ESTIMATING THE MEAN AND VARIANCE OF NORMAL POPULATIONS FROM SINGLY TRUNCATED AND DOUBLY TRUNCATED SAMPLES[1]

By A. C. Cohen, Jr.

The University of Georgia

1. Summary. This paper is concerned with the problem of estimating the mean and variance of normal populations from singly and doubly truncated samples having known truncation points. Maximum likelihood estimating equations are derived which, with the aid of standard tables of areas and ordinates of the normal frequency function, can be readily solved by simple iterative processes. Asymptotic variances and covariances of these estimates are obtained from the information matrices. Numerical examples are given which illustrate the practical application of these results. In Sections 3 to 8 inclusive, the following cases of doubly truncated samples are considered: I, number of unmeasured observations unknown; II, number of unmeasured observations in each 'tail' known; and III[2], total number of unmeasured observations known, but not the number in each 'tail'. In Section 9, singly truncated samples are treated as special cases of I and II above.

2. Introduction. In practice, truncated samples arise with various types of experimental data in which recorded measurements are available over only a partial range of the variable. Such samples are usually classified according to the form of the population (complete) distribution; according to whether the truncation points are known or unknown; and according to whether the number of unmeasured (missing) observations is known or unknown. In this paper, the further classification of singly truncated or doubly truncated is made, accordingly as one or both 'tails' of the sample have been removed. Pearson and Lee [1, 2], Fisher [3], Hald [4][3], and this writer [5] studied singly truncated normal samples with a known truncation point when the number of unmeasured observations is unknown. Stevens [6], Cochran [7], and Hald [4] studied similar samples with a known number of unmeasured observations. Stevens [6] also considered doubly truncated normal samples with known truncation points when the number of unmeasured observations in each 'tail' is known. In each of these papers, equations were derived with which maximum likelihood estimates of the population mean and variance can be computed from samples of the type considered. With the exception of [5], which uses standard tables of the normal frequency

[1] Based on papers presented before the American Mathematical Society, Durham, North Carolina, April 2, 1949, and before a joint meeting of the Institute of Mathematical Statistics and the Biometric Society, Chapel Hill, North Carolina, March 18, 1950.

[2] The problem involved in this case was recently called to the writer's attention by Churchill Eisenhart.

[3] Reference [4] appeared while this paper was awaiting publication. Minor revisions have been made in view of Hald's results.

function, practical application of the various estimating equations involves use of special tables which may frequently be unavailable.

3. Case I. Number of unmeasured observations unknown. Let x_0' designate the left truncation point, $x_0' + R$ the right truncation point, and hence R the sample range. Let n_0 be the number of measured observations with values equal to or between the truncation points. In this case, the number of unmeasured observations is assumed to be unknown. We translate the origin to the left terminus by the change of variable $x = x' - x_0'$, and designate the left and right truncation points in standard units of the population (complete distribution) as ξ' and ξ'', respectively. We can write the probability density function for this case as

$$f(x) = \frac{1}{(I_0' - I_0'')\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\xi' + x/\sigma)^2}, \qquad 0 \le x \le R, \tag{1}$$

where

$$I_0' = \frac{1}{\sqrt{2\pi}} \int_{\xi'}^{\infty} e^{-t^2/2}\, dt, \qquad I_0'' = \frac{1}{\sqrt{2\pi}} \int_{\xi''}^{\infty} e^{-t^2/2}\, dt, \tag{2}$$

and

$$\mu = x_0' - \sigma\xi'. \tag{3}$$

Thus $(I_0' - I_0'')$ is the area under the normal curve between ordinates erected at ξ' and ξ'' respectively. Moreover $(I_0' - I_0'') = P(x_0' \le x' \le x_0' + R)$. The likelihood function for such a sample is

$$P(x_1, x_2, \cdots, x_{n_0}) = \left(\frac{1}{(I_0' - I_0'')\sigma\sqrt{2\pi}}\right)^{n_0} e^{-\frac{1}{2}\sum_{1}^{n_0}(\xi' + x_i/\sigma)^2}. \tag{4}$$

Since R is the truncated range, and since ξ' and ξ'' are in standard units, we have

$$\xi'' = \xi' + R/\sigma. \tag{5}$$

It should be understood that ξ' is considered throughout this paper, as the independent parameter of location. The mean, μ, cf. (3), is a linear function of ξ'.

In the derivations which follow, we employ the Fisher I_n functions, where $I_0(\xi)$ is defined by (2) and

$$I_n(\xi) = \int_{\xi'}^{\infty} I_{n-1}(t)\, dt, \tag{6}$$

and hence

$$\frac{dI_n}{d\xi} = -I_{n-1}.$$

These functions satisfy the recurrence formula

$$(n + 1)I_{n+1} + \xi I_n - I_{n-1} = 0, \qquad n \ge -1. \tag{7}$$

$I_n(\xi)$ is ordinarily abbreviated to I_n in this paper. Where no confusion seems likely to occur, similar abbreviations are used for other functions of ξ.

We now obtain certain relations for use in subsequent derivations. Equations (2), (5), and (6) enable us to write

$$(8)\quad \frac{\partial I_0'}{\partial \xi'} = -I_{-1}' = -\varphi(\xi'), \quad \frac{\partial I_0''}{\partial \xi'} = -I_{-1}'' = -\varphi(\xi''), \quad \frac{\partial I_0''}{\partial \sigma} = -I_{-1}''\frac{\partial \xi''}{\partial \sigma},$$

where $\varphi(\xi)$ is the ordinate of the normal frequency curve; i.e., $\varphi(\xi) = \frac{1}{\sqrt{2\pi}}e^{-\xi^2/2}$. Ordinarily we abbreviate $\varphi(\xi')$ to φ' and $\varphi(\xi'')$ to φ''. On differentiating (5) we have

$$(9)\qquad \frac{\partial \xi''}{\partial \sigma} = -\frac{R}{\sigma^2}$$

and hence from (8)

$$\frac{\partial I_0''}{\partial \sigma} = \varphi''\frac{R}{\sigma^2}.$$

Taking logarithms of (4), differentiating with the aid of (8) and (9), and equating to zero, we obtain the maximum likelihood estimating equations

$$(10)\qquad \begin{aligned} \frac{\partial L}{\partial \xi'} &= \frac{n_0(\varphi' - \varphi'')}{I_0' - I_0''} - \sum_1^{n_0}\left(\xi' + \frac{x_i}{\sigma}\right) = 0, \\ \frac{\partial L}{\partial \sigma} &= \left(\frac{n_0\varphi''}{I_0' - I_0''}\right)\frac{R}{\sigma^2} - \frac{n_0}{\sigma} + \frac{1}{\sigma^2}\sum_1^{n_0}\left\{x_i\left(\xi' + \frac{x_i}{\sigma}\right)\right\} = 0. \end{aligned}$$

If we define

$$(11)\qquad Z_1 = \frac{\varphi'}{I_0' - I_0''}, \qquad Z_2 = \frac{\varphi''}{I_0' - I_0''},$$

and substitute these values in (10), the estimating equations become

$$(12)\qquad \begin{aligned} \sigma[Z_1 - Z_2 - \xi'] - \nu_1 &= 0, \\ \sigma^2[1 - \xi'(Z_1 - Z_2 - \xi') - Z_2R/\sigma] - \nu_2 &= 0, \end{aligned}$$

where ν_1 and ν_2 are the first and second sample moments referred to the left terminus; i.e., $\nu_k = \sum_1^{n_0} x_i^k/n_0$.

To obtain the required estimates $\hat{\sigma}$ and $\hat{\xi}'$, it is necessary to solve the two equations of (12) simultaneously. As illustrated in Section 7, this can be accomplished without too much difficulty with the aid of the normal curve tables by using a modified Newton-Raphson method for solving two equations in two unknowns. This method is described in greater detail by Whittaker and Robinson [8]. Note that Z_1 and Z_2, cf. (11), involve only the normal curve ordinates φ' and φ'' and the areas I_0' and I_0''. Consequently they can be evaluated for any desired values of ξ' and σ from standard tables of the normal frequency function. To determine $\hat{\mu}$, substitute $\hat{\sigma}$ and $\hat{\xi}'$ in (3).

Throught this paper, we designate the maximum likelihood estimates as $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}'$ respectively, whereas corresponding population parameters are designated as μ, σ, and ξ'.

4. Case II. Number of unmeasured observations in each 'tail' known. Let the truncation points, the origin of reference, and the number of measured observations be designated as for Case I. If we let n_1 and n_2 be the number of unmeasured observations in the left and right 'tails' respectively, the likelihood function for a sample of this type is

$$(13)\quad P(x_1, x_2, \cdots, x_{n_1+n_0+n_2}) = K(1 - I_0')^{n_1} \cdot \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n_0} e^{-\frac{1}{2}\sum_1^{n_0}(\xi'+x_i/\sigma)^2} \cdot (I_0'')^{n_2},$$

where K is a constant.

We take the logarithms of (13), differentiate with the help of (8) and (9), and equate to zero to obtain the maximum likelihood estimating equations

$$(14)\quad \begin{aligned} \frac{\partial L}{\partial \xi'} &= n_1 \frac{\varphi'}{1 - I_0'} - n_2 \frac{\varphi''}{I_0''} - \sum_1^{n_0}\left(\xi' + \frac{x_i}{\sigma}\right) = 0, \\ \frac{\partial L}{\partial \sigma} &= n_2 \frac{\varphi''}{I_0''} - \frac{n_0}{\sigma} + \frac{1}{\sigma^2}\sum_1^{n_0}\left\{x_i\left(\xi' + \frac{x_i}{\sigma}\right)\right\} = 0. \end{aligned}$$

Let

$$(15)\quad Y_1 = \frac{n_1}{n_0}\frac{\varphi'}{(1 - I_0')}, \qquad Y_2 = \frac{n_2}{n_0}\frac{\varphi''}{I_0''},$$

and (14) can be written as

$$(16)\quad \begin{aligned} \sigma[Y_1 - Y_2 - \xi'] - \nu_1 &= 0, \\ \sigma^2[1 - \xi'(Y_1 - Y_2 - \xi') - Y_2R/\sigma] - \nu_2 &= 0, \end{aligned}$$

where ν_1 and ν_2 are again the first and second sample moments referred to the left terminus. The estimating equations (16) correspond to equations (12) given for Case I, and the manner of solution is the same for both cases. Y_1 and Y_2 for a given sample are functions of ξ' and σ only. They can be evaluated for any desired values of these variables from ordinary normal curve tables. As in Case I, the mean is estimated from (3).

5. Case III. Total number of unmeasured observations known, but not the number in each tail. Again, let the truncation points, the origin of reference, and the number of measured observations be designated as in the two previous cases. Let N be the total sample size and hence $N - n_0$ the combined number of unmeasured observations in both tails. In the notation of Case II, $N - n_0 = n_1 + n_2$. The likelihood function for a sample of this type is

$$(17)\quad P(x_1, x_2, \cdots, x_N) = K(1 - I_0' + I_0'')^{N-n_0} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n_0} e^{-\frac{1}{2}\sum_1^{n_0}(\xi' + x_i/\sigma)^2}.$$

Taking logarithms of (17), differentiating with the assistance of (8) and (9) and equating to zero, we obtain the maximum likelihood estimating equations

$$(18)\quad \begin{aligned} \frac{\partial L}{\partial \xi'} &= (N - n_0)\left(\frac{\varphi' - \varphi''}{1 - I_0' + I_0''}\right) - \sum_1^{n_0}\left(\xi' + \frac{x_i}{\sigma}\right) = 0, \\ \frac{\partial L}{\partial \sigma} &= (N - n_0)\left(\frac{\varphi''}{1 - I_0' + I_0''}\right)\frac{R}{\sigma^2} - \frac{n_0}{\sigma} + \frac{1}{\sigma^2}\sum_1^{n_0}\left\{x_i\left(\xi' + \frac{x_i}{\sigma}\right)\right\} = 0. \end{aligned}$$

In this instance, let

$$(19)\quad Q_1 = \left(\frac{N - n_0}{n_0}\right)\frac{\varphi'}{1 - I_0' + I_0''}, \qquad Q_2 = \left(\frac{N - n_0}{n_0}\right)\frac{\varphi''}{1 - I_0' + I_0''},$$

and (18) can be written as

$$(20)\quad \begin{aligned} \sigma[Q_1 - Q_2 - \xi'] - \nu_1 &= 0, \\ \sigma^2[1 - \xi'(Q_1 - Q_2 - \xi') - Q_2 R/\sigma] - \nu_2 &= 0. \end{aligned}$$

It will be recognized that equations (20) correspond to (12) and (16) for Cases I and II respectively. Since the manner of solving the estimating equations is identical in all three cases, it will not be discussed further here. For any given sample, Q_1 and Q_2 are functions of ξ' and σ only, and they can be evaluated for any desired values of these arguments from standard normal curve tables. In this case also, the mean is estimated from equation (3).

6. First approximations.

CASE I. In this case, the following relations will usually provide satisfactory first approximations for estimating σ and ξ':

$$(21)\quad \sigma_1 = s_x, \qquad \xi_1' = -\nu_1/s_x,$$

where s_x^2 is the sample variance; i.e., $s_x^2 = (\nu_2 - \nu_1^2)$. It should be remarked that the only penalty involved in beginning with a poor first approximation is to increase slightly the number of steps necessary before arriving at a satisfactory final approximation by the method of Section 7.

CASE II. Since n_1 and n_2 are known in this case, it is more expedient to read first approximations to ξ' and ξ'' directly from standard tables of normal curve areas where we set

$$(22)\quad \frac{n_1}{n_1 + n_0 + n_2} = 1 - I_0' = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\xi'} e^{-t^2/2}\,dt,$$

and

$$(23) \qquad \frac{n_2}{n_1 + n_0 + n_2} = I_0'' = \frac{1}{\sqrt{2\pi}} \int_{\xi''}^{\infty} e^{-t^2/2}\, dt.$$

With ξ' and ξ'' determined from (22) and (23), we obtain a first approximation for estimating σ, from equation (5), which we now write as

$$(24) \qquad \sigma_1 = R/(\xi_1'' - \xi_1').$$

Case iii. For a first approximation in this case, it will usually be satisfactory, in the absence of contrary information, to assume that the unmeasured observations are divided equally between the two tails, and then proceed as in Case II.

7. Numerical examples. As previously mentioned, a modified Newton-Raphson method for solving two equations in two unknowns is satisfactory in each of the three cases considered, for solving the estimating equations to obtain $\hat{\sigma}$ and ξ' in practical applications. A random sample from a normal population with $\mu = 0$, and $\sigma = 1$, selected from Mahalanobis's tables [9] will serve to illustrate the solution in each case.

Case i. For the sample selected, $n_0 = 32$; $\nu_1 = 1.244625$; $\nu_2 = 2.105275$; $x_0' = -1.000000$; and $R = 2.750000$. The estimating equations to be solved simultaneously for ξ' and $\hat{\sigma}$ are thus

$$\sigma[Z_1 - Z_2 - \xi'] - 1.244625 = 0,$$

$$\sigma^2[1 - \xi'(Z_1 - Z_2 - \xi') - 2.750000\, Z_2/\sigma] - 2.105275 = 0.$$

For first approximations, we employ (21) to obtain; $\sigma_1 = s_x = 0.75$; and $\xi_1' = -1.244625/0.75 = -1.66$. Beginning with these approximations, we subsequently obtain the results displayed in Table 1.

TABLE 1

Solution of estimating equations in Case I

σ	ξ' from ν_1	ξ' from ν_2	Difference
1.536313	−0.5389	−0.5387	−0.0002
1.527778	−0.5455	−0.5460	+0.0005

Interpolating in this table, we obtain $\hat{\sigma} = 1.534$ and $\hat{\xi}' = -0.541$. On substituting these values in (3) we obtain $\hat{\mu} = -0.170$. Even though the first approximations in this instance proved to be considerably in error, no appreciable increase was experienced in the number of steps necessary to arrive at the final values given.

Case ii. Solution of estimating equations (16) for this case can also be illustrated with the same sample which was used in Case I. In this instance, however, we have the additional information; $n_1 = 7$ and $n_2 = 1$. The equations to be solved are:

$$\sigma[Y_1 - Y_2 - \xi'] - 1.244625 = 0,$$

$$\sigma^2[1 - \xi'(Y_1 - Y_2 - \xi') - 2.750000\ Y_2/\sigma] - 2.105275 = 0.$$

From (22), (23) and (24) we obtain the first approximations: $\xi_1' = -0.935$; $\xi_1'' = 1.960$; and hence $\sigma_1 = 0.950$. Beginning with these values, we proceed as in Case I, and after several trials obtain the results displayed in Table 2.

TABLE 2

Solution of estimating equations in Case II

σ	ξ' from ν_1	ξ' from ν_2	Difference
1.041667	−0.9381	−0.9360	−0.0021
1.000000	−0.9820	−1.0094	+0.0274

Interpolating, we have $\hat{\sigma} = 1.039$ and $\hat{\xi}' = -0.941$. From (3) we then obtain. $\hat{\mu} = -0.022$.

CASE III. Again we use the same sample that was employed to illustrate Cases I and II. In this instance, however, we assume that the only information available about the unmeasured observations is that their total number is 8. In the notation of Section 5, we have $N = 40$, $n_0 = 32$, and hence $N - n_0 = 8$. The estimating equations in this situation are

$$\sigma[Q_1 - Q_2 - \xi'] - 1.244625 = 0,$$

$$\sigma^2[1 - \xi'(Q_1 - Q_2 - \xi') - 2.750000\ Q_2/\sigma] - 2.105275 = 0.$$

Under the assumption that 4 unmeasured observations are in each 'tail', equations (22), (23) and (24) give first approximations: $\xi_1' = -1.28$; $\xi_1'' = 1.28$; and hence $\sigma_1 = 1.074$. Starting with these values and proceeding as in the two previous cases, we obtain the results displayed in Table 3.

TABLE 3

Solution of estimating equations in Case III

σ	ξ' from ν_1	ξ' from ν_2	Difference
1.000000	−1.0794	−1.2091	+0.1297
1.100000	−1.0118	−0.9739	−0.0379

By interpolation, we have $\hat{\sigma} = 1.077$ and $\hat{}' = -1.027$. From equation (3), we then compute $\hat{\mu} = 0.106$.

8. Precision of estimates. To determine asymptotic variances of $\hat{\sigma}$ and $\hat{}'$, we construct the variance-covariance matrices. This requires that we obtain the second partial derivatives of logarithms of the likelihood function in each of the three cases considered. Results stated in (8) and (9) are involved in these derivatives.

CASE I. The second partial derivatives in this case are

$$(25)\quad \frac{\partial^2 L}{\partial \xi'^2} = n_0 f_1(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \xi' \partial \sigma} = \frac{n_0}{\sigma} f_2(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \sigma^2} = \frac{n_0}{\sigma^2} f_3(\xi', \xi'');$$

where

$$(26)\quad \begin{aligned} f_1(\xi', \xi'') &= -[1 + \xi' Z_1 - \xi'' Z_2 - (Z_1 - Z_2)^2], \\ f_2(\xi', \xi'') &= \left\{\frac{R}{\sigma} Z_2[(Z_1 - Z_2) - \xi''] + [Z_1 - Z_2 - \xi']\right\}, \\ f_3(\xi', \xi'') &= \left\{\left(\frac{R}{\sigma}\right)^2 Z_2(Z_2 + \xi'') - \left[2 - \xi'(Z_1 - Z_2 - \xi') - Z_2 \frac{R}{\sigma}\right]\right\}. \end{aligned}$$

Subsequently we obtain

$$(27)\quad V(\hat{\sigma}) = \frac{\sigma^2}{n_0}\left[\frac{-f_1}{f_1 f_3 - f_2^2}\right], \qquad V(\hat{\xi}') = \frac{1}{n_0}\left[\frac{-f_3}{f_1 f_3 - f_2^2}\right], \qquad r_{\hat{\sigma},\hat{\xi}'} = \frac{f_2}{\sqrt{f_1 f_3}}.$$

CASE II. In this case the second partial derivatives are

$$(28)\quad \frac{\partial^2 L}{\partial \xi'^2} = n_0\, g_1(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \xi' \partial \sigma} = \frac{n_0}{\sigma} g_2(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \sigma^2} = \frac{n_0}{\sigma^2} g_3(\xi', \xi''),$$

where

$$(29)\quad \begin{aligned} g_1(\xi', \xi'') &= -\left[1 + \xi' Y_1 - \xi'' Y_2 + \frac{n_0}{n_1} Y_1^2 + \frac{n_0}{n_2} Y_2^2\right], \\ g_2(\xi', \xi'') &= \left\{\frac{R}{\sigma} Y_2\left[\frac{n_0}{n_2} Y_2 - \xi''\right] + [Y_1 - Y_2 - \xi']\right\}. \\ g_3(\xi', \xi'') &= \left\{\left(\frac{R}{\sigma}\right)^2 Y_2\left(\xi'' - \frac{n_0}{n_2} Y_2\right) - [2 - \xi'(Y_1 - Y_2 - \xi') - Y_2 R/\sigma]\right\}. \end{aligned}$$

Finally we can write

$$(30)\quad V(\hat{\sigma}) = \frac{\sigma^2}{n_0}\left[\frac{-g_1}{g_1 g_3 - g_2^2}\right], \qquad V(\hat{\xi}') = \frac{1}{n_0}\left[\frac{-g_3}{g_1 g_3 - g_2^2}\right], \qquad r_{\hat{\sigma},\hat{\xi}'} = \frac{g_2}{\sqrt{g_1 g_3}}.$$

CASE III. This time, the second partial derivatives are

$$(31)\quad \frac{\partial^2 L}{\partial \xi'^2} = n_0 h_1(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \xi' \partial \sigma} = \frac{n_0}{\sigma} h_2(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \sigma^2} = \frac{n_0}{\sigma^2} h_3(\xi', \xi''),$$

where

$$
(32)\qquad
\begin{aligned}
h_1(\xi', \xi'') &= -\left[1 + \xi' Q_1 - \xi'' Q_2 + \frac{n_0}{N - n_0}(Q_1 - Q_2)^2\right],\\
h_2(\xi', \xi'') &= \left\{\frac{R}{\sigma} Q_2 \left[\left(\frac{n_0}{N - n_0}\right)(Q_2 - Q_1) - \xi''\right] + [Q_1 - Q_2 - \xi']\right\},\\
h_3(\xi', \xi'') &= \left\{\left(\frac{R}{\sigma}\right)^2 Q_2 \left(\xi'' - \frac{n_0}{N - n_0} Q_2\right)\right.\\
&\qquad \left. - \left[2 - \xi'(Q_1 - Q_2 - \xi') - Q_2 \frac{R}{\sigma}\right]\right\}.
\end{aligned}
$$

Accordingly we obtain

$$
(33)\quad V(\hat{\sigma}) = \frac{\sigma^2}{n_0}\left[\frac{-h_1}{h_1 h_3 - h_2^2}\right], \quad V(\hat{\xi}') - \frac{1}{n_0}\left[\frac{-h_3}{h_1 h_3 - h_2^2}\right], \quad r_{\hat{\sigma}, \hat{\xi}'} = \frac{h_2}{\sqrt{h_1 h_3}}.
$$

Note that variances of the estimates for each case considered, can be computed for given values of ξ' and σ from standard normal tables of areas and ordinates.

9. Singly truncated samples. If only the left 'tail' is missing from the samples thus far considered, then $\xi'' = \infty$, $n_2 = 0$, $\varphi'' = 0$, $I_0'' = 0$, and hence Z_2, Y_2, and Q_2 each equal zero. Upon substituting these values in (12), (16), and (20) respectively, estimating equations applicable to singly truncated samples are obtained as special cases of the estimating equations for doubly truncated samples. Of course Cases II and III become identical when samples are singly truncated. When $Y_2 = Q_2 = 0$, then $Y_1 = Q_1$, cf. (15) and (19).

Case i. With $Z_2 = 0$, the estimating equations (12) become

$$
(34)\qquad
\begin{aligned}
\sigma[Z_1 - \xi'] &= \nu_1,\\
\sigma^2[1 - \xi'(Z_1 - \xi')] &= \nu_2.
\end{aligned}
$$

Eliminating σ between these two equations we have

$$
(35)\qquad \frac{\nu_2}{\nu_1^2} = \frac{1}{Z_1 - \xi'}\left(\frac{1}{Z_1 - \xi'} - \xi'\right),
$$

which is recognized as the Pearson-Lee-Fisher equation in a form which was previously given by the author [5].

Case ii. With $Y_2 = 0$, the estimating equations (16) become

$$
(36)\qquad
\begin{aligned}
\sigma[Y_1 - \xi'] &= \nu_1\\
\sigma^2[1 - \xi'(Y_1 - \xi')] &= \nu_2.
\end{aligned}
$$

Eliminating σ between the above equations, we obtain

$$
(37)\qquad \frac{\nu_2}{\nu_1^2} = \frac{1}{Y_1 - \xi'}\left(\frac{1}{Y_1 - \xi'} - \xi'\right),
$$

which is in a form completely analogous to (35). Furthermore, this equation can be solved for $\hat{\xi}'$ in the same manner as (35), cf. [5]. Since σ can be eliminated between estimating equations in singly truncated cases, but not in doubly

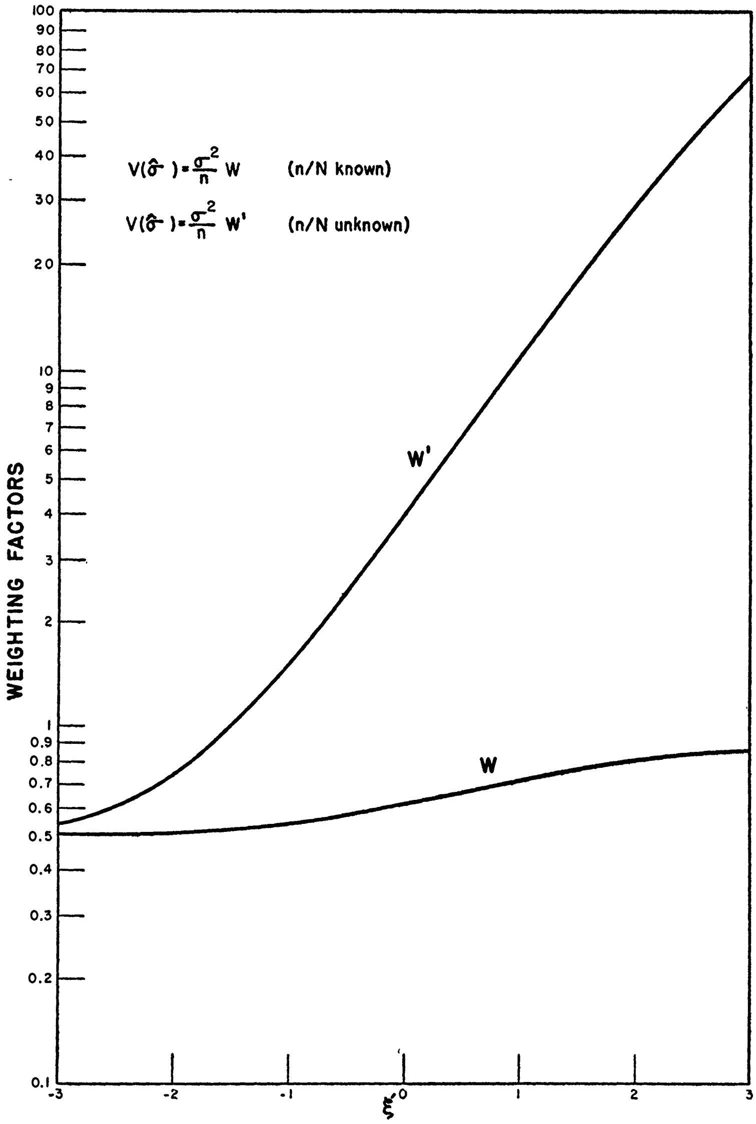

Fig. 1. Weighting factors for use in determining the variance of $\hat{\sigma}$.

truncated cases, the numerical computations are much simpler and less laborious for singly truncated samples.

If the right rather than the left tail is missing from singly truncated samples, applicable estimating equations can be obtained from (12) and (16) by translating the origin to the terminus on the right and setting Z_1 and Y_1 equal to zero rather than Z_2 and Y_2.

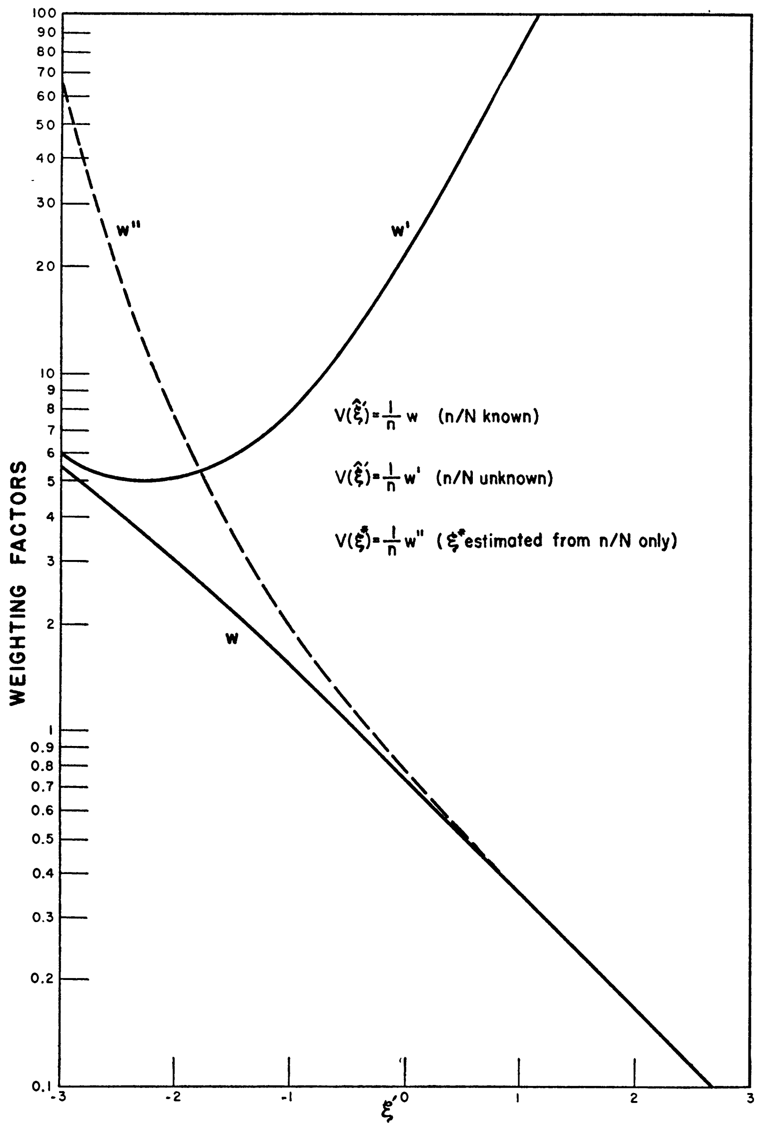

FIG. 2. Weighting factors for use in determining the variances of $\hat{\xi}'$ and ξ^*.

The variance formulas (25) and (28) likewise assume more simple forms with singly truncated samples. Substitute $Z_2 = 0$ in (25) and the variance formulas applicable with singly truncated samples when the number of unmeasured observations is unknown, become identical in form with those previously given by the writer in [5]. When the number of unmeasured observations in a singly truncated sample is known, the applicable variance formulas (28), on setting $Y_2 = 0$, become

$$V(\hat{\sigma}) = \frac{\sigma^2}{n} W(\xi') \quad \text{and} \quad V(\hat{\xi}') = \frac{1}{n} w(\xi'), \tag{38}$$

where W and w may be regarded as weighting functions defined by

$$W(\xi') = \frac{1 + Y_1(Y_1 n_0/n_1 + \xi')}{[2 - \xi'(Y_1 - \xi')][1 + Y_1(Y_1 n_0/n_1 + \xi')] - [Y_1 - \xi']^2} \tag{39}$$

and

$$w(\xi') = \frac{2 - \xi'(Y_1 - \xi')}{[2 - \xi'(Y_1 - \xi')][1 + Y_1(Y_1 n_0/n_1 + \xi')] - [Y_1 - \xi']^2}. \tag{40}$$

Similarly, the correlation between sampling errors of $\hat{\sigma}$ and $\hat{\xi}'$ in this case becomes

$$r_{\hat{\sigma},\hat{\xi}'} = \frac{Y_1 - \xi'}{\sqrt{[2 - \xi'(Y_1 - \xi')][1 + Y_1(Y_1 n_0/n_1 + \xi')]}}. \tag{41}$$

A comparison of the variances (38), with those applicable when the number of unmeasured observations is unknown, serves to indicate the extent to which information contained in a singly truncated sample is increased by adding knowledge of the number of unmeasured observations. To facilitate such comparisons, W, w, and corresponding functions W' and w' applicable when the number of unmeasured observations is unknown, are displayed graphically in Figures 1 and 2. In computing the plotted values of W and w, the ratio n/N in (39) and (40) was replaced by I_0. This ratio is, of course, an estimate of I_0, and for n and N sufficiently large, the substitution is amply justified. Equations for W' and w' can be found in [5]. For further comparisons, a graph of w'' applicable in determining the variance $V(\xi^*)$, where ξ^* is estimated from n/N alone is also included in Figure 2. This latter function is defined as

$$w''(\xi^*) = \frac{I_0^2(1 - I_0)}{\varphi^2}. \tag{42}$$

It follows from the well known formula for the variance of ξ^*:

$$V(\xi^*) = \frac{1}{N}\left\{\frac{I_0(1 - I_0)}{\varphi_2}\right\} = \frac{1}{n}\left\{\frac{I_0^2(1 - I_0)}{\varphi_2}\right\}. \tag{43}$$

An examination of Figures 1 and 2 discloses that except when the omitted portion of the distribution is small ($\xi' < -3$), the variances of the estimates of σ and ξ' based on singly truncated normal samples are substantially less when the number of unmeasured observations is known than when this information is lacking.

REFERENCES

[1] K. Pearson and A. Lee, "On the generalized probable error in multiple normal correlation", *Biometrika*, Vol. 6 (1908), pp. 59–68.

[2] A. Lee, "Table of Gaussian 'tail' functions when the 'tail' is larger than the body", *Biometrika*, Vol. 10 (1915), pp. 208–215.

[3] R. A. Fisher, "Properties and applications of H_h functions", *Mathematical Tables*, Vol. 1, pp. xxvi–xxxv, British Association for the Advancement of Science, 1931.

[4] 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. 119–134.

[5] A. C. Cohen, Jr., "On estimating the mean and standard deviation of truncated normal distributions", *Jour. Am. Stat. Assn.*, Vol. 44 (1949), pp. 518–525.

[6] W. L. Stevens, "The truncated normal distribution", appendix to "The Calculation of the Time-Mortality Curve" by C. I. Bliss, *Annals of Applied Biology*, Vol. 24 (1937), pp. 815–852.

[7] W. G. Cochran, "Use of IBM equipment in an investigation of the truncated normal problem", *Proc. Research Forum*, International Business Machines Corp., 1946, pp. 40–43.

[8] E. T. Whittaker and G. Robinson, *The Calculus of Observations*, Second Ed., Blackie and Son, Ltd., London and Glasgow, 1929, pp. 88–91.

[9] P. C. Mahalanobis, "Tables of random samples from a normal population", *Sankhyā*, Vol. 1 (1934), pp. 289–328.