The Annals of Statistics
1973, Vol. 1, No. 3, 562-567

A NOTE ON THE OPTIMUM SPACING OF SAMPLE QUANTILES FROM THE SIX EXTREME VALUE DISTRIBUTIONS

BY GUNNAR KULLDORFF

Texas A & M University and University of Umeå

The problems of asymptotically optimum spacing of sample quantiles in connection with asymptotically best linear unbiased estimation of the location parameter of an extreme value distribution of Type I or the scale parameter of an extreme value distribution of Type II or III, are essentially the same. Previous results for the scale parameter of an exponential distribution can therefore be used.

1. The extreme value distributions. It is well known (Gnedenko (1943), Gumbel (1958)) that if a nondegenerate limiting distribution of the largest value in a random sample, after suitable linear standardization, exists, it must be one of the following three distributions

$$F_\nu(x) = G_\nu\left(\frac{x-\alpha}{\beta}\right) \qquad (\nu = 1, 2, 3)\,, \tag{1.1}$$

where

$$G_1(x) = \exp(-e^{-x}) \qquad (-\infty < x < \infty)\,, \tag{1.2}$$

$$G_2(x) = \exp(-x^{-\gamma}) \qquad (0 \leqq x < \infty)\,, \tag{1.3}$$

$$G_3(x) = \exp[-(-x)^\gamma] \qquad (-\infty < x \leqq 0)\,. \tag{1.4}$$

The scale parameter β and the shape parameter γ are positive. The distribution $F_\nu(x)(\nu = 1, 2, 3)$ is often called the extreme value distribution of Type ν for largest values. $F_1(x)$ is sometimes called *the* extreme value distribution for largest values.

Similarly we have only three possible limiting distributions of the smallest value in a random sample, after suitable standardization, namely

$$F_\nu(x) = 1 - G_{\nu-3}\left(\frac{\alpha-x}{\beta}\right) \qquad (\nu = 4, 5, 6)\,. \tag{1.5}$$

The distribution $F_\nu(x)(\nu = 4, 5, 6)$ is often called the extreme value distribution of Type $\nu - 3$ for smallest values. $F_4(x)$ is sometimes called *the* extreme value distribution for smallest values, and it has also been called the log-Weibull distribution. $F_6(x)$ is often called the Weibull distribution, and for $\gamma = 1$ it becomes the exponential distribution with

$$F(x) = 1 - \exp\left(-\frac{x-\alpha}{\beta}\right) \qquad (\alpha \leqq x < \infty)\,. \tag{1.6}$$

Received February 1971.

AMS 1970 *subject classification.* Primary 62F10.

Key words and phrases. Order statistics, sample quantiles, asymptotically best linear unbiased estimate, optimum spacing, extreme value distributions, exponential distribution, Weibull distribution, censored sample.

The extreme value distributions $F_\nu(x)(\nu = 1, 2, \cdots, 6)$ have many fields of application, including floods, droughts, rainfalls, snowfalls, temperatures, wind speeds, earthquakes, aeronautics, naval engineering, corrosion, strength of material and survival times of microorganisms. See Gumbel (1958) and Johnson and Kotz (1970, Chapters 20 and 21) for references.

2. Estimation based on sample quantiles. Suppose that a large number of independent observations are made on a random variable with the distribution function $G[(x - \alpha)/\beta]$, where $\beta > 0$ and the function $G(x)$ is known for all x and has the derivative $g(x)$. We also assume that one of the parameters α and β is known, and that the unknown parameter is to be estimated by a linear function of a small number of sample quantiles. Let

$n =$ the sample size,
$k =$ the number of sample quantiles on which to base the estimate,
$\lambda_0, \lambda_1, \cdots, \lambda_{k+1} =$ real numbers such that $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_{k+1} = 1$,
$n_i =$ the greatest integer not greater than $n\lambda_i + 1 (i = 1, 2, \cdots, k)$,
$x_i =$ the n_ith order statistic of the sample $(i = 1, 2, \cdots, k)$,
$u_i = G^{-1}(\lambda_i) (i = 1, 2, \cdots, k)$, where the function $G^{-1}(x)$ is the inverse of $G(x)$,
$g_i = g(u_i)$ $(i = 1, 2, \cdots, k)$, $g_0 = g_{k+1} = 0$,
$\Delta_i = \lambda_i - \lambda_{i-1}$ $(i = 1, 2, \cdots, k + 1)$,
$\Delta_{1i} = g_i - g_{i-1}$ $(i = 1, 2, \cdots, k + 1)$,
$\Delta_{2i} = u_i g_i - u_{i-1} g_{i-1}$ $(i = 1, 2, \cdots, k + 1)$, where $u_0 g_0 = u_{k+1} g_{k+1} = 0$,
$K_{rs} = \sum_{i=1}^{k+1} \Delta_{ri} \Delta_{si} / \Delta_i$ $(r, s = 1, 2)$.

Assuming that $\lambda_1, \lambda_2, \cdots, \lambda_k$ are fixed quantities and that $g(x)$ is positive and differentiable in neighborhoods of $x = u_i$ $(i = 1, 2, \cdots, k)$, it is well known (Ogawa (1951, 1962)) that the asymptotically best linear unbiased estimate (ABLUE) of α is

$$\hat{\alpha} = \sum_{i=1}^{k} a_i x_i - a_0 \beta \, , \tag{2.1}$$

where $a_0 = K_{12}/K_{11}$ and

$$a_i = \frac{g_i}{K_{11}} \left(\frac{\Delta_{1i}}{\Delta_i} - \frac{\Delta_{1,i+1}}{\Delta_{i+1}} \right) \qquad (i = 1, 2, \cdots, k) \, , \tag{2.2}$$

and that the variance of the asymptotic distribution of $\hat{\alpha}$ is

$$\mathrm{AV}(\hat{\alpha}) = \frac{\beta^2}{nK_{11}} \, . \tag{2.3}$$

Similarly, the ABLUE of β is

$$\hat{\beta} = \sum_{i=1}^{k} b_i x_i - b_0 \alpha \, , \tag{2.4}$$

where $b_0 = K_{12}/K_{22}$ and

$$b_i = \frac{g_i}{K_{22}} \left(\frac{\Delta_{2i}}{\Delta_i} - \frac{\Delta_{2,i+1}}{\Delta_{i+1}} \right) \qquad (i = 1, 2, \cdots, k) \, , \tag{2.5}$$

and

$$\text{AV}(\hat{\beta}) = \frac{\beta^2}{nK_{22}} . \tag{2.6}$$

3. Optimum spacing for the extreme value distributions. Let us now consider the estimation problem of Section 2 for the extreme value distributions defined in Section 1. Consideration will be limited to the cases of estimating (i) the location parameter α for $F_1(x)$ or $F_4(x)$ when β is known, (ii) the scale parameter β for $F_\nu(x)$ ($\nu = 2, 3, 5, 6$) when α and γ are known. We shall be particularly concerned with the problem of finding the optimum spacing, i.e. the set of values of $\lambda_1, \lambda_2, \cdots, \lambda_k$ which minimizes the asymptotic variance of the ABLUE.

When the quantities g_i, u_i, a_i, b_i, K_{11} and K_{22} are used for the distribution $F_\nu(x)$ ($\nu = 1, 2, \cdots, 6$) we shall add ν as an upper index. The reader should be careful not to interpret these superscripts as exponents. We then have

$$g_i^1 = -\lambda_i \log \lambda_i \qquad (i = 1, 2, \cdots, k) , \tag{3.1}$$

$$u_i^2 g_i^2 = -\gamma\lambda_i \log \lambda_i \qquad (i = 1, 2, \cdots, k) , \tag{3.2}$$

$$u_i^3 g_i^3 = \gamma\lambda_i \log \lambda_i \qquad (i = 1, 2, \cdots, k) , \tag{3.3}$$

$$g_i^4 = -(1 - \lambda_i) \log (1 - \lambda_i) \qquad (i = 1, 2, \cdots, k) , \tag{3.4}$$

$$u_i^5 g_i^b = \gamma(1 - \lambda_i) \log (1 - \lambda_i) \qquad (i = 1, 2, \cdots, k) , \tag{3.5}$$

$$u_i^6 g_i^6 = -\gamma(1 - \lambda_i) \log (1 - \lambda_i) \qquad (i = 1, 2, \cdots, k) , \tag{3.6}$$

and the relations

$$\gamma^2 K_{11}^1 = K_{22}^2 = K_{22}^3 , \tag{3.7}$$

$$\gamma^2 K_{11}^4 = K_{22}^5 = K_{22}^6 , \tag{3.8}$$

$$K_{11}^1(\lambda_1, \cdots, \lambda_k) = K_{11}^4(1 - \lambda_k, \cdots, 1 - \lambda_1) . \tag{3.9}$$

The last relation is intended to mean that if we replace λ_i by $1 - \lambda_{k-i+1}$($i = 1, 2, \cdots, k$) in the function $K_{11}^4(\lambda_1, \cdots, \lambda_k)$, we obtain the same function as K_{11}^1.

The optimum spacing $\lambda_1, \lambda_2, \cdots, \lambda_k$ is therefore the same for maximizing K_{11}^1, K_{22}^2 or K_{22}^3, and is equal to $1 - \lambda_k^E, 1 - \lambda_{k-1}^E, \cdots, 1 - \lambda_1^E$, where $\lambda_1^E, \lambda_2^E, \cdots, \lambda_k^E$ is the optimum spacing for the ABLUE of β for the exponential distribution. The optimum spacing for maximizing K_{11}^4, K_{22}^5 or K_{22}^6 is $\lambda_1^E, \lambda_2^E, \cdots, \lambda_k^E$. If K_{22}^E denotes the maximum value of K_{22} for the exponential distribution, the maximum value of K_{11}^1 and K_{11}^4 is also K_{22}^E and the maximum value of K_{22}^ν($\nu = 2, 3, 5, 6$) is $\gamma^2 K_{22}^E$. It might be interesting to note that the optimum spacing for maximizing K_{22}^ν ($\nu = 2, 3, 5, 6$) is independent of the shape parameter γ.

The problem of optimum spacing of sample quantiles and the related problem of asymptotically optimum grouping has been considered for the exponential distribution by Hughes (1949), Kulldorff (1958, 1961, 1963), Ogawa (1960), Sarhan, Greenberg and Ogawa (1963), and Saleh and Ali (1966). The optimum

u_i-values can be written as

$$u_i^E = \sum_{j=k-i+1}^{k} \delta_j \qquad (i = 1, 2, \cdots, k), \tag{3.10}$$

where

$$\delta_1 = h^{-1}(2),\ \delta_i = h^{-1}[2 + \delta_{i-1} - h(\delta_{i-1})] \qquad (i = 2, 3, \cdots, k) \tag{3.11}$$

and the function $h^{-1}(x)$ is the inverse of $h(x) = x/(1 - e^{-x})$. The corresponding values of the coefficients $b_i(i = 0, 1, \cdots, k)$ and K_{22} can be written as

$$b_0^E = \left(2 - \frac{\delta_k}{e^{\delta_k} - 1}\right)^{-1}, \tag{3.12}$$

$$b_i^E = 2e^{-u_i^E}[h(\delta_{k-i+1}) - 1]/K_{22}^E \qquad (i = 1, 2, \cdots, k) \tag{3.13}$$

and

$$K_{22}^E = 1 - \left(1 - \frac{\delta_k}{e^{\delta_k} - 1}\right)^2 = 1 - (1 + \sum_{i=1}^{k} (-1)^i \delta_i)^2 . \tag{3.14}$$

These optimum values of u_i^E, $\lambda_i^E = 1 - \exp(-u_i^E)$, $b_i^E(i = 1, 2, \cdots, k)$ and K_{22}^E have been tabulated in some of the papers referred to above for $k = 1, 2, \cdots, 15$.

The simple relationships between the optimum values of $\lambda_i(i = 1, 2, \cdots, k)$ and K_{11} or K_{22} for the six extreme value distributions and the exponential distribution were established above. It is easily seen that in the optimum situations we also have the relations

$$u_{k-i+1}^1 = -u_i^4 = -\log u_i^E \qquad (i = 1, 2, \cdots, k), \tag{3.15}$$

$$u_{k-i+1}^2 = -u_i^5 = (u_i^E)^{-1/\gamma} \qquad (i = 1, 2, \cdots, k), \tag{3.16}$$

$$u_{k-i+1}^3 = -u_i^6 = -(u_i^E)^{1/\gamma} \qquad (i = 1, 2, \cdots, k), \tag{3.17}$$

$$a_{k-i+1}^1 = a_i^4 = b_i^E u_i^E \qquad (i = 1, 2, \cdots, k), \tag{3.18}$$

$$b_{k-i+1}^2 = -b_i^5 = b_i^E (u_i^E)^{1+1/\gamma} \qquad (i = 1, 2, \cdots, k), \tag{3.19}$$

$$b_{k-i+1}^3 = -b_i^6 = -b_i^E (u_i^E)^{1-1/\gamma} \qquad (i = 1, 2, \cdots, k), \tag{3.20}$$

$$a_0^1 = -a_0^4 = \sum_{i=1}^{k} a_i^1 u_i^1 = -\sum_{i=1}^{k} b_i^E u_i^E \log u_i^E , \tag{3.21}$$

$$b_0^2 = -b_0^5 = \sum_{i=1}^{k} b_i^2 = \sum_{i=1}^{k} b_i^E (u_i^E)^{1+1/\gamma} , \tag{3.22}$$

$$b_0^3 = -b_0^6 = \sum_{i=1}^{k} b_i^3 = -\sum_{i=1}^{k} b_i^E (u_i^E)^{1-1/\gamma} . \tag{3.23}$$

4. Censored samples. If the sample is censored at one or both ends of the distribution this introduces a restriction on the spacing of sample quantiles,

$$A \leqq \lambda_1 < \cdots < \lambda_k \leqq B , \tag{4.1}$$

where $A \geqq 0$ and $B \leqq 1$. If the optimum spacing for a complete sample is such that (4.1) is not fulfilled, we have a new problem of finding the optimum spacing for the censored sample. This problem has been dealt with by Saleh (1966) for the exponential distribution.

Let $\lambda_1^E, \lambda_2^E, \cdots, \lambda_k^E$ be the optimum spacing for maximizing K_{22} for the exponential distribution subject to (4.1). Relations (3.7)—(3.9) then imply that the optimum spacing for maximizing K_{11}^4, K_{22}^5 or K_{22}^6 subject to (4.1) is also $\lambda_1^E, \lambda_2^E, \cdots, \lambda_k^E$, and that the optimum spacing for maximizing K_{11}^1, K_{22}^2 or K_{22}^3 subject to the restrictions

$$(4.2) \qquad 1 - B \leqq \lambda_1 < \cdots < \lambda_k \leqq 1 - A$$

is $1 - \lambda_k^E, 1 - \lambda_{k-1}^E, \cdots, 1 - \lambda_1^E$. The relations (3.15)—(3.23) also hold in these situations.

It should be noticed that the results obtained by Saleh (1966) for the ABLUE of β based on optimally spaced quantiles from a left censored sample from an exponential distribution, are incorrect. When using the optimum spacing

$$(4.3) \qquad \lambda_i = A + (1 - A)\lambda_{i-1}' \qquad (i = 1, 2, \cdots, k),$$

where $\lambda_0' = 0$ and $\lambda_1', \lambda_2', \cdots, \lambda_{k-1}'$ is the optimum spacing of $k - 1$ sample quantiles from a complete sample, we have

$$(4.4) \qquad K_{22} = (1 - A)\left[K_{22}' + \frac{1}{A}\log^2(1 - A)\right],$$

where K_{22}' is the maximum value of the function K_{22} for $k - 1$ sample quantiles, and the coefficients of the ABLUE of β are

$$(4.5) \qquad b_0 = -\log(1 - A)/[\log^2(1 - A) + AK_{22}'],$$

$$(4.6) \qquad b_1 = b_0 - b_0'\bigg/\left(1 + \frac{\log^2(1 - A)}{AK_{22}'}\right),$$

$$(4.7) \qquad b_i = b_{i-1}'\bigg/\left(1 + \frac{\log^2(1 - A)}{AK_{22}'}\right) \qquad (i = 2, 3, \cdots, k),$$

where $b_i'(i = 0, 1, \cdots, k - 1)$ are the coefficients of the ABLUE of β based on $k - 1$ sample quantiles with optimum spacing. The relative asymptotic efficiency of this ABLUE of β compared to the BLUE of β based on all the observations in the censored sample is

$$(4.8) \qquad \mathrm{RAE}(\hat{\beta}) = \frac{AK_{22}' + \log^2(1 - A)}{A + \log^2(1 - A)}.$$

5. Some comments on related work. The problem of optimum spacing of sample quantiles for the ABLUE of α in $F_1(x)$ has been considered by Hassanein (1968) for a complete sample and by Chan and Kabir (1969) for a censored sample.

Hassanein (1968) tabulated the optimum spacing $\lambda_i^1(i = 1, 2, \cdots, k)$ and the corresponding values of $u_i^1(i = 1, 2, \cdots, k)$, $a_i^1(i = 0, 1, \cdots, k)$, K_{11}^1 and $1/K_{11}^1$ for $k = 1, 2, \cdots, 15$. He was obviously not aware of the fact that his Tables 1—3 are easily obtainable from previous tables for the exponential distribution by using the relations $\lambda_i^1 = 1 - \lambda_{k-i+1}^E(i = 1, 2, \cdots, k)$, $K_{11}^1 = K_{22}^E$, (3.15), (3.18) and (3.21).

Chan and Kabir (1969) tabulated the optimum spacing $\lambda_i^1(i = 1, 2, \cdots, k)$

and the corresponding values of $a_i^1 (i = 0, 1, \cdots, k)$ and K_{11}^1 for $k = 2, 3, 4, 5$ and for some selected values of A and B. Their Table 3 is closely related to Saleh's (1966) Table I through $\lambda_i^1 = 1 - \lambda_{k-i+1}^E (i = 1, 2, \cdots, k)$, $K_{11}^1 = K_{22}^E$, (3.18) and (3.21). It follows from (4.3) that the optimum spacing in their Table 2 for a right censored sample is $\lambda_i = B\lambda_i' (i = 1, 2, \cdots, k)$, where $\lambda_1', \lambda_2', \cdots, \lambda_{k-1}'$ is the optimum spacing of $k - 1$ sample quantiles from a complete sample as tabulated by Hassanein (1968) and $\lambda_k' = 1$.

REFERENCES

CHAN, LAI K. and KABIR, A. B. M. LUTFUL (1969). Optimum quantiles for the linear estimation of the parameters of the extreme value distribution in complete and censored samples. *Naval Res. Logist. Quart.* **16** 381–404.

GNEDENKO, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math.* **44** 423–453.

GUMBEL, E. J. (1958). *Statistics of Extremes.* Columbia University Press.

HASSANEIN, KHATAB M. (1968). Analysis of extreme-value data by sample quantiles for very large samples. *J. Amer. Statist. Assoc.* **63** 877–888.

HUGHES, HARRY M. (1949). Estimation of the variance of the bivariate normal distribution. *Univ. Calif. Publ. Statist.* **1** 37–51.

JOHNSON, NORMAN L. and KOTZ, SAMUEL (1970). *Continuous Univariate Distributions*, **1**. Houghton Mifflin, Boston.

KULLDORFF, GUNNAR (1958). Maximum likelihood estimation for the exponential distribution when the sample is grouped. Department of Statistics, Univ. of Lund, Sweden.

KULLDORFF, GUNNAR (1961). *Contributions to the Theory of Estimation from Grouped and Partially Grouped Samples.* Almqvist & Wiksell, Stockholm.

KULLDORFF, GUNNAR (1963). On the optimum spacing of sample quantiles from an exponential distribution. Department of Statistics, Univ. of Lund, Sweden.

OGAWA, JUNJIRO (1951). Contributions to the theory of systematic statistics, I. *Osaka J. Math.* **3** 175–213.

OGAWA, JUNJIRO (1960). Determination of optimum spacings for the estimation of the scale parameter of an exponential distribution based on sample quantiles. *Ann. Inst. Statist. Math.* **12** 135–141.

OGAWA, JUNJIRO (1962). Estimation of the location and scale parameters by sample quantiles (for large samples). Chapter 5 in *Contributions to Order Statistics* (Ahmed E. Sarhan and Bernard G. Greenberg, eds.). Wiley, New York.

SALEH, A. K. MD. EHSANES (1966). Estimation of the parameters of the exponential distribution based on optimum order statistics in censored samples. *Ann. Math. Statist.* **37** 1717–1735.

SALEH, A. K. MD. EHSANES and ALI, MIR M. (1966). Asymptotic optimum quantiles for the estimation of the parameters of the negative exponential distribution. *Ann. Math. Statist.* **37** 143–151.

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

INSTITUTE OF MATHEMATICS AND STATISTICS
UNIVERSITY OF UMEÅ
S-901 87 UMEÅ, SWEDEN