# THE DISTRIBUTION OF QUASI-RANGES IN SAMPLES FROM A NORMAL POPULATION

By J. H. Cadwell

*Ordnance Board, Great Britain*[1]

**1. Summary.** A method is developed for the evaluation of the probability density function of the statistic:

$$w_r = x_{n-r} - x_{r+1}$$

where $x_1, x_2, \cdots, x_n$ are ordered values in a sample of $n$ from a normal population.

It is shown that, up to $n = 17$, $w_0$ is the most efficient statistic of this type for the estimation of population standard deviation. Beyond this point $w_1$ is optimum up to $n = 31$, where $w_2$ becomes better. Tables of moment constants and percentage points are given for $w_1$ over the range 10 to 30.

Similar methods are used to determine the efficiencies of two estimates of the form $w_r + \lambda w_s$.

The approximation used is compared with three other published approximations in the case of range ($r = 0$).

Godwin [5] and Nair [11] have discussed problems of this kind for sample sizes up to 10, using exact values of the first two moments. Karl Pearson [12], Mosteller [10] and Jones [9] have considered the large sample case. The methods of the present paper go some way towards filling the gap between these approaches. Moreover, they are not restricted to consideration of mean and variance only.

**2. Introduction.** The use of range as a rapid means of estimating population standard deviation is usually restricted to sample sizes below 20. There are several reasons for this restriction. Beyond this point the efficiency of such an estimate, when compared with one based on sample standard deviation, falls off rapidly. For larger samples the ratio of mean sample range to population standard deviation depends rather critically on the form of the tails of the parent distribution. Thus estimates based upon a normal model may be misleading. Finally the probability of the presence of a "rogue" observation will increase with the sample size. Such a freak observation is likely to lead to an unusually large value of range.

One method of overcoming these drawbacks consists in splitting the sample into smaller groups and finding the average range of these groups. Such a process is not unique and this may be a drawback in certain circumstances. In addition,

Received 2/6/53.

[1] This work arose during analyses made by the Board's Statistics Group. The author is indebted to the Chief Scientist, Ministry of Supply, for permission to publish this paper.

unless the order in which values are recorded is known to be independent of these values, a randomising process is necessary. This will rob the method of speed, one of the main assets of range as an estimator of standard deviation.

In a recent paper Mosteller [10] drew attention to the possible use of $w_r$, calling such statistics quasi-ranges. He mentioned an earlier result of K. Pearson's [12] concerning the use of inter-quantile distances as estimators in large samples from a normal population. In this case an optimum efficiency of 65 per cent is attained when points at a proportion 0.07 of the sample size from either end are used. For small samples, range is the most efficient estimator of this kind. Evidently, as $n$ increases, the optimum value of $r$ will cease to be zero at some point. Since $w_1$ does not depend on the values of the extreme observations, it is likely to be less affected by departures from normality or by the possible presence of an occasional "rogue" observation. Thus it should be preferable to range beyond a certain sample size.

Godwin [4], [5], in discussing a more general type of estimator, found the first two moments of $w_r$ for values of $n$ up to 10. His method depends on a series of double quadratures and becomes very laborious as $n$ increases. Below we find an asymptotic series for the p.d.f. of $w_r$. As sometimes happens (e.g. with Stirling's series) results of high accuracy are obtained for small as well as for large $n$.

**3. Derivation of the asymptotic series.** It is shown in [1] that a series expansion of the appropriate integrand leads to close approximations to moments of quantiles in the normal case. Thus, for odd $n$, the median of a set of $n$ values has a variance close to:

$$\frac{\pi}{(\pi + 2n - 2)}\left\{1 + \frac{4(\pi - 3)(n - 1)}{(\pi + 2n - 2)^2}\right\}.$$

For $n = 3$ this is in error by 0.001, and as $n$ increases the error rapidly sinks to zero.

A similar method can be applied to the p.d.f. of $w_r$ when the parent distribution is symmetrical. We denote the p.d.f. of the parent by $\varphi(x)$, and need the functions:

$$\Phi(x) = \int_0^x \varphi(x)\,dx$$

$$\theta^{(l)}(x) = \left(\frac{d}{dx}\right)^l \varphi(x) \Big/ \Phi(x), \qquad \psi^{(l)}(x) = \left(\frac{d}{dx}\right)^l \varphi(x) \Big/ \{\tfrac{1}{2} - \Phi(x)\}.$$

The p.d.f. of $w_r$ will be given by the integral:

$$(1) \quad f(w_r) = \frac{n!}{(n - 2r - 2)!\,r!\,r!} \int_{-\infty}^{\infty} [\tfrac{1}{2} + \Phi(x)]^r [\tfrac{1}{2} - \Phi(x + w_r)]^r \cdot [\Phi(x + w_r) - \Phi(x)]^{n-2r-2} \varphi(x)\varphi(x + w_r)\,dx.$$

Since $\varphi(x)$ is symmetrical, this integrand will have its maximum value at $-\frac{1}{2}w_r$, and will fall rapidly to zero on either side of this point. This suggests expanding the integrand in terms of: $t = x + \frac{1}{2}w_r$.

Except where otherwise specified, wherever a Greek function letter appears the argument will be $\frac{1}{2}w_r$. Consequently this argument is omitted in the interests of simplicity.

After taking logarithms of the appropriate series, and again expanding, we find:

$$\Phi(x + w_r) - \Phi(x) = 2\Phi \exp \{\tfrac{1}{2}\theta' t^2 + \cdots\} \tag{2}$$

$$[\tfrac{1}{2} + \Phi(x)][\tfrac{1}{2} - \Phi(x + w_r)] = (\tfrac{1}{2} - \Phi)^2 \exp \{-(\psi^2 + \psi')t^2 + \cdots\} \tag{3}$$

$$\varphi(x)\varphi(x + w_r) = \varphi^2 \exp\left\{-\left[\left(\frac{\varphi'}{\varphi}\right)^2 - \frac{\varphi''}{\varphi}\right] t^2 + \cdots\right\} \tag{4}$$

Thus, apart from a constant factor, the integrand can be written in the form:

$$\{1 + At^4 + Bt^6 + \cdots\} \exp - \left\{\left(\frac{\varphi'}{\varphi}\right)^2 - \frac{\varphi''}{\varphi} + r(\psi^2 + \psi') - \tfrac{1}{2}(n - 2r - 2)\theta'\right\} t^2.$$

Using the form of Watson's lemma given by Jeffreys [8], we see that term-by-term integration of this series yields an asymptotic series for the p.d.f. The first term of this series is:

$$f(w_r) \sim \frac{C\varphi^2(\frac{1}{2} - \Phi)^{2r}(2\Phi)^{n-2r-2}}{\left\{2\left(\frac{\varphi'}{\varphi}\right)^2 - 2\left(\frac{\varphi''}{\varphi}\right) + 2r(\psi^2 + \psi') - (n - 2r - 2)\theta'\right\}^{\frac{1}{2}}}.$$

The constant $C$ is chosen to make the area under the approximate p.d.f. equal to unity.

If we now consider the case where $\varphi(x)$ is the normal density, (4) becomes:

$$\varphi(x)\varphi(x + w_r) = \varphi^2 \exp - t^2.$$

With this modification the asymptotic series becomes:

$$f(w_r) \sim C\varphi^2(2\Phi)^{n-2r-2}(\tfrac{1}{2} - \Phi)^{2r} k\left\{1 - \frac{(n - 2r - 2)}{8}[3(\theta')^2 - \theta''']k^4 + \cdots\right\} \tag{5}$$

where
$$\frac{1}{k^2} = 2 + 2r(\psi^2 + \psi') - (n - 2r - 2)\theta'.$$

The second term of the series in brackets is of order $1/n$. The next term, of order $1/n^2$, is:

$$\frac{r}{4}[3(\psi')^2 - \psi''' - 4\psi\psi'' - 6(\psi^2 + \psi')^2]k^4$$
$$+ \frac{(n - 2r - 2)}{48}[\theta^{\mathrm{V}} - 15\theta'''\theta' + 30(\theta')^3]k^6 + \frac{35}{384}(n - 2r - 2)^2[3(\theta')^2 - \theta''']^2 k^8.$$

While these expressions are rather involved, $n$ and $r$ enter into them quite simply. Thus, once the basic combinations of the $\theta$, $\varphi$ and $\psi$ functions have been evaluated, the time taken to evaluate $f(w_r)$ is very much less than would be needed if the exact expression (1) were used.

**4. Accuracy of the dominant term.** A convenient measure of accuracy is provided by comparing the mean value of $w_r$, found from the approximate p.d.f., with the exact value:

$$E(w_r) = 2(r+1)\binom{n}{r+1}\int_{-\infty}^{\infty} x[\tfrac{1}{2} - \Phi(x)]^r[\tfrac{1}{2} + \Phi(x)]^{n-r-1}\varphi(x)\,dx. \tag{6}$$

This integral is easily evaluated by quadrature; for $r = 0$ values are given to 5 decimal places by K. Pearson [13]. The accuracy of the higher moments, and the use of further terms of the expansion, are considered in subsequent paragraphs.

For fixed $n$, $k^2$ decreases as $r$ increases, provided $w_r$ is greater than 1.8. For values of $w_r$ below 1.8, the quantity:

$$3(\theta')^2 - \theta''' \tag{7}$$

is very small. Thus it is evident from (5), that an increase of $r$ should lead to greater accuracy in the dominant term. We find that, when $n = 30$, values are as follows.

| | Exact value | Error of dominant term |
|---|---|---|
| $E(w_0)$ | 4.0855 | +0.0095 |
| $E(w_1)$ | 3.2312 | +0.0019 |
| $E(w_2)$ | 2.7296 | +0.0006 |

For fixed $r$, the effect of the second term will depend on the position of the mode of the p.d.f. The expression (7) rises from zero at $w_r = 0$ to a maximum value near $w_r = 4.2$, and then falls to zero again as $w_r$ increases.

For small $n$, all but the right tail of the distribution will lie in a region where (7) is small. For large $n$, the mode of the distribution will be well beyond the region where (7) has any appreciable effect.

Thus, as $n$ increases, the effect of the error in the dominant term will first increase and later fall to zero. Since the mean value of $w_r$ increases very slowly with $n$, the position of greatest error will occur for a sample size of the order of a hundred.

For $E(w_0)$ we find the results:

| $n$ | 2 | 8 | 20 | 30 | 60 | 100 |
|---|---|---|---|---|---|---|
| Error | 0 | 0.0058 | 0.0086 | 0.0095 | 0.0107 | 0.0114 |
| % error | 0 | 0.202 | 0.229 | 0.232 | 0.232 | 0.227 |

Thus, while the error in the mean is still increasing slowly beyond $n = 60$, the maximum percentage error occurs near 50.

For $w_1$ the mode will be smaller than for $w_0$, for the same sample size, and consequently the increasing error will be longer maintained. Errors in the mean value are given below.

| $n$ | 5 | 10 | 20 | 30 | 40 | 60 |
|---|---|---|---|---|---|---|
| Error........ | 0.00019 | 0.00083 | 0.00153 | 0.00185 | 0.00205 | 0.00227 |
| % error...... | 0.0188 | 0.0414 | 0.0543 | 0.0573 | 0.0584 | 0.0586 |

It seems likely that the maximum percentage error is little in excess of its value at $n = 60$.

**5. Application to range.** We first compare the accuracy of the dominant term with that of three other asymptotic expressions.

Gumbel [6] has found the asymptotic distribution of range in the general symmetric population. His result is:

$$f(R) = 2e^{-R}K_0(2e^{-\frac{1}{2}R}) \tag{8}$$

where

$$R = n\varphi(u)\{w_0 - 2u\}, \qquad \Phi(u) = 1 - 1/n.$$

Elfving [3] derives an asymptotic expression for the normal case, it is:

$$f(\xi) = \xi K_0(\xi), \tag{9}$$

where

$$\xi = 2n\left\{1 - \Phi\left(\frac{w_0}{2}\right)\right\}.$$

In formulae (8) and (9) $K_0$ represents a modified Bessel function of the second kind. Using the method of steepest descent, Cox [2] derives the result:

$$f(w_0) \sim \frac{n^{\frac{3}{2}}\sqrt{2\pi}\,\varphi^2\left(\frac{w_0}{2}\right)\left[2\Phi\left(\frac{w_0}{2}\right)\right]^{n-\frac{3}{2}}}{\left[-2\varphi'\left(\frac{w_0}{2}\right)\right]^{\frac{1}{2}}}. \tag{10}$$

In the case of range, the dominant term of the series considered here is:

$$f(w_0) \sim \frac{n(n-1)\sqrt{\pi}\,\varphi^2\,[2\Phi]^{n-\frac{3}{2}}}{[2\Phi - (n-2)\varphi']^{\frac{1}{2}}}. \tag{11}$$

This is asymptotically identical with (10). However, as $\varphi'$ is small for quite moderate values of $w_0$, $n$ has to be very large for good agreement between (10) and (11).

Cox examines the errors of (8), (9) and (10) when $n = 20$. His results are compared with those obtained from (11) in the following table.

| | Mean | St. Devn. | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|
| Exact value | 3.7350 | 0.7287 | 0.1627 | 3.259 |
| Error of (8) | +0.12 | +0.15 | +0.49 | +0.94 |
| Error of (9) | +0.03 | +0.04 | −0.09 | −0.18 |
| Error of (10) | +0.10 | +0.05 | +0.05 | +0.16 |
| Error of (11) | +0.0086 | +0.0025 | −0.0043 | −0.019 |

We now consider the effect of taking two terms of the series for $n = 20$, 60 and 100. For $n = 20$ the exact values are given by Hartley and Pearson [7], for 60 and 100 values are given by K. Pearson [13]. Errors are as follows:

| $n$ | Mean | St. Devn. | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|
| 20 | −0.0026 | −0.0010 | +0.0009 | +0.006 |
| 60 | −0.0042 | −0.002 | +0.001 | +0.02 |
| 100 | −0.0050 | −0.002 | +0.003 | +0.02 |

It will be seen that two terms of the series give a very good approximation to the true distribution.

**6. The first quasi-range.** The following figures illustrate the effect of taking into account successive terms of the series. They refer to the mean value of $w_1$ when $n = 30$.

| | |
|---|---|
| Exact value | 3.23120 |
| Error of first term | +0.00185 |
| Error of first two terms | −0.00025 |
| Error of first three terms | +0.00006 |
| Error of first four terms | −0.00002 |

It appears that the use of three terms gives a high degree of accuracy, at least for the mean value.

In order to examine the accuracy of higher moments and percentage points, exact values of the p.d.f. were found by quadrature for $n = 30$. The behaviour of the approximation, using three terms of the series, is shown by the following figures.

| | Exact Value | Error |
|---|---|---|
| Variance | 0.25879 | +0.00004 |
| $\beta_1$ | 0.0865 | +0.0001 |
| $\beta_2$ | 3.142 | <0.0005 |
| 95% point | 4.107 | +0.001 |
| 99.9% point | 5.021 | +0.001 |

Below the 95 per cent point errors in percentage points are always less than 0.001.

For $n = 10$, Godwin's five place value for the variance agreed exactly with that obtained when using three terms of the series.

**7. Efficiency values.** The efficiency is defined as:

$$100 \text{ var } s \bigg/ \frac{\text{var } w_r}{[E(w_r)]^2}$$

where $s$ is the unbiassed estimate of population standard deviation:

$$s = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)} \sqrt{\Sigma(x_i - \bar{x})^2}.$$

For range, values of efficiency were obtained from the tables of reference [13]. For $w_1$ and $w_2$, Godwin's values were available up to sample size 10. Beyond this, values were computed from the approximate expressions. Some values, rounded to the nearest 0.5 per cent, are given below.

| $n$ | 10 | 20 | 30 | 40 | 60 |
|---|---|---|---|---|---|
| Efficiency of $w_0$ | 85.0 | 70.0 | 60.5 | 54.0 | 44.5 |
| Efficiency of $w_1$ | 67.0 | 73.0 | 70.0 | 66.0 | 59.5 |
| Efficiency of $w_2$ | 44.0 | 65.5 | 69.5 | 69.0 | 65.5 |

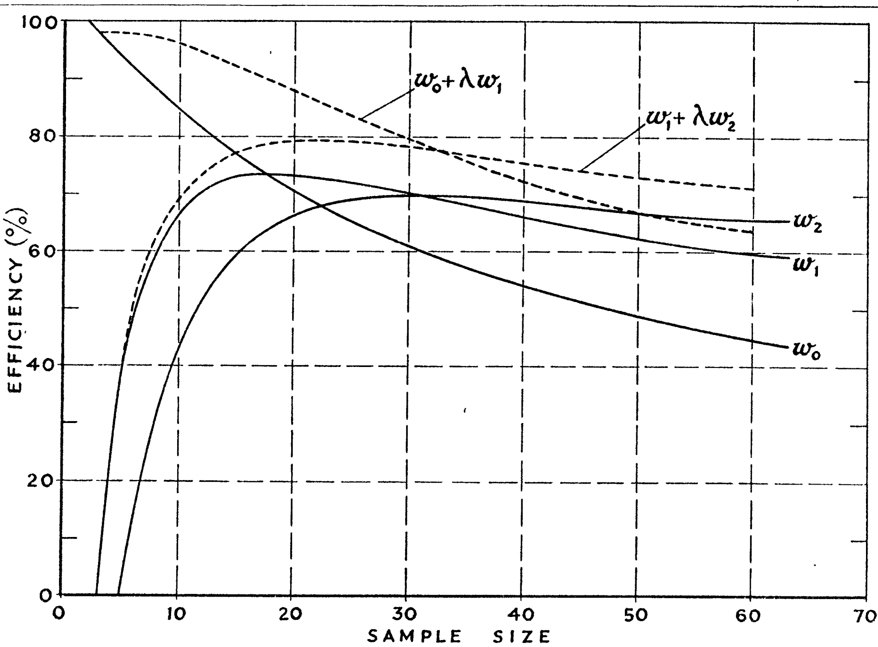


FIG. 1. Efficiencies of the various estimators considered

The general behaviour is illustrated by Fig. 1. It appears that, as far as efficiency is concerned, range is best up to samples of 17. Then the first quasi-range is optimum up to $n = 31$. It seems likely that $w_3$ becomes optimum round about sample size 50. In any case, the difference between $w_2$ and $w_3$ will probably be small over quite a wide range of sample sizes.

**8. Table of values for first quasi-range.** The p.d.f. was computed, using three terms of the series, for even values of $n$ from 10 to 20, and also for 25 and 30. Values of moment constants and percentage points were found for these sample sizes. Values for other sizes in the range 10 to 30 were then found by interpolation. As a check, mean values for each sample size were evaluated using (6).

Throughout this work, as in all the earlier quadratures, ordinates were found at intervals of 0.2 in $w_r$. This spacing was close enough to ensure that the difference of odd and even ordinate sums was of the order of one part in a hundred thousand. This proved a valuable check on the computation of p.d.f. values. It was found to fail only for sample sizes below 10. In such cases there is no longer a high enough degree of contact between the p.d.f. and the axis at the origin to secure this balancing of odd and even sums.

TABLE I

*Constants for first quasi-range*

| $n$ | Moments | | | | Percentage Points | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Var. | $\beta_1$ | $\beta_2$ | 0.1 | 1.0 | 2.5 | 5.0 | 95.0 | 97.5 | 99.0 | 99.9 |
| 10 | 2.0027 | 0.3423 | 0.145 | 3.125 | 0.57 | 0.83 | 0.97 | 1.11 | 3.03 | 3.25 | 3.52 | 4.09 |
| 11 | 2.1238 | 0.3362 | 0.131 | 3.119 | 0.67 | 0.95 | 1.09 | 1.23 | 3.14 | 3.36 | 3.62 | 4.18 |
| 12 | 2.2315 | 0.3300 | 0.120 | 3.117 | 0.77 | 1.05 | 1.21 | 1.35 | 3.23 | 3.45 | 3.70 | 4.27 |
| 13 | 2.3282 | 0.3240 | 0.112 | 3.117 | 0.87 | 1.15 | 1.31 | 1.45 | 3.32 | 3.53 | 3.78 | 4.34 |
| 14 | 2.4158 | 0.3183 | 0.106 | 3.118 | 0.96 | 1.25 | 1.40 | 1.54 | 3.39 | 3.60 | 3.86 | 4.41 |
| 15 | 2.4959 | 0.3128 | 0.102 | 3.119 | 1.04 | 1.33 | 1.49 | 1.63 | 3.46 | 3.67 | 3.92 | 4.47 |
| 16 | 2.5695 | 0.3076 | 0.098 | 3.120 | 1.12 | 1.41 | 1.57 | 1.71 | 3.53 | 3.74 | 3.98 | 4.52 |
| 17 | 2.6376 | 0.3028 | 0.096 | 3.122 | 1.19 | 1.49 | 1.64 | 1.78 | 3.59 | 3.79 | 4.04 | 4.57 |
| 18 | 2.7008 | 0.2982 | 0.094 | 3.124 | 1.26 | 1.56 | 1.71 | 1.85 | 3.64 | 3.85 | 4.09 | 4.62 |
| 19 | 2.7599 | 0.2938 | 0.092 | 3.125 | 1.32 | 1.62 | 1.78 | 1.92 | 3.70 | 3.90 | 4.14 | 4.66 |
| 20 | 2.8152 | 0.2898 | 0.091 | 3.126 | 1.39 | 1.68 | 1.84 | 1.98 | 3.74 | 3.94 | 4.18 | 4.71 |
| 21 | 2.8672 | 0.2859 | 0.090 | 3.128 | 1.44 | 1.74 | 1.90 | 2.04 | 3.79 | 3.99 | 4.23 | 4.75 |
| 22 | 2.9163 | 0.2823 | 0.089 | 3.130 | 1.50 | 1.80 | 1.95 | 2.09 | 3.83 | 4.03 | 4.27 | 4.78 |
| 23 | 2.9627 | 0.2788 | 0.088 | 3.132 | 1.55 | 1.85 | 2.00 | 2.14 | 3.87 | 4.07 | 4.30 | 4.82 |
| 24 | 3.0067 | 0.2754 | 0.088 | 3.134 | 1.60 | 1.90 | 2.05 | 2.19 | 3.91 | 4.11 | 4.34 | 4.85 |
| 25 | 3.0486 | 0.2723 | 0.087 | 3.135 | 1.65 | 1.95 | 2.10 | 2.24 | 3.95 | 4.14 | 4.37 | 4.88 |
| 26 | 3.0885 | 0.2694 | 0.087 | 3.137 | 1.70 | 2.00 | 2.14 | 2.28 | 3.98 | 4.18 | 4.41 | 4.91 |
| 27 | 3.1265 | 0.2665 | 0.087 | 3.138 | 1.74 | 2.04 | 2.19 | 2.32 | 4.02 | 4.21 | 4.44 | 4.94 |
| 28 | 3.1629 | 0.2638 | 0.087 | 3.140 | 1.78 | 2.08 | 2.23 | 2.36 | 4.05 | 4.24 | 4.47 | 4.97 |
| 29 | 3.1978 | 0.2612 | 0.087 | 3.141 | 1.82 | 2.12 | 2.27 | 2.40 | 4.08 | 4.27 | 4.50 | 5.00 |
| 30 | 3.2312 | 0.2588 | 0.086 | 3.142 | 1.86 | 2.16 | 2.31 | 2.44 | 4.11 | 4.30 | 4.52 | 5.02 |

All values were computed to one further place than that given in Table I. Errors should not exceed one unit in the last place quoted, except for variance from 20 to 30. Here the wider interval of interpolation may result in errors up to 1.5 units in the last figure.

We can find constants $c$, $\lambda$, $\nu$ so that $cw_1^\lambda$ has approximately a $\chi^2$ distribution with $\nu$ degrees of freedom. Errors do not vary greatly over the range $n = 10$ to $n = 30$. Thus we find for $n = 30$,

$$c = 16.231 \qquad \lambda = 1.1409 \qquad \nu = 62.$$

In this case the maximum error in the probability integral deduced from that of $\chi^2$ is 0.0004, occurring near the 20 per cent point; errors are smaller at the tails. From the 5 per cent point to the 97.5 per cent point errors in percentage points of $w_1$, deduced from the corresponding values for $\chi^2$, are always less than 0.001. The error at the 0.1 per cent point is 0.008, while that at the 99.9 per cent point is 0.005.

This type of transformation enables Bartlett's test to be used as an approximate test of homogeneity of a set of values of $w_1$, each for the same sample size. It has been found to give results of similar accuracy for range and average range. It is hoped that a detailed study of this transformation will be completed and published shortly.

**9. Linear combinations of quasi-ranges.** Godwin [5] determines the optimum linear combination of $w_0$, $w_1$, $w_2$, $\cdots$, for the estimation of standard deviation. Such an estimator uses all the possible quasi-ranges, and for $n = 10$ gives an efficiency of 99.0 per cent.

For rapid estimation, attention must be restricted to a few values of $w_r$. Mosteller [10] considers certain unweighted sums of two values. His investigation is restricted to large samples, where the $w_r$ are replaced by inter-quantile distances. Nair [11] considers the sum of the first $k$ quasi-ranges for sample sizes up to 10, while Jones [9] investigates it in the large sample case.

Using the methods of Section 3, we can derive approximations to the covariances of pairs of $w_r$ values. Thus, for $w_0$ and $w_1$ we find that the quadruple integral for $E(w_0 w_1)$ is approximately equal to:

$$K \int_{-\infty}^{\infty} w_1 \varphi^3(\tfrac{1}{2} - \Phi)(2\Phi)^{n-4}\, k\, dw_1$$

where

$$K = 2\sqrt{2\pi}\, n(n-1)(n-2)(n-3)$$

and

$$\frac{1}{k^2} = 3 - \psi' - \tfrac{1}{4}w_1^2 - (n-4)\theta'.$$

Here, all Greek letters are functions of $\frac{1}{2}w_1$. When $n = 10$, comparison with Godwin's values [4] shows that the error in the covariance is 0.0028. The use of another term of the series reduces this error to 0.0009 units.

In deriving these errors, the exact value of $K$ given above was not used. Instead, the integral was first evaluated with the factor $w_1$ omitted. This is an approximation to $E(w_0)$, and $K$ can be found to satisfy this condition exactly. This device is analogous to the determination of $C$ in (5) so as to make the area under the p.d.f. equal to unity. As in the previous case it gives appreciably better results than does the use of the theoretically correct constant.

For $w_1$ and $w_2$ we have:

$$E(w_1 w_2) = L \int_{-\infty}^{\infty} w_2 \varphi^2 (\tfrac{1}{2} - \Phi)^4 [\psi - \tfrac{1}{2}h(w_2)](2\Phi)^{n-6} k \, dw_2. \tag{12}$$

where

$$h(w_2) = \frac{\dfrac{1}{2} - \Phi\left(\dfrac{w_2}{\sqrt{2}}\right)}{\sqrt{\pi}\left[\dfrac{1}{2} - \Phi\left(\dfrac{w_2}{2}\right)\right]^2} \tag{13}$$

and

$$\frac{1}{k^2} = 2 - (n - 6)\theta' - \frac{\psi\psi' + \psi'' + 2\psi^3 + (\psi' - \psi^2)h(w_2)}{\psi - \frac{1}{2}h(w_2)}. \tag{14}$$

All functions without an argument shown are to be taken with argument $\frac{1}{2}w_2$. Comparison with Godwin's value shows that this expression yields $\operatorname{cov}(w_1, w_2)$ with an error of 0.0005 units, when $n = 10$.

The two estimators:

$$\text{(a)} \quad \frac{w_0 + \lambda w_1}{E(w_0) + \lambda E(w_1)} \qquad \text{and} \qquad \text{(b)} \quad \frac{w_1 + \lambda w_2}{E(w_1) + \lambda E(w_2)}$$

were considered, the constant $\lambda$ being chosen to maximise efficiency. Results are shown by the dotted curves in the figure. Values, expressed to the nearest 0.5 per cent, are given below.

| $n$ | 10 | 20 | 30 | 40 | 60 |
|---|---|---|---|---|---|
| (a) | | | | | |
| Efficiency | 96.5 | 87.0 | 79.5 | 72.5 | 63.5 |
| $\lambda$ | 0.85 | 1.47 | 1.93 | 2.30 | 2.95 |
| (b) | | | | | |
| Efficiency | 69.0 | 79.5 | 78.5 | 76.0 | 71.0 |
| $\lambda$ | 0.36 | 0.81 | 1.13 | 1.40 | 1.76 |

As is to be expected, the value of $\lambda$ used is not at all critical. Thus, a convenient integer can be used with little loss of efficiency. For instance, using the estimator (a) with $\lambda = 2$ when $n = 40$, results in a loss of 1.1 per cent.

It is evident from the figure that the use of both $w_0$ and $w_1$ offers an appreciable advantage for sample sizes from 10 to 30. The combination of $w_1$ and $w_2$ is not so impressive. It seems likely that $w_2$ used with a higher order quasi-range might be better from 30 to 60.

The p.d.f. of such estimates will be a trivariate integral. While the methods used above allow this to be replaced approximately by a single integral, the labour of evaluation for a set of values would still be considerable. However, for some purposes, confidence limits based on a normal approximation will be satisfactory.

Alternatively, the approximate evaluation of $\beta_1$ is possible. Thus, if the analogy with $w_0$ and $w_1$ can be relied upon, an approximation based on a power of $\chi^2$ should give a degree of accuracy sufficient for most purposes.

I should like to thank Mr. D. F. Mills, who carried out the computations necessary for the tabled values of constants for first quasi-range.

## REFERENCES

[1] J. H. Cadwell, "The distribution of quantiles of small samples," *Biometrika*, Vol. 39 (1952), pp. 207–211.

[2] D. R. Cox, "A note on the asymptotic distribution of range," *Biometrika*, Vol. 35 (1948), pp. 310–315.

[3] G. Elfving, "The asymptotic distribution of range in samples from a normal population," *Biometrika*, Vol. 34 (1947), pp. 111–119.

[4] H. J. Godwin, "Some low moments of order statistics," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 279–85.

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

[6] E. J. Gumbel, "The distribution of the range," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 384–412.

[7] H. O. Hartley and E. S. Pearson, "Moment constants for the distribution of range in normal samples," *Biometrika*, Vol. 38 (1951), pp. 463–464.

[8] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, 2nd ed., Cambridge University Press 1950, pp. 501–503.

[9] A. E. Jones, "A useful method for the routine estimation of dispersion from large samples," *Biometrika*, Vol. 33 (1946), pp. 274–282.

[10] F. Mosteller, "On some useful inefficient statistics," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 377–408.

[11] K. R. Nair, "Efficiencies of certain linear systematic statistics for estimating dispersion from normal samples," *Biometrika*, Vol. 37 (1950), pp. 182–183.

[12] K. Pearson, "On the probable errors of frequency constants," Pt. III, *Biometrika*, Vol. 13 (1920), pp. 113–132.

[13] K. Pearson, *Tables for Statisticians and Biometricians*, Part II, Biometrika Office, University College, London, 1931, p. cxvii and p. 165.