1. Introduction

Let X be a variate having the cumulative distribution function (Cdf)

$$P = F(x) = 1 - e^{-((x - \mu)/\beta)^m} \tag{1}$$

from which

$$\log(-\ln(1 - P)) = m.\log(x - \mu) - m.\log \beta \tag{2}$$

Already from the first applications of this distribution to actual sets of observations (1), it has been a common praxis to check the assumption that the observed data x_i of a sample of size N belong to the population, defined by eq. (1), by plotting the values $\log(x_i - \mu)$ against the values $\log(-\ln(1 - P_i))$.

Since P, by definition, denotes the probability that a value drawn at random from the population is equal to or less than x, the "plotting position" P_i tends to i/N as N tends to infinity.

For finite sample sizes a modified expression of P_i has to be introduced. Three such formulas have been used, viz.

$$P_i = (i - 0.5)/N \quad P_i = i/(N + 1) \quad P_i = (i - 0.3)/(N + 0.4) \tag{3}$$

The third formula, introduced by Benard & Bos-Levenbach (2), has been proposed by L.G. Johnson (3) as plotting position. It yields a remarkably good approximation of the percentage points of the medians of the order statistics, as will be demonstrated in the following.

If now, after a proper choice of μ, if unknown, the data points fall, with due consideration to unavoidable sampling errors, on a straight line, then it is accepted that the observations belong to a population defined by eq. (1).

In order to estimate the parameters from the graph, it has been usual to fit a straight line by eye to the data, or, to eliminate subjectiveness, by use of the least-squares method. Both methods will, in principle, yield the same result, because in the first alternative the line is intuitively drawn so as to provide the smallest deviations from the fitted line of the data points, which are assumed to have the same "weights".

The slope of the fitted line is then taken as an estimate of the parameter m, and the intersection with the P-axis as an estimate of m.log. That value of μ, which yields the smallest deviations from a straight line, is taken as an estimate of μ.

This method of analyzing given test data is very convenient and recommendable as a first orientation, because it is adapted for disclosing whether the observations emanate from one single population or not.

However, as a method of estimating the distribution parameters it suffers from serious imperfections due to the following causes:

a) The estimates will be biased, because none of the plotting positions in eq. (3) is generally acceptable, the correct positions being the expected values of the order statistics y_i.

b) The estimates have a very poor efficiency for two reasons: The data points are assumed to have the same weights, measured by the reciprocal of their variances, which is very far from the fact. For instance, for N = 20 the variance $V(y_1) = 0.31025$ while $V(y_{17}) = 0.01431$ and for N = 100 we have $V(y_1) = 0.31025$ while $V(y_{80}) = 0{,}00290$. Neglecting the difference in weights may in some cases reduce the efficiency by 90 %, that is, down to 10 %. The second reason of efficiency is explained by the neglection of the mutual dependencies of the order statistics, that is, the assumption that all $Cov(y_i, y_j) = 0$.

Furthermore, the confidence limits of the order statistics have, up to now, been unknown, which makes it difficult to decide whether observed deviations are acceptable or not.

The scope of the present report is to eliminate the imperfections mentioned above by deducing formulas for distribution functions, expected values, variances and covariances of the order statistics y_i, as demonstrated in Chapter 2, and, by use of these formulas, to arrive at linear, unbiased, minimum-variance estimates of the parameters, as demonstrated in Chapter 3.

2. Properties

2.1 General properties

The standardized variate Z is defined by

$$z = (x - \mu)/\beta \tag{4}$$

Introducing z into eq. (1), the Cdf of the variate Z becomes

$$F(z) = 1 - e^{-z^m} \tag{5}$$

This function involves one parameter only, the shape parameter m.

Introducing a new variate Y, defined by

$$y = \log(z^m) \tag{6}$$

it follows from (2), (4) and (6) that

$$y = m.\log(x - \mu) - m.\log \beta \tag{7}$$

and from (1), (4) and (6) that the Cdf of Y becomes

$$F(y) = 1 - e^{-10^y} \tag{8}$$

and its density function (Df)

$$f(y) = 10^y.e^{-10^y}/\log e \tag{10}$$

This distribution is parameter-free, which is of great practical importance, because the order statistics y_i depend on i and N only, whereas the order statistics z_i depend also on the shape parameter m. The table work required for z_i consequently is ten times as large as that required for y_i.

Formulas for the distribution functions, variances and covariances of the order statistic y_i have been developed and they are presented in the following sections.

2.2 Distribution functions, percentiles and confidence limits

The Cdf of the order statistic y_i is

$$F(y_i) = \frac{N!}{(i-1)!(N-i)!} \int_{-\infty}^{y_i} F^{i-1}(y)(1-F(y))^{N-i} \cdot f(y)dy \tag{10}$$

where $F(y)$ is the Cdf of the variate Y.

Putting

$$P = F(y) \tag{11}$$

and

$$N!/(i-1)!(N-i)! = i \cdot C_i^N \tag{12}$$

eq. (10) may be transformed into

$$F(P_i) = i \cdot C_i^N \int_0^{P_i} P^{i-1}(1-P)^{N-i} dP \tag{13}$$

The percentage point $P_{i,p}$ corresponding to the percentile $y_{i,p}$ is equal to the root of the equation

$$i \cdot C_i^N \int_0^{P_{i,p}} P^{i-1}(1-P)^{N-i} dP = p \tag{14}$$

It is interesting to note that $P_{i,p}$ does not require any specification of the function $F(y)$ and the formula (14) is valid for any continouous distribution as, for instance, eq. (1) and (5), that is, for the order statistics x_i and z_i.

Substituting

$$Q = 1 - P \qquad dP = - dQ \tag{15}$$

into eq. (14) we have the percentage point $Q_{j,q}$ defined by

$$- j \cdot C_j^N \int^{Q_{j,q}} (1-Q)^{j-1} \cdot Q^{N-j} dQ = q \tag{16}$$

If $j = N + 1 - i$, then, as easily verified, $j \cdot C_j^N = i \cdot C_i^N$ and from eq. (14) and (16)

$$i \cdot C_i^N \left(\int_{Q_{j,q}} Q^{i-1}(1-Q)^{N-i} dQ + \int^{P_{i,p}} P^{i-1}(1-P)^{N-i} dP \right) = p+q \tag{17}$$

Hence, if

$$Q_{j,q} = 1 - P_{j,q} = P_{i,p} \tag{18}$$

then

$$i \cdot C_i^N \int_0 P^{i-1}(1-P)^{N-1} dP = 1 = p+q$$

or

$$q = 1 - p \tag{19}$$

and

$$P_{N+1-i,1-p} = 1 - P_{i,p} \tag{20}$$

This formula cuts the calculations in half. The integral in eq. (13) is identical with the Beta-function, which has been tabulated by K. Pearson. (4)

Values of $P_{i,p}$ have been computed and are presented in Table 1 for

$$p = .5, 25, 50, 75, 95 \quad \text{and} \quad N = 1(1)25$$

The percentage points $\check{P}_i$ are very closely approximated by the Benard formula in eq. (3) as demonstrated in Table 2.

From the known values $P_{i,p}$ the corresponding percentiles can be computed for any variate T with given Cdf F(t) by the formula

$$t_{i,p} = F^{-1}(P_{i,p}) \tag{21}$$

where F^{-1} is the inverse function of F.

Applying this formula to the distribution functions (5) and (8) we have

$$z_{i,p} = (-\ln(1 - P_{i,p}))^{\alpha} \tag{22}$$

and

$$y_{i,p} = \log(-\ln(1 - P_{i,p})) \tag{23}$$

Some values of $y_{i,p}$ are presented in Table 3, from which also the limits for the 50 % and 90 % confidence levels of $y_{i,p}$ can be read.

For the <u>special case i = 1</u> equation (13) takes the simple form

$$F(P_1) = N \int_0^P (1-P)^{N-1} dP = 1 - (1-P_1)^N \tag{24}$$

from which

$$P_{1,p} = 1 - (1 - p)^{1/N} \tag{25}$$

and

$$y_{i,p} = \log \ln(1/(1 - p)) - \log N \tag{26}$$

From eq. (26) it can be concluded that the confidence intervals of y_1, for any confidence level, are independent of the sample size N, and that the location of the limits are fixed in relation to the median $\breve{y}_1$, and, as will be proved in the next section, also in relation to the expected value Ey_1.

For p = 50 % it follows from eq. (25) that

$$\breve{P}_1 = 1 - 0.5^{1/N} \tag{27}$$

Hence

$$\breve{P}_1 \longrightarrow 0.69315/N \quad \text{as} \quad N \longrightarrow \infty \tag{28}$$

which is very close to the Benard value, which becomes

$$\breve{P}_1 \longrightarrow 0.7/N \quad \text{as} \quad N \longrightarrow \infty$$

Further

$$\breve{y}_1 = \log(0.69315/N) = - 0.159175 - \log N \tag{29}$$

while the Benard formula gives

$$\breve{y}_1 = - 0.154902 - \log N$$

2.3 Expected values and plotting positions

The expected values of y_i, denoted by $E(y_i)$, are equal to the first moment of y_i. Considering the Cdf (10) we thus have

$$E(y_i) = i.C_i^N.\int_{-\infty}^{\infty} y.F^{i-1}(y)(1-F(y))^{N-i}f(y)dy \tag{30}$$

or by (8) and (9)

$$E(y_i) = i.C_i^N.\int_{-\infty}^{\infty} y.(1-e^{-10^y})^{i-1}.e^{-(N-i)10^y}.10^y.e^{-10^y}dy/\log e \tag{31}$$

Introducing

$$t = 10^y \; ; \; y = \log t \; ; \; dy = \log e.dt/t \tag{32}$$

the expected value becomes

$$E(y_i) = i.C_i^N.\log e.\int_0^{\infty} \ln t(e^t-1)^{i-1}.e^{-Nt}dt \tag{33}$$

Considering that

$$(e^t-1)^{i-1} = \sum_{r=0}^{i-1}(-1)^r.C_r^{i-1}.e^{-(N+1-i+r)t}dt \tag{34}$$

we have

$$E(y_i) = i.C_i^N.\log e.\int_0^{\infty}\ln t.\sum_0^{i-1}(-1)^r.C_r^{i-1}.e^{-(N+1-i+r)t}dt \tag{35}$$

The integral

$$I = \int_0^{\infty}\ln t.e^{-kt}dt = \int_0^{\infty}(\ln t - \ln k)e^{-t}dt/k$$

and from the Appendix

$$I = (Y(1,0) - \ln k)/k \tag{36}$$

Since

$$i \cdot C_i^N \cdot \sum_{0}^{i-1} (-1)^r \cdot C_r^{i-1}/(N+1-i+r) = 1 \qquad (37)$$

and denoting

$$Y(1,0).\log e = C_1 = -\ 0.250681578 \qquad (38)$$

it follows that

$$E(y_i) = C_1 - i \cdot C_i^N \cdot \sum_{r=0}^{i-1} (-1)^r \cdot C_r^{i-1} . \log(N+1-i+r)/(N+1-i+r) \qquad (39)$$

In particular we have

$$\left.\begin{aligned} E(y_1) &= C_1 - \log N \\ E(y_2) &= C_1 - N(N-1)(\log(N-1)/(N-1) - \log N/N) \end{aligned}\right\} \qquad (40)$$

Observing that the sum of the unordered and the ordered values in a sample must be equal or

$$\sum_{i=1}^{N} y_{i,N} = \sum_{i=1}^{N} y_{i,1} \qquad (41)$$

also their expected values are equal, that is,

$$\sum_{i=1}^{N} E(y_{i,N}) = N \cdot E(y_{1,1}) = N \cdot C_1 \qquad (42)$$

This formula may be used as a control of the computations.

Comparing eq. (29) and (40) it is readily seen that, for any sample size

$$\breve{y}_1 - E(y_1) = 0.091507 \qquad (43)$$

For moderate sample sizes the values of $E(y_i)$ can be computed by use of the formula (39). For large sample sizes, the binomial coefficients will become unpracticably large. In that case the expected values can be computed by applying the Simpson rule to the integral in eq. (31) or (33). Using these two methods the expected values have been computed for

$$N = 1(1)25(5)30(10)50(25)125 \text{ and selected values of}$$

$$N = 1.000$$

as presented in Table 4.

As proved in Chapter 3, the expected values $E(y_i)$ will provide the proper plotting positions in so far as the estimates will be unbiased.

In order to simplify the plotting procedure, when probability papers are available, the percentage points $\bar{P}_i$, corresponding to $E(y_i)$ have been computed by use of the formula

$$E(y_i) = \log(-\ln(1 - \bar{P}_i)) \tag{44}$$

For <u>the special case i = 1</u> we have from eq. (41) and (44)

$$\bar{P}_1 = 1 - e^{-0.561\ 4595/N} \tag{45}$$

from which it follows that

$$\bar{P}_1 \longrightarrow 0.561\ 4595/N \quad \text{as} \quad N \longrightarrow \infty \tag{46}$$

For sample sizes other than those tabulated, a formula giving the values of P_i with sufficient accuracy is very desirable. It should be fitted for use in computers.

After having compared the values of $E(y_i)$ corresponding to the values of $\bar{P}_i$ given in eq. (3), it was found that the first formula in most cases came closer to the exact values of $E(y_i)$ than the two other formulas. Starting from this observation, a correction term c, defined by

$$\overline{P}_i = (i - c)/N \quad \text{or} \quad c = i - N.\overline{P}_i \tag{47}$$

has been computed for several combinations of i and N. It was found that the function

$$c = f((i-1)/(N-1)) \tag{48}$$

tends to a fixed form for increasing N, as demonstrated in Table 5 for N = 21(10)121 and values of i corresponding to (i-1)/(N-1) = 0(0.1)1.0.

For any sample size, the function (48) can be approximated by a straight line. For sample sizes from N = 50 and above, a very close approximation is attained by the formula

$$c = 0.49 + 0.15\, i/N \tag{49}$$

The accuracy of this formula with regard to the deviations from the exact values $E(y_i)$ is demonstrated in Table 6, in which also the values corresponding to c = 0 and c = 0.5 are presented. For sample sizes larger than N = 100 the deviations from the true plotting positions will be smaller than those given in Table 6.

2.4 Second moments, variances and weights of the observations

In analogy with eq. (33) the second moment becomes

$$E(y_i^2) = i.C_i^N.\log^2 e \int_0^\infty \ln^2 t.(e^t - 1)^{i-1}.e^{-Nt}dt \tag{50}$$

or, using the expansion eq. (34)

$$E(y_i^2) = i.C_i^N.\log^2 e \int_0^\infty \ln^2 t. \sum_{r=0}^{i-1}(-1)^r.e^{-(N+1-i+r)t}dt \tag{51}$$

After some obvious manipulations the integral

$$J = \int_0^\infty \ln^2 t.e^{-kt}dt = (Y(2,0).\log^2 e - 2 \ln k + \ln^2 k)/k \tag{52}$$

Hence

$$E(y_i^2) = C_2 - i.C_i^N.\sum_0^{i-1}(-1)^r(2\ C_1\ \log k - \log^2 k)/k \tag{53}$$

where

$$\begin{aligned} k &= N+1-i+r \\ C_2 &= Y(2,0).\log e = 0.373\ 095\ 059 \\ 2\ C_1 &= 2Y(1,0).\log e = -0.501\ 363\ 156 \end{aligned} \tag{54}$$

Since the values of $E(y_i^2)$ have been computed, the variance

$$V(y_i) = E(y_i^2) - E^2(y_i) \tag{55}$$

In particular we have <u>for i = 1</u>

$$V(y_1) = C_2 - C_1^2 = 0.310\ 254\ 405 \tag{56}$$

for any sample size, and <u>for $i = 2$</u>

$$V(y_2) = C_2 - C_1^2 - N(N-1)(\log N - \log(N-1))^2$$

and for $N = 100$

$$V(y_2) = 0.121\ 6443$$

and for $N \rightarrow \infty$

$$V(y_2) = 0.121\ 6427 \qquad (57)$$

Values of the variances $V(y_i)$ for $N = 1(1)25;30(10)50(25)125$ and for selected order numbers i of the sample $N = 1{,}000$ have been computed and are presented in Table 4. Their reciprocals may be used as the weights w_i of the observations as demonstrated in Section 3.1.2.

For sample sizes equal to or larger than $N = 50$ the weights can be approximated according to the following formulas:

$$w_i = i \qquad \text{for} \qquad i \leq 0.67N$$

$$w_i = 0.67N \qquad \text{for} \qquad 0.67N \leq i \leq 0.90N$$

$$w_i = 0.67N - 5(i - 0.90N) \qquad \text{for} \qquad 0.90N \leq i$$

2.5 Product moments and covariances

In analogy with eq. (33) the product moment becomes

$$E(y_i,y_j) = \frac{N!}{(i-1)!(j-i-1)!(N-j)!} \int_0^\infty \int_0^{t_j} \ln t_i . \ln t_j . (1-e^{-t_i})^{i-1} . (e^{-t_i}-e^{-t_j})^{j-i-1} . e^{-(N+1-j)t_j} . e^{-t_i} . dt_i dt_j \tag{58}$$

and in particular for $i = 1$; $j = 2$

$$E(y_1,y_2) = N(N-1) \int_0^\infty \int_0^{t_2} \ln t_1 . \ln t_2 . e^{-t_1} . e^{-(N-1)t_2} . dt_1 dt_2 \tag{59}$$

The covariances are then obtained by

$$Cov(y_i,y_j) = E(y_i,y_j) - E(y_i).E(y_j) \tag{60}$$

Covariance matrices have been computed by use of Simpson's rule for $N = 2(1)5(5)20$.

A control of the computations is obtained in the following way: Considering that the sum of the ordered and the unordered sample values are identical, it is obvious that the same rule holds also for their variances. Hence,

$$V(\sum y_i) = N.V(y_1) \tag{61}$$

which implies that

$$2.\sum(Cov(y_i,y_j)) = 0.310\ 254\ N - \sum(V(y_i))(i<j) \tag{62}$$

In particular we have <u>for N = 2</u>

$$Cov(y_1,y_2) = 0.09\ 062$$

3. Estimation of parameters

3.1 Location parameter μ known

3.1.1 Linear, unbiased estimates with minimum variance

Eq. (7) may be written

$$\log(x_i - \mu) = y_i/m + \log \beta \tag{63}$$

Substituting

$$\alpha = 1/m \; ; \; u = \log(x - \mu) \; ; \; b = \log \beta \tag{64}$$

we have

$$u_i = \alpha . y_i + b \tag{65}$$

Let now the estimates of the parameters α, b be linear functions of the observed values u_i (μ being known), then

$$\left.\begin{aligned} \hat{\alpha} &= \Sigma\, a_i . u_i \\ \hat{b} &= \Sigma\, b_i . u_i \end{aligned}\right\} \tag{66}$$

These estimates are unbiased on the condition that

$$\left.\begin{aligned} E\hat{\alpha} &= \Sigma\, a_i . Eu_i = \alpha\Sigma\, a_i . Ey_i + b . \Sigma\, a_i = \alpha \\ E\hat{b} &= \Sigma\, b_i . Eu_i = \alpha\Sigma\, b_i . Ey_i + b . \Sigma\, b_i = b \end{aligned}\right\} \tag{67}$$

which requires that

$$\left.\begin{aligned} &\Sigma(a_i . Ey_i) = 1 \qquad \Sigma\, a_i = 0 \\ &\Sigma(b_i . Ey_i) = 0 \qquad \Sigma\, b_i = 1 \end{aligned}\right\} \tag{68}$$

The variances of the estimates (66) are, considering that

$$\mathrm{Cov}(u_i, u_j) = \alpha^2 . \mathrm{Cov}(y_i, y_j) \tag{69}$$

equal to

$$\left.\begin{aligned} V(\hat{a}) &= \alpha^2 \Sigma\ a_i a_j \cdot \sigma_{ij} \\ V(\hat{b}) &= \alpha^2 \Sigma\ b_i b_j \cdot \sigma_{ij} \end{aligned}\right\} \quad (70)$$

where

$$\sigma_{ij} = \mathrm{Cov}(y_i, y_j) \quad \text{and } \sigma_{ii} = V(y_i) \quad (71)$$

Derivating eq. (70) with respect to a_i and b_i and taking eq. (68) as side conditions, it follows that the coefficients a_i, b_i are the roots of the following two systems of equations

$$\left.\begin{array}{l} a_1 + a_2 + \ldots + 0 + 0 = 0 \\ a_1 \cdot E(y_1) + a_2 \cdot E(y_2) + \ldots + 0 + 0 = 1 \\ a_1 \cdot \sigma_{11} + a_2 \cdot \sigma_{12} + \quad + k_1 + k_2 \cdot E(y_1) = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_1 \cdot \sigma_{n1} + a_2 \cdot \sigma_{n2} + \ldots + k_1 + k_2 \cdot E(y_n) = 0 \end{array}\right\} \quad (72)$$

and

$$\left.\begin{array}{l} b_1 + b_2 + \ldots + 0 + 0 = 1 \\ b_1 \cdot E(y_1) + b_2 \cdot E(y_2) + \ldots + 0 + 0 = 0 \\ b_1 \cdot \sigma_{11} + b_2 \cdot \sigma_{12} + \ldots + k_3 + k_4 \cdot E(y_n) = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ b_1 \cdot \sigma_{n1} + b_2 \cdot \sigma_{n2} + \ldots + k_3 + k_4 \cdot E(y_n) = 0 \end{array}\right\} \quad (73)$$

Multiplying the third equation of eq. (72) by a_1, the fourth by a_2, and so on, adding and considering the first two equations, it follows that

$$k_2 = - \Sigma\ a_i a_j\ \sigma_{ij} = - V(\hat{a})/\alpha^2$$

In the same way the complete covariance matrix of the estimates is obtained as

$$V(\hat{\alpha}) = -k_2 . \alpha^2 \;;\; V(\hat{b}) = -k_3 . \alpha^2 \;;\; Cov(\hat{\alpha}, \hat{b}) = -k_1 \alpha^2 = -k_4 \alpha^2 \qquad (74)$$

The coefficients a_i, b_i and k have been computed for some selected sample sizes. The results are presented in Table 7.

3.1.2 Least weighted-squares method

Knowing that the covariance matrices are very tedious to compute, it is desirable to have formulas not requiring them. For large samples, in particular such with grouped data, the efficiency is but slightly reduced by neglecting the dependency of the order statistics, that is, by putting $\sigma_{ij} = 0\ (i \neq j)$. The system of equations (72) then takes the form

$$\Sigma\, a_i = 0\ ;\ \Sigma(a_i . E(y_i)) = 1\ ;\ a_i = -w_i(k_1 + k_2 . E(y_i)) \tag{75}$$

where the weights w_i are defined by

$$w_i = 1/\sigma_i \tag{76}$$

The coefficients are easily computed by use of (75). It is, however, of interest to note that identical results are obtained by estimating the parameters in such a way that the sum of the weighted squares of deviations

$$M = \Sigma\, w_i(u_i - \alpha . E(y_i) - b)^2 / \Sigma\, w_i = \text{minimum} \tag{77}$$

Hence, after derivating M with respect to α and b, the estimates become the roots of the system

$$\left.\begin{aligned} \hat{\alpha} . \Sigma(w_i . E^2(y_i)) + \hat{b} . \Sigma(w_i . E(y_i)) &= \Sigma(w_i u_i . E(y_i)) \\ \hat{\alpha} . \Sigma(w_i . E(y_i)) + \hat{b} . \Sigma\, w_i &= \Sigma(w_i u_i) \end{aligned}\right\} \tag{78}$$

from which

$$\left.\begin{aligned} \hat{\alpha} &= \frac{\Sigma\, w_i . \Sigma(w_i u_i . E(y_i)) - \Sigma(w_i u_i) . \Sigma(w_i . E(y_i))}{\Sigma\, w_i . \Sigma(w_i . E^2(y_i)) - (\Sigma(w_i . E(y_i)))^2} \\ b &= \frac{\Sigma(w_i u_i . \Sigma(w_i . E(y_i)) - \Sigma(w_i . E(y_i)) . \Sigma(w_i u_i . E(y_i))}{\Sigma\, w_i . \Sigma(w_i . E^2(y_i)) - (\Sigma(w_i . E(y_i)))^2} \end{aligned}\right\} \tag{79}$$

The minimum value of M is obtained by introducing the estimated values of the parameters into eq. (77), that is,

$$M_{min} = \Sigma(w_i(u_i - \hat{\alpha}E(y_i) - \hat{b})^2)/\Sigma\, w_i \tag{80}$$

The advantage of this way of deducing the formulas lies in the fact that M_{min} can be used as a criterion whether an actual set of observed data belongs to a Weibull population or not. To this purpose it is, however, necessary to know the expected value of M_{min} and the probability of a given deviation from this value, that is, the Cdf of M_{min}.

This problem can be solved by means of a Monte-Carlo study as follows. In the expression (80) the values of $\hat{E}(y_i)$ and w_i are uniquely determined by the sample size N, whereas $\hat{\alpha}$ and $\hat{b}$ are linear functions of u_i Random values $u_{i,N}$ are obtained from the formula

$$u_{i,N} = \alpha.\log(-\ln(1-P_{i,N})) + b \tag{81}$$

Considering that $P_{i,N}$ is a variate, rectangularly distributed within the interval (0,1), we have, denoting by $r_{i,N}$ real numbers drawn at random from the interval (0,1).

$$u_{i,N} = \alpha.\log(-\ln(1-r_{i,N})) + b \tag{82}$$

The values $u_{i,N}$ thus defined are the elements of a random sample drawn from the population defined by the parameters α, β, μ. Introducing (82) into (66) we have

$$\begin{aligned} \hat{\alpha} &= \alpha\, \Sigma\, a_i.\log(-\ln(1-r_i)) \\ \hat{b} &= \alpha\, \Sigma\, b_i.\log(-\ln(1-r_i)) \end{aligned} \tag{83}$$

and

$$M_{min} = \alpha^2\, \Sigma\, w_i \left[\log(-\ln(1-r_i)) - E(y_i).\Sigma\, a_i.\log(-\ln(1-r_i)) - \Sigma\, b_i.\log(-\ln(r_i))\right]^2$$

3.1.3 Least unweighted-squares method

This approximation is obtained by assuming that the observations have equal weights. The formulas are simplified by putting

$$w_i = 1 \; ; \; \Sigma \, w_i = N \qquad (85)$$

This method is, in principle, identical with the commonly used graphical method. It should, however, be pointed out that the efficiency of the estimates, thus computed, is, in most cases, extremely poor. As an example it may be mentioned that an estimation of the shape parameter from the eleven order statistics nrs 1,10,20,....,90,100 in a sample of size $N = 100$ has an efficiency of 12.5 % only, compared with that of the method not neglecting the differences in weight. This implies that a sample eight times as large is required to obtain the same accuracy.

3.2 Locations on parameter μ unknown

3.2.1 Least-squares method

For a set of properly chosen values of μ the values of M_{min} are computed according to eq. (80) and, by interpolation, that value of μ which corresponds to the least value of M_{min} is accepted as the estimate of μ.

By use of this value the values of u_i are computed according to eq. (64) and finally the estimates $\hat{a}$ and $\hat{b}$ from eq. (66).

3.2.2 Combination of two linear estimators

In the Int. Sci. Rep. Nr 12 of AF 61(052)-522 linear estimators for the parameters β and μ are given on the condition that the parameter α is known. They may be written

$$\hat{\mu} = \Sigma\, c_i \cdot x_i \;;\quad \hat{\beta} = \Sigma\, d_i \cdot x_i \tag{86}$$

Introducing the first estimate into the first of eq. (66) an equation which involves only the shape parameter α is obtained, namely

$$\hat{\alpha} = \Sigma\, a_i \log(x_i - \Sigma\, c_i \cdot x_i) \tag{87}$$

It should be observed that the coefficients c_i depend not only on i and N but also on α, so the solution of the equation is performed by taking some values of α, computing the corresponding values of c_i and then $\hat{\alpha}$ and proceding until $\hat{\alpha}$ equates the selected value of α. From this value $\hat{\mu}$ and $\hat{\beta}$ are computed by use of eq. (86).

<ins>Note</ins>: All the tables in this report have been computed on an IBM Computer by use of programs written in FORTRAN 4 by Laborator Göran W. Weibull, Försvarets Forskningsanstalt, Stockholm 80, from whom copies may be obtained if required.

Appendix: The Integral $Y(k,x) = \int_0^\infty (\ln t)^k . t^x . e^{-t} dt$

This integral can be transformed into combinations of three known and tabulated functions, namely, the Gamma, Digamma and Trigamma functions which are, respectively, defined by

$$x! = \int_0^\infty t^x . e^{-t} dt \tag{A1}$$

$$\Psi(x) = d(\ln x!)/dx = x!'/x! \tag{A2}$$

$$\Psi'(x) = d^2(\ln x!)/dx^2 = x!''/x! - (x!'/x!)^2 \tag{A3}$$

Values of these functions are given to ten decimal places for $x = 0(0.005)1$ in Handbook of Mathematical Functions, Table 6.1

For $x > 1$ the following recurrence formulas may be used.

$$(x+1)! = (x+1).x! \tag{A4}$$

$$\Psi(x+1) = \Psi(x) + 1/(x+1) \tag{A5}$$

$$\Psi'(x+1) = \Psi'(x) - 1/(x+1)^2 \tag{A6}$$

Other notations for $x!$ are $\pi(x)$ and $\Gamma(x+1)$.

For $k = 0$ the integral becomes

$$Y(0,x) = \int_0^\infty t^x . e^{-t} dt \tag{A7}$$

that is, identical with the Gamma function or

$$Y(0,x) = x! \tag{A8}$$

In particular for $x = 0$

$$Y(0,0) = 1 \tag{A9}$$

From (4) it follows that

$$Y(0,x+1) = (x+1).Y(0,x) \qquad \text{(A10)}$$

For k = 1 the integral becomes

$$Y(1,x) = \int_0^{\infty} \ln t . t^x . e^{-t} dt \qquad \text{(A11)}$$

Derivating the members of (1) with respect to x we have

$$x!' = \int_0^{\infty} \ln t . t^x . e^{-t} dt$$

Hence

$$Y(1,x) = x!' \qquad \text{(A12)}$$

and by (2)

$$Y(1,x) = x!.Y(x) \qquad \text{(A13)}$$

In particular for $x = 0$

$$Y(1,0) = Y(0) \qquad \text{(A14)}$$

and the constant

$$C_1 = Y(1,0).\log e \qquad \text{(A15)}$$

By (4), (8) and (12) we obtain the recurrence formula

$$Y(1,x+1) = (x+1).Y(1,x) + Y(0,x) \qquad \text{(A16)}$$

For k = 2 the integral becomes

$$Y(2,x) = \int_0^{\infty} (\ln t)^2 . t^x . e^{-t} dt \qquad \text{(A17)}$$

Derivating the members of (1) twice with respect to x we have

$$x!'' = \int_0^\infty (\ln\ t)^2 . t^x . e^{-t} dt \qquad \text{(A18)}$$

Hence

$$\Upsilon(2,x) = x!'' \qquad \text{(A19)}$$

and by (2) and (3)

$$\Upsilon(2,x) = x! \left[\Upsilon'(x) + (\Upsilon(x))^2\right] \qquad \text{(A20)}$$

In particular for $x = 0$

$$\Upsilon(2,0) = \Upsilon'(0) + (\Upsilon(0))^2 \qquad \text{(A21)}$$

and the constant

$$C_2 = \Upsilon(2,0).\log^2 e = \left[\Upsilon'(0) + (\Upsilon(0))^2\right] .\log\ e \qquad \text{(A22)}$$

The following recurrence formula has been derived

$$\Upsilon(2,x+1) = (x+1).\Upsilon(2,x) + 2\Upsilon(1,x) \qquad \text{(A23)}$$

The following constants are required:

$\log e = 0.434\ 294\ 482$

$\log^2 e = 0.188\ 611\ 697$

$\Upsilon(1,0) = -0.577\ 215\ 665$

$C_1 = -0.250\ 681\ 578$

$\Upsilon(2,0) = 1.978\ 111\ 991$

$C_2 = 0.373\ 095\ 059$

$C_2 - C_1^2 = 0.310\ 253\ 806$

References

1. Weibull, W. "A statistical theory of strength of materials." I.V.A. handl.No.151,1939.

2. Benard, A. & Bos-Levenbach, E.C. The plotting of observations on probability paper (in Dutch), Statistica 7,1953,163-173.

3. Johnson, L.G. "Median ranks of sample values in their population with an application to certain fatigue studies," Ind.Math.Soc.2,1951,1-9.

4. Pearson, K. "Tables of the incomplete Beta-Function, Cambridge,1934.