OPTIMUM ESTIMATORS OF THE PARAMETERS OF NEGATIVE EXPONENTIAL DISTRIBUTIONS FROM ONE OR TWO ORDER STATISTICS

By M. M. Siddiqui

Boulder Laboratories, National Bureau of Standards

1. Introduction and summary. Let

$$f_1(x) = \sigma^{-1} \exp(-x/\sigma), \text{ if } x \geqq 0; 0, \text{ otherwise};$$

$$f_2(x) = \sigma^{-1} \exp[-(x - \alpha)/\sigma], \text{ if } x \geqq \alpha; 0, \text{ otherwise}.$$

Let x_k denote the kth order statistic of a random sample of size n. Harter [1] discusses the following three problems designated here as P_1, P_2, and P_3:

P_1 : Best unbiased estimator of the form $c_k x_k$ for σ of $f_1(x)$;

P_2 : Best unbiased estimator of the form $c_l x_l + c_m x_m$ for σ of $f_1(x)$;

P_3 : Best unbiased estimators of the form $c_l x_l + c_m x_m$ for σ, α, and the mean, μ, of $f_2(x)$. For P_3 he shows that the optimum l is equal to 1 and that the same m is optimum for all three parameters. In each problem, after setting up the equation for the relative efficiency of a linear combination of one or two order statistics, he remarks that he is not aware of any analytical method for determining the best combination, and hence finds them by exhaustive numerical computations for n up to 100. In this paper an analytical method for his problems will be presented. For P_1 and P_3 the correct optimum values of k and m are readily determined *for all n*. These will be given in Sections 2 and 3. The equations for P_2, however, are quite difficult to solve. The analytical formulation of P_2 and an approximate solution, arrived at by trial and error, will be presented in Section 4.

The method is based on the Euler-Maclaurin formula

$$\sum_{r=0}^{k-1} f(r) = \int_0^k f(x)\,dx - \frac{1}{2}[f(k) - f(0)] + \left(\frac{1}{12}\right)[f^{(1)}(k) - f^{(1)}(0)] - \left(\frac{1}{720}\right)[f^{(3)}(k) - f^{(3)}(0)] + \cdots. \tag{1.1}$$

For a discussion of the remainder after a finite number of terms on the right we refer to [2]. It is sufficient to note here that this is an asymptotic expansion and the most accurate result is obtained by taking the sum to one-half of the smallest term.

2. Estimating σ of $f_1(x)$ from one x_k. $c_k x_k$ is an unbiased estimator of σ, where [1]

$$c_k = 1/\sum_1^k a_i, \qquad a_i = 1/(n - i + 1), \tag{2.1}$$

Received February 21, 1962.

with efficiency relative to the sample mean given by

$$E_k = \left(\sum_1^k a_i\right)^2 \Big/ n \sum a_i^2. \tag{2.2}$$

Using (1.1) and setting $y = (n - k + 1)/(n + 1)$, we have

$$\begin{cases} \displaystyle\sum_1^k a_i = \sum_{i=0}^{k-1} \frac{1}{n - k + 1 + i} = -\ln y + \frac{1}{2(n+1)} \frac{1-y}{y} \\ \displaystyle\qquad\qquad + \frac{1}{12(n+1)^2} \frac{1-y^2}{y^2} + \cdots, \\ \displaystyle\sum_1^k a_i^2 = \frac{1}{n+1} \frac{1-y}{y} + \frac{1}{2(n+1)^2} \frac{1-y^2}{y^2} \\ \displaystyle\qquad\qquad + \frac{1}{6(n+1)^3} \frac{1-y^3}{y^3} + \cdots. \end{cases} \tag{2.3}$$

To form an idea of the error of approximation involved, for example, in taking only three terms on the right of each equation in (2.3), we mention that: (1) for a fixed n the error is maximum when $k = n$; (2) for $n = 4$ the maximum relative error for $\sum a_i$ is less than 0.005, and for $\sum a_i^2$ less than 0.015; (3) in each case, the maximum relative error decreases as n increases; (4) the relative error in E_k (when three term approximations for $\sum a_i$ and $\sum a_i^2$ are used) does not exceed the relative error for $\sum a_i^2$.

Now let $g(y) = nE_k/(n + 1)$, so that from (2.2) and (2.3)

$$\begin{aligned} &\left\{\left(\frac{1-y}{y}\right) + \frac{1}{2(n+1)}\left(\frac{1-y^2}{y^2}\right) + \cdots\right\} g(y) \\ &\qquad = \left\{-\ln y + \frac{1}{2(n+1)}\left(\frac{1-y}{y}\right) + \cdots\right\}^2. \end{aligned} \tag{2.4}$$

Considering $g(y)$ as a function of a real variable y, we wish to determine y_0, $(n + 1)^{-1} \leqq y_0 \leqq n(n + 1)^{-1}$, such that $g(y_0)$ is a maximum. The optimum k is then found by taking $n - k + 1$ to be the nearest integer to $(n + 1)y_0$. Differentiating (2.4), and setting $g'(y) = 0$, we obtain, after some simplification, the equation

$$\begin{aligned} &\left\{1 + \frac{1}{(n+1)y} + \cdots\right\}\left\{-\ln y + \frac{1}{2(n+1)}\left(\frac{1-y}{y}\right) + \cdots\right\} \\ &\quad = 2\left\{y + \frac{1}{2(n+1)} + \cdots\right\}\left\{\left(\frac{1-y}{y}\right) + \frac{1}{2(n+1)}\left(\frac{1-y^2}{y^2}\right) + \cdots\right\}. \end{aligned} \tag{2.5}$$

y_0 is a solution to this equation. If y is fixed and $n \to \infty$, (2.5) reduces to $-\ln y = 2(1 - y)$, which has a solution $y_1 = 0.20319$ correct to five decimal places. The other solution, $y = 1$, which is also a solution to (2.5), is rejected as it leads to the value $k = 0$. If $n \geqq 4$, this asymptotic solution y_1 is in the desired range $[(n + 1)^{-1}, n(n + 1)^{-1}]$.

We then observe that if $n \geqq 4$ an evaluation of $g'(y)$ at y_1 and $y_2 = y_1 + 0.5(n+1)^{-1}$ shows that $g'(y_1) > 0$ and $g'(y_2) < 0$. Hence

$$y_1 < y_0 < y_1 + 0.5(n+1)^{-1}. \tag{2.6}$$

We then develop y_0 in the series $y_0 = y_1 + a_1(n+1)^{-1} + a_2(n+1)^{-2} + \cdots$, and obtain $a_1 = 0.39841$, $a_2 = -1.16312$. Thus

$$y_0 \cong 0.20319 + 0.39841(n+1)^{-1} - 1.16312(n+1)^{-2}. \tag{2.7}$$

This determination of y_0 together with (2.6) is sufficiently accurate for our purposes to yield the optimum value of k. The optimum k is the nearest integer to

$$(n+1)(1-y_0) \cong 0.79681(n+1) - 0.39841 + 1.16312(n+1)^{-1}. \tag{2.8}$$

As a check one may compare the values of k thus determined with Harter's values for $n = 4$ through 100 and find that they are always correct. They are correct even for $n = 2$ and 3. Only on very rare occasions, when the fractional part of $(n+1)(1-y_0)$ as calculated from (2.8) is very close to 0.5, there may be an ambiguity whether to take the integer just above or just below $(n+1)(1-y_0)$. In practice either of the two may be considered optimum as the efficiencies of the estimates corresponding to these integers will be almost the same. In any case a further term in the series of y_0, or a comparison of the efficiencies of the corresponding estimators, can decide between the two.

3. Estimating the parameters of $f_2(x)$. We will postpone P_2 to the following section. In this section we will consider P_3 due to its similarity with P_1. From Harter's discussion after his equation (33), it is evident that the problem of optimum estimators of α, σ and the mean, μ, of $f_2(x)$ reduces to finding the optimum m which maximizes

$$E_{\hat{\sigma}} = \left(\sum_2^m a_i\right)^2 \bigg/ \left[(n-1)\left(\sum_2^m a_i^2\right)\right]. \tag{3.1}$$

Using (1.1) to approximate the summations involved we end up again with equations (2.4) and (2.5), this time with $y = (n-m+1)/n$, $g(y) = (n-1)E_{\hat{\sigma}}/n$, and $(n+1)$ replaced by n elsewhere. Thus the optimum y_0 has the same development as in (2.7) with $n+1$ replaced by n, i.e.,

$$y_0 \cong 0.20319 + 0.39841n^{-1} - 1.16312n^{-2}. \tag{3.2}$$

The optimum m is determined by taking $n - m + 1$ to be the closest integer to ny_0, i.e., m is the closest integer to

$$n - ny_0 + 1 \cong 0.79681n + 0.60159 + 1.16312n^{-1}. \tag{3.3}$$

For example, if $n = 6$, 14, 34, and 88, the optimum values of m are 6, 12, 28, and 71 respectively, which compare exactly with the values found by Harter [1, pp. 1088–89].

4. Estimating σ of $f(x)$ from two order statistics. The same technique as above when applied to the problem of finding the optimum among the unbiased estimators of the type $c_l x_l + c_m x_m$ for σ of $f_1(x)$, leads to quite a complicated pair of equations. The problem is [1, p. 1080] to find the optimum l and m such that

$$E_{lm} = \left(\sum_1^l a_i + \lambda \sum_1^m a_i\right)^2 \Big/ n\left[(1 + 2\lambda) \sum_1^l a_i^2 + \lambda^2 \sum_1^m a_i^2\right] \tag{4.1}$$

is maximized, where $n \geqq m > l \geqq 1$, and

$$\lambda = \sum_{l+1}^m a_i \sum_1^l a_i^2 \Big/ \left(\sum_1^l a_i \sum_1^m a_i^2 - \sum_1^m a_i \sum_1^l a_i^2\right). \tag{4.2}$$

If a summation is approximated by only the corresponding integral in (1.1), then, setting $x = (n - l + 1)/(n + 1)$, $y = (n - m + 1)/(n + 1)$, we have

$$\frac{n}{n+1} E_{lm} \cong \frac{(\ln x + \lambda \ln y)^2 xy}{(1 + 2\lambda)(1 - x)y + \lambda^2 x(1 - y)} = g(x, y), \quad \text{say,} \tag{4.3}$$

$$\lambda \cong \frac{y(1 - x)\ln(x/y)}{y(1 - x)\ln y - x(1 - y)\ln x}. \tag{4.4}$$

To find (x_0, y_0), $0 < y_0 < x_0 < 1$, such that $g(x_0, y_0)$ is a maximum we set $\partial g/\partial x = 0$, and $\partial g/\partial y = 0$. The resulting equations seemed to be intractable. However, an examination of these equations near $y = 0$ indicated that for the required solution $\ln y$ should be taken near -2.5, and $\ln x$ near -1, i.e., y near 0.08 and x near 0.37. A numerical study of the values of $g(x, y)$ near the point (0.37, 0.08) then indicated that $x_0 = 0.361$, $y_0 = 0.073$, with $g(x_0, y_0) = .820262$. The surface is quite flat near this point, for example, $g(.362, .074) = .820261$, $g(.362, .075) = .820241$; hence a very exact determination of (x_0, y_0) is not very essential. For an asymptotic solution of P_2 we then take l to be the nearest integer to $0.639(n + 1)$ and m to $0.927(n + 1)$. For example, if $n = 11, 37, 56, 81$, and 94 we obtain $(l, m) = (8, 11), (24, 35), (36, 53), (52, 76)$ and (61, 88), respectively. A comparison with Harter's values, which are (8, 11), (24, 35), (36, 52), (53, 76), and (61, 88) respectively, shows that the asymptotic solution is either optimum or very near to the optimum.

5. Acknowledgment. The referee has brought to the writer's attention a paper by Sarhan, Greenburg and Ogawa [3], [4] in which they discuss the more general problem of obtaining estimators for the parameters of an exponential distribution which are linear in arbitrary number (not necessarily one or two) of order statistics. In that paper the problem of optimization is solved by maximizing the asymptotic, rather than the exact, relative efficiency of an estimator. The referee also suggests that, for P_2, if we take l to be the nearest integer to $0.6386\ (n + \frac{1}{2})$ and m to $0.9266\ (n + \frac{1}{2})$ the approximation is improved.

REFERENCES

[1] HARTER, H. LEON (1961). Estimating the parameters of negative exponential population from one or two order statistics. *Ann. Math. Statist.* **32** 1078–1090.

[2] JEFFREYS, SIR HAROLD and LADY (1956). *Methods of Mathematical Physics*, 3rd edition. Cambridge. 279–281.

[3] SARHAN, A. E., GREENBURG, B. G. and OGAWA, JUNJIRO (1960). Simplified estimates for the exponential distribution. Interim Technical Report No. 1, University of North Carolina.

[4] SARHAN, A. E., GREENBURG, B. G. and OGAWA, JUNJIRO (1963). Simplified estimates for the exponential distribution. *Ann. Math. Statist.* **34** 102–116.