})$ converges to $(1/\sqrt{2\pi}) \exp -\frac{1}{2}y^2$ for every fixed y. If now we turn our attention to $U(t, \lambda + ay)$, we see that U can be expanded near $t = 0$ in powers of t to terms of order t^2 plus a remainder of order $o(t^2)$, since the second moment of the distribution corresponding to U exists for every fixed y. Similar remarks apply to $W(t, \lambda + ay)$. Thus, by manipulation of the type employed in obtaining an asymptotic representation of (3.5.5), we find that the similar result for the integrand of (3.5.14) is

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\left\{y^2 - \frac{2i}{K\sqrt{pq}}\left(\int_0^\lambda \frac{\partial f(x, \theta)}{\partial\theta}\, dx\right) ty\right.\right. \\ &\qquad \left.\left. + \int_0^\lambda \left(\frac{\partial \log f(x, \theta)}{\partial\theta}\right)^2 f(x, \theta)\, dx - \frac{1}{q}\left(\int_0^\lambda \frac{\partial f(x, \theta)}{\partial\theta}\, dx\right)^2 \frac{t^2}{K^2}\right\}\right]. \end{aligned} \tag{3.5.15}$$

Since (3.5.14) meets the conditions indicated for validity of (3.4.9), we can for any fixed t, carry out a convergence argument as indicated in Lemma 2, to obtain

$$(3.5.16)\quad \begin{aligned}\lim_{n\to\infty} \phi(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}t^2\right) \\ &\cdot \exp\left[-\frac{1}{2}\left\{y + \frac{it}{K\sqrt{pq}} \int_0^{\lambda} \frac{\partial f(x,\theta)}{\partial\theta}\, dx\right\}^2\right] dy = \exp\left(-\tfrac{1}{2}t^2\right).\end{aligned}$$

Theorem 1 follows.

4. Generalizations. One can generalize the discussion of Section 3 to the case of several parameters and show that maximum likelihood estimation of $\theta = (\theta_1, \cdots, \theta_p)$ from $h(x_1, \cdots, x_r)$ is a best estimation procedure in the sense that

(a) the maximum likelihood equations have a set of solutions $(\hat{\theta}_1, \cdots, \hat{\theta}_p)$ which are consistent;

(b) $\sqrt{n}(\hat{\theta} - \theta)$ has a multivariate normal limit law;

(c) the covariance matrix of the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is the best matrix in the sense of Cramér [1].

In connection with (c) we mean specifically that the concentration ellipsoid corresponding to the covariance matrix of the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is identical with

$$(4.2)\quad \sum_{i,j=1}^{p} E \frac{\partial \log h}{\partial\theta_i} \frac{\partial \log h}{\partial\theta_j} (\mu_i - \theta_i)(\mu_j - \theta_j) = p + 2.$$

The ellipsoid (4.2) is shown by Cramér [1] to lie wholly within the ellipsoid corresponding to any set of regular unbiased estimates of $\theta_1, \cdots, \theta_p$. The meaning of "regular" here is precisely in the sense of Cramér [1] as applied to the joint frequency function $h(x_1, \cdots, x_r)$. The assumptions necessary to obtain the result are the natural analogues of Assumptions A–D. Thus A, B, D are extended by imposing similar conditions upon the various derivatives up to third order, that is those with respect to each θ_i and also the mixed derivatives. The condition C becomes a requirement that the matrix with elements

$$\begin{aligned}A_{ij} = \int_0^{\lambda} &\left(\frac{\partial \log f(x,\theta)}{\partial\theta_i}\right)\left(\frac{\partial \log f(x,\theta)}{\partial\theta_j}\right) f(x,\theta)\, dx \\ &+ \frac{1}{p}\left(\int_0^{\lambda} \frac{\partial f(x,\theta)}{\partial\theta_i}\, dx\right)\left(\int_0^{\lambda} \frac{\partial f(x,\theta)}{\partial\theta_j}\, dx\right), \qquad i, j = 1, 2, \cdots, p,\end{aligned}$$

be positive definite. The additional assumption on $p(x_r, \theta)$ specified in Theorem 1 remains unchanged except that θ is taken to be a vector parameter.

Under the assumptions outlined above the proof of (a), (b), (c) follows the lines of Section 3.

A direction of further generalization is to the case of several points of truncation, each truncation point being a sample percentage point. This work has not been carried out in detail, but due to the asymptotic joint normality of sample percentage points, it appears clear that results of the nature of (a), (b), (c) would hold under conditions analogous to these given by Assumptions A–D.

5. Small-sample estimation. For samples of the type considered in Section 3, one can obtain small-sample results for two important special choices of $f(x, \theta)$.

CASE A. $$f(x, \theta) = \theta e^{-\theta x}, \qquad 0 \leqq x < \infty, \theta > 0.$$

For this case we have from (3.1)

$$(5.1) \quad h(x_1, \cdots, x_r) = \frac{n!}{(n-r)!}\theta^r \exp\left\{-\theta\left[\sum_{i=1}^{r-1} x_i + (n-r+1)x_r\right]\right\}.$$

From (3.4) it appears that in any regular estimation case an estimate θ^* will be such that

$$(5.1.1) \qquad E(\theta^* - \theta)^2 \geqq \frac{\theta^2}{nq},$$

for θ^* unbiased. From (5.1) we obtain the maximum likelihood estimate of θ as

$$(5.1.2) \qquad \hat{\theta} = \frac{r}{\sum_{i=1}^{r-1} x_i + (n-r+1)x_r} = \frac{r}{y}, \quad \text{say.}$$

It is easy to show by calculating its moment generating function that the random variable, $2\theta y$, has a chi-square distribution with $2r$ degrees of freedom. It follows that

$$(5.1.3) \qquad E\hat{\theta} = \frac{r\theta}{r-1} \sim \theta, \qquad \operatorname{Var} \hat{\theta} = \frac{\theta^2}{r-1} \sim \frac{\theta^2}{nq}.$$

Thus $\hat{\theta}$ is a best estimate in the sense of Theorem 1. $\hat{\theta}$ can, of course, be corrected for bias, the variance of the adjusted estimate then being $\theta^2/(r-2)$.

CASE B. $$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right], \qquad -\infty < x < \infty.$$

This case is of marked interest since it is frequently assumed in life testing that the logarithm of time to death is normally distributed. Essentially this case has also been considered by Hald [3] and Cohen [5]. We indicate the solution for completeness. It can be shown that

$$(5.2) \qquad \hat{\sigma} = \tfrac{1}{2}(x_r - \bar{x}_r)(-\hat{h} + \sqrt{\hat{h}^2 + V^2}), \qquad \mu = x_r - \hat{h}\hat{\sigma},$$

where

$$\bar{x}_r = \frac{1}{r}\sum_{i=1}^{r} x_i, \qquad V^2 = 4\left[1 + \frac{\sum_{i=1}^{r-1}(x_i - \bar{x}_r)^2}{r(x_r - \bar{x}_r)^2}\right],$$

and $\hat{h}$ is the solution of

$$(5.2.1)\qquad \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}h^2\right)}{\frac{1}{\sqrt{2\pi}} \int_h^\infty \exp\left(-\frac{1}{2}z^2\right) dz} = \frac{-r}{n-r} h + \frac{2r}{n-r} \frac{h + \sqrt{h^2 + V^2}}{V^2}.$$

It is easy to show that the right and left sides of (5.2.1) are monotone decreasing and increasing respectively. This implies the uniqueness of the solution and also affords a simple method of solving (5.2.1) with the aid of a table of ordinates and areas of the standardized normal distribution. Despite the fairly formidable appearance of (5.2.1) the solution goes quickly.

It is also simple and interesting to calculate the asymptotic efficiency of the estimate of μ from a truncated sample when σ is known (the efficiency being considered relative to a completely known sample). For selected values of $q = r/n$, approximate efficiencies are given in thet following table. Approximate efficiency of $x_{[qn]}$, the sample percentage point, in estimating μ is also given to indicate the extent to which one gains by using the other $(r - 1)$ observations in the estimation procedure.

q	.1	.2	.3	.4	.5	.6	.7	.8	.9
Eff $\hat{\mu}$	.36	.53	.66	.75	.82	.88	.91	.95	.98
Eff $x_{[qn]}$	.33	.49	.57	.62	.64	.62	.57	.49	.33

Acknowledgment. The author is indebted to Dr. A. M. Mood who suggested this problem.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] H. Cramér, "A contribution to the theory of statistical estimation," *Skandinavisk Aktuarietidskrift*, Vol. 29 (1946), pp. 85–94.

[3] A. Hald, "Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point," *Skandinavisk Aktuarietidskrift*, Vol. 32 (1949), pp. 119–34.

[4] A. M. Mood, "An asymptotic distribution for a mortality problem," *Rand Corporation document RM-111*, 1949, unpublished.

[5] A. C. Cohen, "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 557–569.