# THE USE OF GENERALIZED PROBABILITY PAPER FOR CONTINUOUS DISTRIBUTIONS[1]

By Herman Chernoff and Gerald J. Lieberman

*Stanford University*

**1. Summary.** The problem of plotting on probability paper is extended to continuous distributions which are completely specified except for scale and location parameters. Necessary and sufficient conditions are given to ensure that the plot which is optimal for estimating the scale parameter is also optimal for estimating each of the percentiles.

**2. Introduction.** In a previous paper [2], the question of how to plot a sample from a normal population on normal probability paper was raised. The main purpose of that paper was to illustrate that the optimal construction of a graph depends on the use to which the graph would be put. In particular, the best plot for estimating the mean and standard deviation was discussed. Although the proposed method of plotting was considered to be merely an illustration of the above-mentioned principle, considerable comment about its usefulness was aroused. Therefore, it was decided to extend the problem to a general continuous distribution with finite variance which is specified except for a location and scale parameter. Special examples of interest are the exponential and extreme-value distributions.

The optimization methods used in this paper are applications of the method of Lagrange multipliers, and they essentially reproduce some of the results given by Downton [4][5], Godwin [7], Lloyd [11], and Sarhan [12][13].

**3. Preliminaries.** Let $x_1, x_2, \cdots, x_n$ be the ordered observations on a continuous chance variable $X$, where

$$X = \mu + \sigma Y \tag{1}$$

and where $Y$ has mean 0 and variance 1. By a suitable monotonic transformation of the vertical scale, it is possible to transform the c.d.f. of $Y$ and of all linear functions of $Y$ to straight lines. In fact, this is accomplished by plotting the $p$ percentile at a distance $v = F^{-1}(p)$ above the $x$-axis where $F$ is the c.d.f. of $Y$.

We shall use the term "plot" to represent a choice of $n$ numbers $p_1, p_2, \cdots, p_n$ (or the corresponding $v$'s, $v_1, v_2, \cdots, v_n$) which are attached to $x_1, x_2, \cdots, x_n$, respectively. It will be understood that the use of a "plot" corresponds to the plotting of the points $(x_1, p_1), (x_2, p_2), \cdots, (x_n, p_n)$. Three examples of such plots are $(x_1, 1/n), (x_2, 2/n), \cdots, (x_n, n/n)$; $(x_1, 1/(n+1)), (x_2, 2/(n+1)), \cdots, (x_n, n/(n+1))$; and $(x_1, 1/2n), (x_2, 3/2n), \cdots, (x_n, (2n-1)/2n)$. It will frequently be more convenient to consider the points in the linear scale, i.e., the points $(x_1, v_1), (x_2, v_2), \cdots, (x_n, v_n)$. In the first example mentioned above,

Received May 31, 1955.

[1] Work done under the sponsorship of the Office of Naval Research.

the plot in the linear scale is represented by $(x_1, F^{-1}(1/n))$, $(x_2, F^{-1}(2/n))$, $\cdots$, $(x_n, F^{-1}(n/n))$. Since there is no obvious rationale for choosing a plot, there arises the problem of selecting an "optimum" plot.

A "plot" is to be used to estimate the scale parameter or the percentiles of the $X$ distribution in the following fashion. Visually fit a straight line through the $n$ points. We shall assume that this fitted straight line is a good approximation to the line which would be obtained by minimizing the sum of the squares of the horizontal deviations of the points to the line. We take horizontal deviations because the $x_i$ are the random variables. Suppose that this fitted straight line is given by

$$x = a + bv. \tag{2}$$

An estimate of the standard deviation, $\sigma$, is given by $b$. If it is desired to estimate the $p_0$ percentile, we may use $x_0 = a + bv_0$, where $v_0 = F^{-1}(p_0)$. Graphically, these procedures are described as follows. To estimate $\sigma$, take the differences of the abscissas on two points of the line where the ordinates are the c.d.f. values corresponding to $\mu$ and $\mu + \sigma$. Since these c.d.f. values are $F(0)$ and $F(1)$, the ordinates in the $v$ scale are 0 and 1. To estimate the $p_0$ percentile, take that value of $x$ where the line has ordinate $p_0$ ($v_0$ in the linear scale).

The problem of estimating $\mu$ may be regarded as that of estimating the $p_0$ percentile, where $p_0 = F(0)$. One can treat the mode or other location parameters similarly.

To each plot there are associated estimates of $\sigma$ and the percentiles. If we assume that the visually fitted straight line is actually the least-squares line, these estimates are of a special type. In fact,

$$\hat{\sigma} = b = \frac{\sum_{i=1}^{n} x_i(v_i - \bar{v})}{\sum_{i=1}^{n} (v_i - \bar{v})^2}, \tag{3}$$

$$a = \bar{x} - b\bar{v}, \tag{4}$$

$$\hat{x}_0 = \bar{x} + (v_0 - \bar{v}) \cdot \frac{\sum_{i=1}^{n} x_i(v_i - \bar{v})}{\sum_{i=1}^{n} (v_i - \bar{v})^2}. \tag{5}$$

The estimate of $\sigma$ is a *contrast* in the ordered observations (i.e., a linear function of the $x_i$, the sum of whose coefficients is zero). The estimate of $x_0$ is a weighted average of the ordered observations. Let

$$u_i = \frac{(v_i - \bar{v})}{\sum_{i=1}^{n} (v_i - \bar{v})^2},$$

$$w_i = (v_0 - \bar{v})u_i, \tag{6}$$

$$\beta_i = E(y_i), \tag{7}$$

where $y_i$ is $i$th ordered observation of a sample of $n$ observations on $Y$, and let

$$\sigma_{ij} = E\{(y_i - \beta_i)(y_j - \beta_j)\}. \tag{8}$$

Let $u$, $w$, $\beta$, and $x$ be column vectors whose elements are $u_i$, $w_i$, $\beta_i$ and $x_i$, respectively. Let $\Sigma = \| \sigma_{ij} \|$ and let $e$ be the column vector, all of whose elements are $1/n$.

Then we may write

$$\hat{\sigma} = u'x, \tag{9}$$

$$\hat{x}_0 = (e + w)'x. \tag{10}$$

Note that the definition of $u$ imposes the sole restriction $e'u = 0$ on $u$. Similarly, the definition of $w$ imposes the sole restriction $e'w = 0$ on $w$ unless $v_0 - \bar{v} = 0$, in which case $w = 0$.

The following relations hold:

$$E(x) = n\mu e + \sigma\beta, \tag{11}$$

$$\beta' e = 0, \tag{12}$$

$$e'\Sigma e = \frac{1}{n} = e'e, \tag{13}$$

$$v_i - \bar{v} = \frac{u_i}{u'u} = \frac{(v_0 - \bar{v})w_i}{w'w}. \tag{14}$$

We may also remark that $\Sigma$ is positive definite.

**4. Estimation of $\sigma$.** In this section we derive the plots which yield the minimum variance unbiased estimate of $\sigma$, and the estimate of $\sigma$ with the minimum second moment about $\sigma$. For the first we minimize

$$E(\hat{\sigma} - \sigma)^2 = \sigma^2[u'\Sigma u] \tag{U.S.D.1}$$ [2]

subject to the restrictions

$$\beta' u = 1, \qquad e'u = 0.$$

We obtain

$$\Sigma u = \lambda_1 \beta + \lambda_2 e$$

or

$$u = \lambda_1 \Sigma^{-1}\beta + \lambda_2 \Sigma^{-1} e, \tag{U.S.D.2}$$

[2] In Section 4 all the equations that are prefixed by U.S.D. indicate that they are applicable to the case of unbiased estimation of the standard deviation. All the equations that are prefixed by B.S.D. indicate that they are applicable to the case of biased estimation of the standard deviation.

where $\lambda_1$ and $\lambda_2$, the Lagrange multipliers determined by the above restrictions, are

(U.S.D.3) $$\lambda_1 = \frac{e'\Sigma^{-1}e}{\Delta},$$

(U.S.D.4) $$\lambda_2 = \frac{-\beta'\Sigma^{-1}e}{\Delta}.$$

(15) $$\Delta = (\beta'\Sigma^{-1}\beta)(e'\Sigma^{-1}e) - (\beta'\Sigma^{-1}e)^2.$$

Thus,

$$E(\hat{\sigma} - \sigma)^2 = \sigma^2(u'\Sigma u) = \sigma^2 u'(\lambda_1\beta + \lambda_2 e) = \sigma^2\lambda_1,$$

or

(U.S.D.5) $$E(\hat{\sigma} - \sigma)^2 = \frac{\sigma^2(e'\Sigma^{-1}e)}{\Delta}.$$

Now let us derive the equations for the plot which yields the estimate of $\sigma$ that has the minimum second moment about $\sigma$. We minimize

(B.S.D.1) $$E(\hat{\sigma} - \sigma)^2 = \sigma^2[u'\Sigma u + (\beta'u - 1)^2]$$

subject to the restriction $e'u = 0$. We have $\Sigma u + (\beta'u - 1)\beta = \lambda e$, where $\lambda$ is the Lagrange multiplier. It then follows that $u'\Sigma u + (\beta'u - 1)\beta'u = 0$, whence $E\{(\hat{\sigma} - \sigma)^2\}/\sigma^2 = u'\Sigma u + (\beta'u - 1)^2 = 1 - \beta'u$. Now,

$$u = \lambda\Sigma^{-1}e + (1 - \beta'u)\Sigma^{-1}\beta,$$

$$\lambda e'\Sigma^{-1}e + (1 - \beta'u)e'\Sigma^{-1}\beta = 0,$$

$$\lambda\beta'\Sigma^{-1}e + (1 - \beta'u)\beta'\Sigma^{-1}\beta = \beta'u = 1 - (1 - \beta'u),$$

$$\lambda = \frac{-e'\Sigma^{-1}\beta}{\Delta^*}, \qquad 1 - \beta'u = \frac{e'\Sigma^{-1}e}{\Delta^*},$$

(16) $$\Delta^* = \Delta + e'\Sigma^{-1}e,$$

(B.S.D.2) $$u = \frac{1}{\Delta^*}[(e'\Sigma^{-1}e)\Sigma^{-1}\beta - (e'\Sigma^{-1}\beta)\Sigma^{-1}e].$$

and

(B.S.D.3) $$E\{(\hat{\sigma} - \sigma)^2\} = \frac{\sigma^2 e'\Sigma^{-1}e}{\Delta^*}.$$

It might be noted that the $u$ vectors for the unbiased and biased estimates are proportional. In fact, the biased estimates could easily be derived from the unbiased by the same arguments as those used by the authors in [2] and by Goodman in [8].

It should be noted that the value of $\bar{v}$ was immaterial. Geometrically, this means merely that raising or lowering the line does not change its slope.

**5. Estimation of the $p_0$ percentile.** In this section we shall discuss the plot which furnishes the minimum variance unbiased estimate of the $p_0$ percentile, $x_0$, and also the plot which furnishes the estimate with the minimum second moment about $x_0$.

For the first problem we minimize

(U.P.1)[3] $$E\{(\hat{x}_0 - x_0)^2\} = \sigma^2(e + w)'\Sigma(e + w)$$

subject to the restrictions

$$e'w = 0, \qquad \beta'w = v_0 .$$

We have

$$\Sigma(e + w) = \lambda_1\beta + \lambda_2 e$$

or

(U.P. 2) $$w = \lambda_1\Sigma^{-1}\beta + \lambda_2\Sigma^{-1}e - e,$$

where $\lambda_1$ and $\lambda_2$ are the Lagrange multiplier given by

(U.P.3) $$\lambda_1 = \frac{v_0 e'\Sigma^{-1}e - \frac{1}{n}\beta'\Sigma^{-1}e}{\Delta}$$

and

(U.P.4) $$\lambda_2 = \frac{\frac{1}{n}\beta'\Sigma^{-1}\beta - v_0 e'\Sigma^{-1}\beta}{\Delta}.$$

Thus,

$$E\{(\hat{x} - x_0)^2\} = \sigma^2(e + w)'(\lambda_1\beta + \lambda_2 e) = \sigma^2(\lambda_1 v_0 + \lambda_2/n)$$

or

(U.P.5) $$E\{(\hat{x}_0 - x_0)^2\} = \frac{\sigma^2}{\Delta}\left(\frac{\beta}{n} - v_0 e\right)'\Sigma^{-1}\left(\frac{\beta}{n} - v_0 e\right).$$

For the problem of minimizing the second moment of $\hat{x}_0$ about $x_0$, we minimize

(B.P.1) $$E\{(\hat{x}_0 - x_0)^2\} = \sigma^2[(e + w)'\Sigma(e + w) + (\beta'w - v_0)^2]$$

subject to $e'w = 0$. We have

$$\Sigma(e + w) + (\beta'w - v_0)\beta = \lambda e,$$

$$(e + w)'\Sigma(e + w) + (\beta'w - v_0)(\beta'w) = \lambda/n,$$

[3] In Sections 5 and 6 all the equations that are prefixed by U.P. indicate that they are applicable to the case of unbiased estimation of the $p_0$ percentile. All the equations that are prefixed by B.P. indicate that they are applicable to the case of biased estimation of the $p_0$ percentile.

whence

$$E\{(\hat{x} - x_0)^2\} = \sigma^2[(e + w)'\Sigma(e + w) + (\beta'w - v_0)^2]$$
$$= \sigma^2[\lambda/n + v_0(v_0 - \beta'w)].$$

Now,

$$\text{(B.P.2)} \quad \begin{aligned} e + w &= \lambda\Sigma^{-1}e + (v_0 - \beta'w)\Sigma^{-1}\beta, \\ \lambda e'\Sigma^{-1}e &+ (v_0 - \beta'w)e'\Sigma^{-1}\beta = 1/n, \\ \lambda\beta'\Sigma^{-1}e &+ (v_0 - \beta'w)\beta'\Sigma^{-1}\beta = \beta'w = v_0 - (v_0 - \beta'w), \end{aligned}$$

$$\text{(B.P.3)} \qquad v_0 - \beta'w = \frac{v_0\,(e'\Sigma^{-1}e) - \frac{1}{n}\,(\beta'\Sigma^{-1}e)}{\Delta^*},$$

$$\text{(B.P.4)} \qquad \lambda = \frac{\frac{\beta'\Sigma^{-1}\beta + 1}{n} - v_0 e'\Sigma^{-1}\beta}{\Delta^*}.$$

Thus,

$$\text{(B.P.5)} \quad E\{(\hat{x} - x_0)^2\} = \frac{\sigma^2}{\Delta^*}\left[\frac{\beta'\Sigma^{-1}\beta + 1}{n^2} - \frac{v_0 e'\Sigma^{-1}\beta}{n} + v_0^2(e'\Sigma^{-1}e) - \frac{v_0}{n}\,(\beta'\Sigma^{-1}e)\right]$$

or

$$\text{(B.P.6)} \quad E\{(\hat{x} - x_0)^2\} = \frac{\sigma^2}{\Delta^*}\left[\left(\frac{\beta}{n} - v_0 e\right)'\Sigma^{-1}\left(\frac{\beta}{n} - v_0 e\right) + \frac{1}{n^2}\right].$$

In each of these problems the value of $\bar{v}$ is not determined. Geometrically, this means that if the $v_0 - v_i$ are multiplied by a factor, the position where the fitted line has ordinate $v_0$ will not be affected. Since $\bar{v}$ is the only undetermined element of the optimal plot, it follows that the optimal weighted average of the ordered observations for estimating $x_0$ is *unique* for both the biased and unbiased cases.

**6. Invariance.** In this section we shall study the conditions which imply that the optimal plot does not depend on the percentile being estimated. In fact, we shall call a plot an *invariant optimal plot* if for each $x_0$ it yields an estimate $\hat{x}_0$ which minimizes $E\{(\hat{x}_0 - x_0)^2\}$. We shall call a plot an *invariant optimal unbiased plot* if for each $x_0$ it yields an unbiased estimate $\hat{x}_0$ which minimizes

$$E[(\hat{x}_0 - x_0)^2].$$

We shall use the terms *optimal weighted average for* $x_0$ and *optimal unbiased weighted average for* $x_0$ similarly.

LEMMA 1. *If there are real numbers c and k such that* $\Sigma e = ce + k\beta$, *it follows that*

$$\Sigma e = e + k\beta, \tag{17}$$

$$\beta'\Sigma^{-1}e = -k\beta'\Sigma^{-1}\beta, \tag{18}$$

$$e'\Sigma^{-1}e = \frac{1}{n} + k^2\beta'\Sigma^{-1}\beta, \tag{19}$$

$$\Delta = \frac{1}{n}\beta'\Sigma^{-1}\beta, \tag{20}$$

$$\Delta^* = \frac{1}{n} + \beta'\Sigma^{-1}\beta\left[\frac{1}{n} + k^2\right]. \tag{21}$$

PROOF. We apply

$$e'\Sigma e = e'e$$

to obtain $c = 1$. Premultiplying

$$e = \Sigma^{-1}e + k\Sigma^{-1}\beta$$

by $\beta'$ and $e'$, we obtain (18) and (19). The remaining equations follow by substitution.

THEOREM 1. *There is an invariant optimal unbiased plot if and only if* $\Sigma e = e + k\beta$. *In that case the invariant optimal plot is unique, has* $\bar{v} = 0$, *and is optimal for the unbiased estimation of* $\sigma$. *Also,*

$$E\{\hat{x}_0 - x_0\}^2 = \sigma^2\left\{\frac{(1 + nv_0k)^2}{n} + \frac{v_0^2}{\beta'\Sigma^{-1}\beta}\right\}. \tag{U.P.6}$$

PROOF. Suppose that there is an invariant optimal unbiased plot. Then (see (U.P. 2, 3, and 4)),

$$w = (v_0 - \bar{v})w^{(1)} + w^{(2)},$$

where

$$w^{(1)} = \frac{e'\Sigma^{-1}e}{\Delta}\Sigma^{-1}\beta - \frac{e'\Sigma^{-1}\beta}{\Delta}\Sigma^{-1}e$$

and

$$w^{(2)} = \frac{\bar{v}(e'\Sigma^{-1}e) - \frac{1}{n}\beta'\Sigma^{-1}e}{\Delta}\Sigma^{-1}\beta + \frac{\frac{1}{n}\beta'\Sigma^{-1}\beta - \bar{v}e'\Sigma^{-1}\beta}{\Delta}\Sigma^{-1}e - e.$$

Applying (14), we have

$$\sum_{i=1}^{n}(v_i - \bar{v})^2 = \frac{(v_0 - \bar{v})^2}{[(v_0 - \bar{v})w^{(1)} + w^{(2)}]'[(v_0 - \bar{v})w^{(1)} + w^{(2)}]}.$$

A necessary and sufficient condition that the plot be invariant is that $w^{(2)} = 0$, which is equivalent to $\Sigma e = ce + k\beta$ or $\Sigma e = e + k\beta$ by Lemma 1. But then

$$\frac{\frac{1}{n}(\beta'\Sigma^{-1}\beta - \bar{v}e'\Sigma^{-1}\beta)}{\Delta} = 1 \qquad \text{or} \qquad \bar{v} = 0$$

and the plot is unique and coincides (see (U.S.D. 2, 3, and 4)) with one which is optimal for the unbiased estimation of $\sigma$. Now suppose only that $\Sigma e = e + k\beta$. Let $\bar{v} = 0$. Then, $w^{(2)} = 0$ and

$$\text{(U.P.7)} \qquad w = v_0\left\{\frac{e'\Sigma^{-1}e}{\Delta}\Sigma^{-1}\beta - \frac{e'\Sigma^{-1}\beta}{\Delta}\Sigma^{-1}e\right\}$$

furnishes an invariant optimal unbiased plot. Substituting in (U.P. 5), we obtain (U.P. 6).

THEOREM 2. *There is an invariant optimal biased plot if and only if* $\Sigma e = e + k\beta$. *In that case the invariant optimal plot is unique, has* $\bar{v} = k$, *and is optimal for the estimation of* $\sigma$. *Also,*

$$\text{(B.P.7)} \qquad E\{\hat{x}_0 - x_0)^2\} = \sigma^2\left\{\frac{(\beta'\Sigma^{-1}\beta)\left(\frac{1}{n} + kv_0\right)^2 + \frac{v_0^2}{n} + \frac{1}{n^2}}{(\beta'\Sigma^{-1}\beta)\left(\frac{1}{n} + k^2\right) + \frac{1}{n}}\right\}.$$

PROOF. As in the proof of Theorem 1, we find that a necessary and sufficient condition that a plot be invariant is that $w^{(2)} = 0$, where now

$$w^{(2)} = \frac{\bar{v}(e'\Sigma^{-1}e) - \frac{1}{n}(e'\Sigma^{-1}\beta)}{\Delta^*}\Sigma^{-1}\beta + \frac{\frac{\beta'\Sigma^{-1}\beta + 1}{n} - \bar{v}e'\Sigma^{-1}\beta}{\Delta^*}\Sigma'e - e.$$

Hence, if a plot is invariant, $\Sigma e = ce + k\beta = e + k\beta$. But if $w_2 = 0$,

$$\frac{\frac{\beta'\Sigma^{-1}\beta + 1}{n} - \bar{v}e'\Sigma^{-1}\beta}{\Delta^*} = 1 \qquad \text{or} \qquad \bar{v} = k,$$

and the plot is unique and coincides (see (B.S.D. 2)) with one which is optimal for the estimation of $\sigma$. Now suppose $\Sigma e = e + k\beta$. Let $\bar{v} = k$. Then, $w^{(2)} = 0$ and

$$\text{(B.P.8)} \qquad w = (v_0 - k)\left[\frac{e'\Sigma^{-1}e}{\Delta^*}\Sigma^{-1}\beta - \frac{e'\Sigma^{-1}\beta}{\Delta^*}\Sigma^{-1}e\right]$$

furnishes an invariant optimal plot. Substituting in (B.P.6), we obtain (B.P.7).

COROLLARY 1. *A necessary and sufficient condition that $\bar{x}$ is the optimal unbiased weighted average for estimating $\mu$ is that $\Sigma e = e + k\beta$.*

PROOF. The sufficiency is trivial. For the necessity, we note that if $\bar{x}$ is optimal, $w = 0$ and hence $\Sigma e = \lambda_1\beta + \lambda_2 e$, and by Lemma 1, $\Sigma e = e + k\beta$.

COROLLARY 2. *A necessary and sufficient condition that $\bar{x}$ is the optimal weighted average for estimating the $p_0 = F(v_0)$ percentile is that $\Sigma e = e + v_0\beta$.*

The proof is similar to that of Corollary 1 and is omitted. Note that as a particular case, $\bar{x}$ is optimal for estimating $\mu$ if and only if $\Sigma e = e$.

The following corollaries are of some interest because they relate notions like sufficiency, completeness, and min-max estimates with the covariance matrix of the ordered observations. On the other hand they may not yield many applications even if they were refined in the more or less obvious ways.

COROLLARY 3. *If $\bar{x}$ is a function of a sufficient statistic for $(\mu, \sigma)$ whose family of distributions is complete, then $\Sigma e = e + k\beta$.*

PROOF. Under the above assumptions, it follows that $\bar{x}$ is a minimum variance unbiased estimate of $\mu$. (See [1] and [10].) By Corollary 1, the result follows.

COROLLARY 4. *If for fixed $\sigma$, $\bar{x}$ is a* min-max *estimate of $\mu$ with respect to the quadratic loss function, then $\Sigma e = e$.*

PROOF. Let $t$ be a weighted average of the ordered observations. Its risk is given by

$$r_t = E\{(t - \mu)^2\} = a_t\sigma^2,$$

where $a_t$ is independent of $\mu$ and $\sigma$. Similarly,

$$r_{\bar{x}} = E\{(\bar{x} - \mu)^2\} = \frac{1}{n}\sigma^2.$$

If $\bar{x}$ is min-max, we must have $a_t \geqq 1/n$. Corollary 2 then yields our result.

This proof would apply equally well if it were given that $\bar{x}$ is a min-max estimate among invariant estimates with respect to the loss function $(t - \mu)^2/\sigma^2$.

Results similar to the above corollaries have been obtained by Lloyd [11] for the special case of symmetric distributions. In that case if $\Sigma e = e + k\beta$, $k$ is necessarily zero.

**7. Examples.**

EXAMPLE 1. *The exponential distribution.* Let us derive the means and covariances of the order statistics of a sample of $n$ independent observations with density

$$f(z) = e^{-z} \quad \text{for } z \geqq 0, \qquad (22)$$
$$= 0 \quad \text{for } z < 0.$$

Following the method of Epstein and Sobel [6], we note first that if $Z$ has the above exponential density, then the conditional density of $Z - a$, given $Z \geqq a$, also has this density. If we let $z_{in}$ be the $i$th order statistic from a sample of size $n$, it follows that

$$z_{2n} - z_{1n},\ z_{3n} - z_{1n},\ \cdots,\ z_{nn} - z_{1n}$$

have the same joint distribution as the order statistics from a sample of size $n - 1$ and are independent of $z_{1n}$. Then we might write

$$z_{in} = z_{1n} + z_{i-1,n-1} \qquad \text{for } i = 2, 3. \cdots, n,$$

where $z_{i-1,n-1}$ is independent of $z_{1n}$. Also

$$E\{z_{in}\} = E\{z_{1n}\} + E\{z_{i-1,n-1}\} \qquad \text{for } i = 2, 3, \cdots, n,$$

$$\sigma_{z_{1n}z_{in}} = \sigma_{z_{1n}z_{1n}} \qquad \text{for } i = 2, 3, \cdots, n,$$

and

$$\sigma_{z_{in}z_{jn}} = \sigma_{z_{1n}z_{1n}} + \sigma_{z_{i-1,n-1}z_{j-1,n-1}} \quad \text{for } i, j = 2, 3, \cdots, n.$$

Now,

$$P\{z_{1n} > a\} = P\{Z > a\}^n = e^{-na},$$

which implies that $E\{z_{1n}\} = 1/n$ and $\sigma^2_{z_{1n}} = 1/n^2$. It follows that

$$E\{z_{in}\} = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-i+1}$$

and

$$\sigma_{z_{in}z_{jn}} = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \cdots + \frac{1}{(n-i+1)^2} \qquad \text{for} \quad i \leqq j.$$

The chance variable $Z$ has mean 1 and variance 1. We normalize to $Y = Z - 1$ and it follows that

(23) $$\beta_i = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-i+1} - 1,$$

(24) $$\sigma_{ij} = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \cdots + \frac{1}{(n-i+1)^2}, \qquad i \leqq j,$$

and therefore

(25) $$\Sigma e = e + \frac{1}{n}\beta.$$

It follows from our results of the preceding section that there are invariant optimal plots for both the biased and unbiased estimation problems, that $\bar{x}$ is the minimum variance unbiased estimate of $\mu$ among the weighted averages of the ordered observations, and that $\bar{x}$ is not the optimal weighted average for estimating $\mu$ if bias is permitted. In fact, $\bar{x}$ is the optimal weighted average for estimating the $1 - e^{-[1+(1/n)]}$ percentile.

In this particular problem it is easy to show that

(26) $$e'\Sigma^{-1} = (n, 0, 0, 0 \ \cdots, 0),$$

(27) $$\beta'\Sigma^{-1} = (1 - n^2, 1, 1, 1, \cdots, 1).$$

For the optimal unbiased plot, we have

$$v' = \left(\frac{1}{n} - 1, \frac{1}{n}, \frac{1}{n} \cdots, \frac{1}{n}\right), \tag{28}$$

$$p' = (1 - e^{-1/n}, 1 - e^{-[1+(1/n)]}, 1 - e^{-[1+(1/n)]}, \cdots, 1 - e^{-[1+(1/n)]}), \tag{29}$$

$$E\{(\hat{\sigma} - \sigma)^2\} = \sigma^2/(n - 1), \tag{30}$$

and

$$E\{(\hat{x}_0 - x_0)^2\} = \sigma^2\left\{\frac{(1 + v_0)^2}{n} + \frac{v_0^2}{n(n - 1)}\right\}. \tag{31}$$

On the other hand, for the optimal biased plot,

$$v' = \left(\frac{1}{n} - 1, \frac{1}{n} + \frac{1}{n - 1}, \frac{1}{n} + \frac{1}{n - 1}, \cdots, \frac{1}{n} + \frac{1}{n - 1}\right), \tag{32}$$

$$p' = (1 - e^{-1/n}, 1 - e^{-[1+(1/n)+1/(n-1)]}, \cdots, 1 - e^{-[1+(1/n)+1/(n-1)]}), \tag{33}$$

$$E\{(\hat{\sigma} - \sigma)^2\} = \sigma^2/n, \tag{34}$$

and

$$E\{(\hat{x}_0 - x_0)^2\} = \frac{\sigma^2}{n}\left[(1 + v_0)^2 - \frac{(1 + 2v_0)}{n} + \frac{1}{n^2}\right]. \tag{35}$$

In this somewhat extraordinary example all order statistics except the first are plotted at the same probability level. Of course, this property, imposed by our criteria for graphing, makes the graph useless for the purpose of testing whether the distribution is exponential. In this case there seems little to be gained by using the above plots instead of algebra.

From another point of view, it is not surprising that these results appeared. In fact for the exponential distributions with unknown scale and location parameter, a sufficient statistic for $(\mu, \sigma)$ is given by $(x_1, \bar{x})$. It is not surprising, then, that the above plots lump all except the first observation at one level.

In fact, it can be shown that the family of distributions of $(x_1, \bar{x})$ is complete, and then Corollary 3 gives us the fact that $\Sigma e = e + k\beta$.

Example 2. *The normal distribution.* It is known [14] that $\bar{x}$ is a min-max estimate of $\mu$. It follows that $\Sigma e = e$. Since the normal distribution is symmetric, $\Sigma e = e$ can also be inferred from Lloyd's paper [11]. It is of interest to note that the result was also obtained by Jones [9] by a method which resembles this approach in that it uses the fact that $\bar{x}$ is a good weighted average of the ordered observations for estimating $\mu$. This result can also be obtained as follows. The differences of unordered observations are independent of $\bar{x}$ and $x_i - x_j$ is a function of these differences. Therefore, $\bar{x}$ is uncorrelated with $(x_i - x_j)$ and

$$\sigma_{x_i\bar{x}} - \sigma_{x_j\bar{x}} = (1/n)\ \Sigma_{r=1}^{n}\ (\sigma_{x_ix_r} - \sigma_{x_jx_r}) = 0.$$

The sum of the elements of rows of the covariance matrix are the same for each row. Hence this sum is one and, finally, $\Sigma e = e$.

Tables have been presented in [2] for the optimal plots using normal probability papers for sample sizes up to $n = 10$. Thes will be extended [3] to $n = 20$.

Note that $\Sigma e = e$ implies that $\beta'\Sigma^{-1}e = \beta' e = 0$. Hence the $v_i$ for the optimal biased and unbiased plots are proportional to $\Sigma^{-1}\beta$. This relation also implies that $e$ is a characteristic vector of $\Sigma$ corresponding to characteristic value 1. Furthermore, since all the elements of $\Sigma$ are positive, 1 is the largest characteristic value of $\Sigma$.

**8. Concluding Remarks.** The problems treated in this paper have been relatively simple. There are a number of questions which are more difficult and are as yet unsolved.

First, it would be very interesting to know under what conditions one can be insured that the optimal $v_i$ are increasing. It would be embarrassing to propose that the smallest observation be plotted at a higher $p$-level than the second smallest. One conjecture is that $\Sigma e = e + k\beta$ implies the desired result.

Second, for distributions which correspond to chance variables which are bounded, the range of $v$ corresponding to values of $p$ between 0 and 1 is similarly bounded. Under what conditions will the optimal plot involve only values of $v$ which correspond to values of $p$ between 0 and 1?

Third, what is the class of continuous distributions for which $\Sigma e = e$ for all $n$? For $n = 2$, all symmetric distributions have this property. Could it be that only the normal distribution has this property for all $n$ or even for any $n > 2$?

Fourth, what are the asymptotic properties of the optimal plots as $n \to \infty$?

Finally, the important question of how to plot in order to furnish a test that the distribution is in the given family is still untreated.

We conclude with the following remark. Suppose that instead of fitting the least-squares line to the plotted points, we fitted a modified least-squares line where the points had specified weights not all equal. Then it can be shown that there is a one-to-one correspondence of plots such that equivalent estimates are obtained by the two methods.

## REFERENCES

[1] D. Blackwell, "Conditional expectation and unbiased sequential estimation," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 105–110.

[2] H. Chernoff and G. J. Lieberman, "Use of normal probability paper," *J. Amer. Stat. Assoc.*, Vol. 49 (1954), pp. 778–784.

[3] H. Chernoff and G. J. Lieberman, "Engineering applications of normal probability paper," unpublished paper.

[4] F. Downton, "A note on ordered least-squares estimation," *Biometrika*, Vol. 40 (1953), pp. 457–458.

[5] F. Downton, "Least-squares estimates using ordered observations," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 303–316.

[6] B. Epstein and M. Sobel, "Life testing," *J. Amer. Stat. Assoc.*, Vol. 48 (1953), pp. 486–502.

[7] H. J. Godwin, "On the estimation of dispersion by linear systematic statistics," *Biometrika*, Vol. 36 (1949), pp. 92–100.

[8] L. A. Goodman, "A simple method for improving some estimators," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 114–117.

[9] H. L. JONES, "Exact lower moments of order statistics in small samples from a normal distribution," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 270–273.

[10] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions, and unbiased estimation. Part I," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[11] E. H. LLOYD, "Least-squares estimation of location and scale parameters using order statistics," *Biometrika*, Vol. 39 (1952), pp. 88–95.

[12] A. E. SARHAN, "Estimation of the mean and standard deviation by order statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 317–328.

[13] A. E. SARHAN, "Estimation of the mean and standard deviation by order statistics, Part II and Part III," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 505–511 and pp. 576–592.

[14] J. WOLFOWITZ, "Min-max estimates of the mean of a normal distribution with known variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 182–197.