THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

ON THE DISTRIBUTIONS OF THE QUASI-RANGE AND MIDRANGE FOR SAMPLES FROM A NORMAL POPULATION

by

G. M. Jones, C. H. Kapadia, D. B. Owen and R. P. Bland

Technical Report No. 20
Department of Statistics THEMIS Contract

Department of Statistics
Southern Methodist University

THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

ON THE DISTRIBUTIONS OF THE QUASI-RANGE AND MIDRANGE FOR SAMPLES FROM A NORMAL POPULATION

by

G. M. Jones, C. H. Kapadia, D. B. Owen and R. P. Bland

Technical Report No. 20
Department of Statistics THEMIS Contract

October 11, 1968

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

On the Distributions of the Quasi-Range and Midrange for Samples from a Normal Population

by

G. M. Jones*, C. H. Kapadia**†, D. B. Owen† and R. P. Bland**†

Southern Methodist University

ABSTRACT

Let $X_1, X_2, \cdots, X_n$ be an ordered sample of size $n(X_1 \le X_2 \le \cdots \le X_n)$ from the standard normal population. If the r largest and r smallest observations are omitted, the range of the remaining n-2r sample values is defined as the rth quasi-range, and is usually written as $W_r = X_{n-r} - X_{r+1}$. Expressions involving multivariate normal probabilities are found for the density function, the cumulative distribution function, and the expected value.

The midrange is defined as $M = (X_1 + X_n)/2$. If the sample values X_1 and X_n are from the standard normal population, it is possible to express the density function of the midrange in terms of multivariate normal probabilities for even values of n.

For certain small values of n and r, the distributions referred to above are expressed in terms of G(h), the univariate normal distribution function, T(h,a) [4], and S(h,a,b) [5]. These functions have been tabulated and are used to evaluate univariate, bivariate, and trivariate normal probabilities, respectively.

*Now with LTV Aerospace Corporation.
**Research partially supported by NIH Grant 2 T01 GM00951-06 EBB.
† Also supported by ONR Contract N00014-68-A-0515.

1. Summary and Introduction

Let $X_1, X_2, \cdots, X_n$ be an ordered sample of size $n(X_1 \le X_2 \le \cdots \le X_n)$ from the standard normal population. If the r largest and r smallest observations are omitted, the range of the remaining n-2r sample values is defined as the rth quasi-range, and is usually written as $W_r = X_{n-r} - X_{r+1}$. Expressions involving multivariate normal probabilities are found for the density function, the cumulative distribution function, and the expected value.

The midrange is defined as $M = (X_1 + X_n)/2$. If the sample values X_1 and X_n are from the standard normal population, it is possible to express the density function of the midrange in terms of multivariate normal probabilities for even values of n.

For certain small values of n and r, the distributions referred to above are expressed in terms of G(h), the univariate normal distribution function, T(h,a) [4], and S(h,a,b) [5]. These functions have been tabulated and are used to evaluate univariate, bivariate, and trivariate normal probabilities, respectively.

Bland, et.al. [1] have found the distribution of the range (r=0) in terms of multivariate normal probabilities for all n, and for the special cases where n = 2,3,4, and 5, they have expressed the distribution of the range in terms of the functions mentioned above. Since the range may be thought of as a special case of the quasi-range, this paper represents a natural extension of their results.

2. Quasi-Range

2.1 Probability density function

The density of the quasi-range [2] in integral form is

$$f_{W_r}(w) = \frac{n!}{(r!)^2(n-2r-2)!} \int_{-\infty}^{+\infty} [G(x)]^r \, [G(x+w) - G(x)]^{n-2r-2}$$

$$\cdot [1-G(x+w)]^r \, G'(x) \, G'(x+w)dx \qquad w \geq 0, \qquad (1)$$

where $G(x) = \int_{-\infty}^{x} G'(t)dt$ and $G'(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$.

If the transformation $y=\sqrt{2}\, x + w/\sqrt{2}$ is made and the integrand in equation (1) above is expanded by using the binomial theorem and morever the result given by Das [3] is used, the above equation reduces to

$$f_{W_r}(w) = \frac{n!}{\sqrt{2}(r!)^2(n-2r-2)!} G'(w/\sqrt{2})$$

$$\cdot \sum_{i=0}^{r} \sum_{k=0}^{n-2r-2} (-1)^{n-2r-2+i-k} \binom{r}{i} \binom{n-2r-2}{k}$$

$$\cdot \Pr[V_1 \leq \Delta_{k+i-1} w/\sqrt{6},\ V_2 \leq \Delta_{k+i-2} w/\sqrt{6}, \qquad (2)$$

$$\ldots,\ V_{n-r+i-2} \leq \Delta_{k-n+r+2} w/\sqrt{6} \mid \rho = 1/3],$$

where

$$\Delta_h = \begin{cases} 1, & \text{if } h \geq 0, \\ -1, & \text{if } h < 0, \end{cases}$$

and $V_1, V_2, \ldots, V_{n-r+i-2}$ have a joint multivariate normal distribution with zero means and unit variances and correlations all equal to one-third. If $r = 0$ is substituted in equation (2), the result given for the range by Bland, et. al. [1] follows.

When n = 4 or 5 and r = 1, the density of the quasi-range can be expressed in a closed form.

For example, when n = 4 and r = 1,

$$f_{W_1}(w) = 12\sqrt{2}\,G'(w/\sqrt{2})[1 - G(w/\sqrt{6}) - 2T(w/\sqrt{6}, \sqrt{2})],$$

where $T(h,a) = \int_0^a \frac{G'(h)G'(hx)}{1+x^2}dx$ is related to the bivariate normal distribution and has been tabulated by Owen [4].

When n = 5 and r = 1, the density may be written as

$$f_{W_1}(w) = 60\sqrt{2}\,G'(w/\sqrt{2})\Big[2G(w/\sqrt{6}) - 1 + 2T(w/\sqrt{6}, 1/\sqrt{2})$$

$$+ 2T(w/\sqrt{6}, \sqrt{2}) - 8S(w/\sqrt{6}, \sqrt{2}, \sqrt{3/5})$$

$$- 4S(w/\sqrt{6}, 1/\sqrt{2}, 3\sqrt{3/5})\Big],$$

where

$$S(h,a,b) = \int_{-\infty}^{h} T(as,b)G'(s)ds$$

is related to the trivariate normal distribution and has been tabulated by Steck [5].

2.2 The Distribution Function

The distribution function of the quasi-range is

$$F_{W_r}(w) = \frac{n!}{(r!)^2(n-2r-2)!}\int_0^w \int_{-\infty}^{\infty} [G(x)]^r[G(x+w) - G(x)]^{n-2r-2}$$

$$\cdot [1 - G(x+w)]^r G'(x)G'(x+w)dx,\ w \geq 0.$$

Changing the order of integration and integrating by parts r times yields

$$F_{W_r}(w) = \sum_{k=0}^{r} \frac{n(n-1)\cdots(n-2r+k)}{r!(r-k)!} \int_{-\infty}^{\infty} [1 - G(x+w)]^{r-k} \cdot [G(x+w) - G(x)]^{n-2r+k-1} \cdot [G(x)]^{r} G'(x)dx. \tag{3}$$

The distribution function may also be written as

$$F_{W_r}(w) = \sum_{k=0}^{r} \frac{n(n-1)\cdots(n-2r+k)}{r!(r-k)!} \sum_{i=0}^{r-k} \sum_{j=0}^{n-2r+k-1} (-1)^{n-2r-1-k+k+i} \cdot \binom{r-k}{i}\binom{n-2r+k-1}{j} \Pr[Y_1 \le \delta_{j+i-1} w/\sqrt{2},\ Y_2 \le \delta_{j+i-2} w/\sqrt{2},\ \ldots,\ Y_{n-r+i+k-1} \le \delta_{j-n+r-k+1} w/\sqrt{2} \mid \rho = 1/2], \tag{4}$$

where

$$\delta_h = \begin{cases} 0, & \text{if } h < 0, \\ 1, & \text{otherwise,} \end{cases}$$

and $Y_1, Y_2, \ldots, Y_{n-r-1+k+i}$ have a joint multivariate normal distribution with zero means, unit variances, and correlations all equal to one-half.

The method used in deriving equation (4) above is parallel to the one which is used in deriving the equation (2) from equation (1). If $r = 0$ is substituted in equation (4) the expression is exactly the same as that given by Bland et.al. [1] for the range.

When $n = 4$ and $r = 1$, the distribution function becomes

$$F_{W_1}(w) = 12[G(w/\sqrt{2},\ 1/\sqrt{3}) - 4S(w/\sqrt{2},\ 1/\sqrt{3},\ \sqrt{2})] - 5.$$

2.3 The Expected Value

Cadwell [2] gives the expected value of the quasi-range which can be simplified as:

$$E(W_r) = 2(r+1)\binom{n}{r+1}\int_{-\infty}^{\infty} x[1 - G(x)]^r [G(x)]^{n-r-1}G'(x)dx$$

$$= (r+1)\binom{n}{r+1}\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}\frac{(n-1-k)}{\sqrt{\pi}}$$

$$\cdot \Pr[Z_i \le 0;\ i = 1,2, \cdots, n-k-2 | \rho = 1/3],$$

where $Z_1, Z_2, \cdots, Z_{n-k-2}$ have a joint multivariate normal distribution with zero means and unit variances and correlations all equal to one-third. This follows from the method used previously.

3. The Midrange

The density of the midrange is

$$f_M(m) = 2n(n-1)\int_{-\infty}^{m} [G(2m-x) - G(x)]^{n-2} G'(x)G'(2m-x)dx \qquad \text{for } -\infty < m < \infty.$$

If the transformation $y = \sqrt{2}(x-m)$ is performed, it can be shown that the integrand is an even function with respect to y for n = 2, 4,6, $\cdots$, so that

$$f_M(m) = \frac{n(n-1)}{\sqrt{2}} G'(\sqrt{2}m)\int_{-\infty}^{\infty} [G(m-y/\sqrt{2}) - G(m+y/\sqrt{2})^{n-2}G'(y)dy \qquad (n = 2, 4, 6, \cdots,).$$

Again using the same type of procedures as for the quasi-range, the density of the midrange can be expressed as follows:

$$f_M(m) = \frac{n(n-1)}{\sqrt{2}} G'(\sqrt{2}\, m) \sum_{k=0}^{n-2} \sum_{i=0}^{n-2-k} (-1)^{k+i} \binom{n-2}{k} \binom{n-2-k}{i}$$

$$\cdot \Pr[V_1 \leq \Delta_{k-1}\sqrt{2/3}\, m,\ V_2 \leq \Delta_{k-2}\sqrt{2/3} m,\ \cdots,$$

$$V_{k+i} \leq \Delta_{-i}\sqrt{2/3} m \,|\, \rho = 1/3] \quad (n = 2,4,6, \cdots),$$

where Δ_h is defined as before and $V_1, V_2, \cdots, V_{k+i}$ have a joint multivariate normal distribution with zero means and unit variances and correlations all equal to one-third.

When n = 4 the density becomes

$$f_M(m) = 24\sqrt{2}\, G'(\sqrt{2}\, m) \;[T(\sqrt{2/3}\, m,\ \sqrt{2}) - T(\sqrt{2/3}\, m,\ 1/\sqrt{2})],$$

which may be integrated to give the distribution function

$$F_M(m) = 24[S(\sqrt{2}\, m,\ 1/\sqrt{3},\ \sqrt{2}) - S(\sqrt{2}\, m,\ 1/\sqrt{3},\ 1/\sqrt{2})].$$

The density of the midrange for samples of size 3 can also be obtained in a closed form, but other procedures must be used. The integral form is

$$f_M(m) = 6\sqrt{2}\, G'(\sqrt{2}\, m) \int_{-\infty}^{0} [G(m-y/\sqrt{2}) - G(m+y/\sqrt{2})]\, G'(y)dy$$

which reduces to

$$f_M(m) = 12\sqrt{2}\, G'(\sqrt{2}\, m)\, T(\sqrt{2/3}\, m,\ 1/\sqrt{2}).$$

The distribution function becomes

$$F_M(m) = 12\, S(\sqrt{2}\, m,\ 1/\sqrt{3},\ 1/\sqrt{2}).$$

LIST OF REFERENCES

1. Bland, R. P., Gilbert, R. D., Kapadia, C. H., and Owen, D. B., "On the Distributions of the Range and Mean Range for Samples from a Normal Distribution," Biometrika, 53, (1966), 245-248.

2. Cadwell, J. H., "The Distribution of Quasi-Ranges from a Normal Population," Annals of Mathematical Statistics, 24, (1953), 603-613.

3. Das, S. C., "The Numerical Evaluation of a Class of Integrals, II," Proceedings of the Cambridge Philosophical Society, 52, (1956), 442-448.

4. Owen, D. B., "Tables for Computing Bivariate Normal Probabilities," Annals of Mathematical Statistics, 27, (1956), 1075-1090.

5. Steck, G. P., "A Table for Computing Trivariate Normal Probabilities," Annals of Mathematical Statistics, 29, (1958), 780-800.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

ORIGINATING ACTIVITY (Corporate author) SOUTHERN METHODIST UNIVERSITY	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP UNCLASSIFIED

REPORT TITLE

On the Distributions of the Quasi-Range and Midrange for Samples from a Normal Population

DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

AUTHOR(S) (First name, middle initial, last name)

G. M. Jones, C. H. Kapadia, D. B. Owen and R. P. Bland

REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 11, 1968	8	5

CONTRACT OR GRANT NO. N00014-68-A-0515 PROJECT NO. NR 042-260 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 20. 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

DISTRIBUTION STATEMENT

No limitations

1. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research

3. ABSTRACT

Let $X_1, X_2, \cdots, X_n$ be an ordered sample of size $n(X_1 \le X_2 \le \cdots \le X_n)$ from the standard normal population. If the r largest and r smallest observations are omitted, the range of the remaining n-2r sample values is defined as the rth quasi-range, and is usually written as $W_r = X_{n-r} - X_{r+1}$. Expressions involving multivariate normal probabilities are found for the density function, the cumulative distribution function, and the expected value.

The midrange is defined as $M = (X_1 + X_n)/2$. If the sample values X_1 and X_n are from the standard normal population, it is possible to express the density function of the midrange in terms of multivariate normal probabilities for even values of n.

For certain small values of n and r, the distributions referred to above are expressed in terms of G(h), the univariate normal distribution function, T(h,a) [*], and S(h,a,b) [†]. These functions have been tabulated and are used to evaluate univariate, bivariate, and trivariate normal probabilities, respectively.

[*] Owen, D. B., "Tables for Computing Bivariate Normal Probabilities," Annals of Mathematical Statistics, 27, (1956), 1075-1090.

[†] Steck, G. P., "A Table for Computing Trivariate Normal Probabilities," Annals of Mathematical Statistics, 29, (1958), 780-800.