CHAPTER 13

Normal Distributions

1 DEFINITION AND TABLES

A random variable X is **normally** distributed if it has the probability density function:

$$\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \xi}{\sigma} \right)^2 \right], \qquad \sigma > 0. \tag{13.1}$$

The probability density function of $U = (X - \xi)/\sigma$ is

$$p_U(u) = (\sqrt{2\pi})^{-1} \exp\left(-\tfrac{1}{2}u^2 \right), \tag{13.2}$$

which does not depend on the parameters ξ, a. This is called the **standard form** of normal distribution. (It is also the **standardized form.**) The random variable U is called a **standard,** or **unit,** normal variable.

Since

$$\Pr[X \le x] = \Pr\left[U \le \frac{x - \xi}{\sigma} \right], \tag{13.3}$$

such probabilities can be evaluated from tables of the cumulative distribution function of U, which is

$$\Phi(u) = \Pr[U \le u] = (\sqrt{2\pi})^{-1} \int_{-\infty}^{u} e^{-x^2/2} \, dx. \tag{13.4}$$

The notation $\Phi(\cdot)$ is widely used, so it will be used in this book. Further it is convenient to have a systematic notation for the quantiles of the distribution of U. We use the system defined by

$$\Phi(U_\alpha) = \alpha$$

so that $U_{1-\alpha}$ is the upper $100\alpha\%$ point, and U_α $(= -U_{1-\alpha})$ is the lower $100\alpha\%$ point of the distribution.

There are other forms of notation that are much less frequently encountered in statistical work. The parameter in (13.1) is sometimes replaced by the *precision* modulus

$$h = \left(\sigma\sqrt{2}\right)^{-1}.$$

Other functions are

$$\text{erf}(x) = 2\pi^{-1/2} \int_0^x e^{-t^2} dt = 2\Phi(x\sqrt{2}) - 1,$$

$$\text{erfc}(x) = 1 - \text{erf}(x)$$

(13.5)

(erf is the error function, or Cramp function, and erfc the error function complement). Other names for the distribution are second law of *Laplace,* *Laplace,* Gaussian, Laplace-Gauss, de Moivre; $\Phi(\cdot)$ is also called the Laplace-Gauss integral, or simply the probability integral, and $\text{erf}(\cdot)$ is also known by this last name and is sometimes called the error integral.

Tables relating to the unit normal distribution are a necessary ingredient of any textbook in statistical theory or its applications. This is because for many decades the normal distribution held a central position in statistics. As pointed out, tables of the unit normal distribution suffice for calculations relating to all normal distributions. Some care is necessary in using these tables; for example, putting

$$\left(\sqrt{2\pi}\right)^{-1} e^{-x^2/2} = Z(x),$$

(13.6)

it is necessary to remember the multiplier σ^{-1} in

$$\left(\sqrt{2\pi}\,\sigma\right)^{-1} \exp\left[-\frac{1}{2}\left\{\frac{x-\xi}{\sigma}\right\}^2\right] = \sigma^{-1} Z\!\left(\frac{x-\xi}{\sigma}\right).$$

(13.7)

But no real difficulties are presented by the extended use of tables of the unit normal distribution. [The symbols $\varphi(x), \phi(x)$ are often used in place of $Z(x)$.]

In most of the tables only positive values of the variable are given. This is all that is necessary, since

$$Z(x) = Z(-x) \quad \text{and} \quad \Phi(x) = 1 - \Phi(-x).$$

(13.8)

Here we give a list of only the more easily available tables. Fuller lists are given in the National Bureau of Standards (1952) (up to 1952) and **Green-**

wood and **Hartley** *(1962)* (up to *1958).* The functions most often tabulated are $\Phi(x)$, $Z(x)$, and U_α, but there are many variants for special uses.

Pearson and **Hartley** *(1948)* give tables based on values originally computed by Sheppard *(1903, 1907).* These contain

1. $\Phi(x)$ and $Z(x)$ to *7* decimal places for $x = 0.00(0.01)4.50$; and to *10* decimal places for $x = 4.50(0.01)6.00$.
2. U_α to *4* decimal places for $a = 0.501(0.001)0.980(0.0001)0.9999$.
3. $Z(U_\alpha)$ to *5* decimal places for $a = 0.500(0.001)0.999$.

Pearson and **Hartley** *(1970)* have also provided these tables.

Fisher and Yates *(1963)* give U_α to *6* decimal places for $a = 0.505(0.005)0.995,$ and to *5* decimal places for $1 - a = 0.0'1$ [$r = 2(1)8$]. These tables include values of "probits" $-(5 + U_\alpha)$—to *4* decimal places for $a = 0.001(0.001)0.980(0.0001)0.9999$ and of $Z(u)$ to 4 decimal places for $u = 0.00(0.01)3.00(0.1)3.9.$

Owen *(1962)* gives $Z(x)$ and $\Phi(x)$ to *6* decimal places, $Z^{(1)}(x)$, $Z^{(2)}(x)$, $Z^{(3)}(x)$, and $\{1 - \Phi(x)\}/Z(x)$ to *5* decimal places, and $\Phi(x)/Z(x)$ to 4 decimal places, for $x = 0.00(0.01)3.99$; also $\{1 - \Phi(x)\}$ to *5* significance figures for $x = 3.0(0.1)6.0(0.2)10.0(1)20(10)100(25)200(50)500$, and U_α and $Z(U_\alpha)$ to *5* decimal places for $a = 0.500(0.001)0.900(0.005)0.990.$

Kelley *(1948)* gives U_α to 8 decimal places for $a = 0.5000(0.0001)0.9999$. Hald *(1952)* gives $Z(x)$ and $\Phi(x)$ to *4* significant figures for

$$x = \pm 0.00(0.01)4.99,$$

and **probits** *(5 + U_α)* to *3* decimal places for

$$\alpha = 0.0001(0.0001)0.0250(0.001)0.9750(0.0001)0.9999.$$

We next describe some tables containing larger number of decimal places, useful for special calculations.

In Zelen and Severo *(1964)* there are tables of $Z(x)$, $\Phi(x)$ and $Z^{(1)}(x)$ to *15* decimal places, $Z^{(2)}(x)$ to *10,* and $Z^{(r)}(x)$ $(r = 3, 4, 5, 6)$ to *8* decimal places for $x = 0.00(0.02)3.00$. For the values $x = 3.00(0.05)5.00,$ $\Phi(x)$ is given to *10* decimal places, $Z(x)$ to *10* significant figures, and $Z^{(r)}(x)$ $(r = 2, \ldots, 6)$ to *8* significant figures. *A* further table gives $Z^{(r)}(x)$ $(r = 7, \ldots, 12)$ to *8* significant figures for $x = 0.0(0.1)5.0$. There are also tables [based on Kelley **(1948)**] of U_α and $Z(U_\alpha)$ to *5* decimal places for $a = 0.500(0.001)0.99$, and of U_α to *5* decimal places for $a = 0.9750(0.0001)0.9999.$

In the National Bureau of Standards tables *(1953)* there are given tables of $Z(x)$ and $2\Phi(x) - 1$ [$= \text{erf}(x/\sqrt{2})$] to *15* decimal places for $x = 0(0.0001)1.0000(0.001)7.800$; and also of $2[1 - \Phi(x)]$ to *7* significant figures for $x = 6.00(0.01)10.00.$

Table 13.1   Percentile Points of Normal Distribution, as Standardized Deviates (Values of U_α)

a	U_α
0.5	0.000000
0.6	0.253347
0.7	0.524401
0.75	0.674490[a]
0.8	0.841621
0.9	1.281552
0.95	1.644854
0.975	1.959964
0.99	2.326348
0.995	2.575829
0.9975	2.807034
0.999	3.090232

"The value of $U_{0.75}$ ($= 0.6745$), the upper quartile of the unit normal distribution, is occasionally called the *probable error* of the distribution, though this nomenclature is seldom used at present. The probable error of distribution (1) is, of course, $U_{0.75}\sigma$.

There are many other publications containing various forms of tables of the normal distribution. Further tables of special functions associated with the normal distribution are used in connection with **probit** analysis. There is no need for extensive tables of the normal distribution to be given here. We confine ourselves, in Table 13.1, to a few commonly used values of U_α.

Tables of random unit normal deviates (representing values of a random variable having a unit normal distribution) have been constructed from tables of random numbers (representing values of a random variable having a discrete rectangular distribution over the integers 0–9). In 1948 Wold (1948) published a set of 25,000 random unit normal deviates (to 3 decimal places), based on Kendall and Babington Smith's (1942) tables of random numbers. A set of 10,400 random unit normal deviates (also to 3 decimal places), based on Tippett's (1927) table of random numbers, was published by Sengupta and Bhattacharya (1958). These replaced an earlier set of tables, first appearing in 1936 [Mahalanobis et al. (1934)] which were found to contain a number of errors.

A set of 100,000 random unit normal deviates, to 3 decimal places, based on the first half-million random numbers produced in 1947, was published by RAND (1955). In Buslenko et al. (1966) there is a table of 1000 random unit normal deviates, to 4 decimal places. These were calculated from the values of five independent random variables R,, ..., R_5 each randomly distributed

over the range 0 to 1 (see Chapter 26), using the formulas

$$U = X - 0.01(3X - X^3),$$

where

$$X = \frac{1}{\sqrt{5}} \sum_{j=1}^{5} \sqrt{3}\,(2R_j - 1).$$

[This formula was suggested by **Bol'shev (1959)**. Note that $\sqrt{3}\,(2R_j - 1)$ has a *standardized* rectangular distribution.]

2 HISTORICAL REMARKS

Because of the importance of the normal distribution, considerable attention has been paid to its historical development. The earliest workers regarded the distribution only as a convenient approximation to the binomial distribution. At the beginning of the nineteenth century appreciation of its broader theoretical importance spread with the work of **Laplace** and Gauss. The normal distribution became widely and uncritically accepted as the basis of much practical statistical work, particularly in astronomy. Around the beginning of the present century, a more critical spirit developed with more attention being paid to systems of "skew (nonnormal) frequency curves" (see Chapter 12). This critical spirit has persisted, but it is offset by developments in both theory and practice. The normal distribution has a unique position in probability theory, and it can be used as an approximation to other distributions. In practice, "normal theory" can frequently be applied, with small risk of serious error, when substantially nonnormal distributions correspond more closely to observed values. This allows us to take advantage of the elegant nature and extensive supporting numerical tables of normal theory.

The earliest published derivation of the normal distribution (as an approximation to a binomial distribution) seems to be that in a pamphlet of de Moivre dated 12 November 1733. This pamphlet was in Latin; in 1738 de Moivre published an English translation, with some additions. [See also Archibald (1926) and Daw (1966).]

In 1774 **Laplace** obtained the normal distribution as an approximation to hypergeometric distribution, and four years later he advocated tabulation of the probability integral [$\Phi(x)$, in our notation]. The work of Gauss in 1809 and 1816 established techniques based on the normal distribution, which became standard methods used during the nineteenth century.

Most theoretical arguments for the use of the normal distribution are based on forms of *central limit theorems*. These theorems state conditions under which the distribution of standardized sums of random variables tends

to a unit normal distribution as the number of variables in the sum increases, that is, with conditions sufficient to ensure an asymptotic unit normal distribution. Gauss's (1816) derivation of the normal distribution, as the resultant of a large number of additive independent errors, may be regarded as one of the earliest results of this kind.

Formal rigorous mathematical discussion of central limit theorems (for independent random variables) may be said to start with the work of Lyapunov (1900). A useful theorem associated with his name states that if $X_1, X_2, \ldots, X_n$ are independent, identically distributed random variables with finite mean and standard deviation then the distribution of the standardized sum

$$\frac{\sum_{j=1}^{n} X_j - nE[X]}{\sqrt{n\,\mathrm{Var}(X)}}$$

tends to the unit normal distribution as n tends to infinity. Lyapunov also obtained an upper bound for the magnitude of the difference between the cumulative distribution functions of the standardized sum and the unit normal. This upper bound was of the form $Cn^{-1/2}\log n$, where C is a constant depending on the variances and third moments of the X_i's. It has subsequently been considerably improved by Cramér (1928), Berry (1941), Esseen (1942), Zahl (1966), and Zolotarev (1967). For the case when the variables $\{X_i\}$ are identically distributed the upper bound obtained by Zolotarev (1967) is

$$0.82\left(\frac{\nu_3}{\sigma^3}\right)n^{-1/2}$$

where

$$\sigma^2 = \mathrm{Var}(X_i),$$

$$\nu_3 = E\left[\left|X_i - E[X_i]\right|^3\right].$$

This result was an improvement on an earlier result of Wallace (1959) [correcting a result of Berry (1941)]. Zahl (1966) has shown that the upper bound

$$0.65\left(\frac{\nu_3}{\sigma^3}\right)n^{-1/2}$$

can be obtained, provided $\nu_3/\sigma^3 \geq 3/\sqrt{2} = 2.22$.

It can be shown by consideration of particular cases that the upper bound must be at least

$$C\left(\frac{\nu_3}{\sigma^3}\right)n^{-1/2},$$

with

$$C = \frac{\sqrt{13}}{6\sqrt{2\pi}} = 0.40974.$$

Zolotarev (1967) has shown that if the variance and absolute third central moment of X_j are σ_j^2, ν_{3j}, respectively $(j = 1, 2, \ldots, n)$, then an upper bound for the magnitude of the difference between cumulative distribution functions is

$$0.9051\left(\sum_{j=1}^{n} \nu_{3j}\right)\left(\sum_{j=1}^{n} \sigma_j^2\right)^{-3/2}$$

For the general case of independent (but not necessarily identically distributed) variables, Lindeberg (1922) showed that putting $\mathbf{Var}(X_i) = \sigma_i^2$ and

$$\sigma_{(n)}^2 = \sum_{i=1}^{n} \sigma_i^2,$$

if

$$\lim_{n \to \infty} \sigma_{(n)}^{-2} \sum_{i=1}^{n} \left(\Pr\{|X_i - E[X_i]| \geq t\sigma_{(n)}\}\right.$$

$$\left.\times E\left[\{X_i - E[X_i]\}^2 \big| |X_i - E[X_i]| \geq t\sigma_{(n)}\right]\right) = 0$$

for all t > 0, the distribution of the standardized sum

$$\sigma_{(n)}^{-1} \sum_{i=1}^{n} (X_i - E[X_i])$$

tends to the unit normal distribution as n tends to infinity. The necessity of Lindeberg's condition was established by Feller **(1935)**. Since then attention has moved to consideration of conditions under which a limiting normal distribution applies to sums of nonindependent random variables. An account of some such conditions can be found in a book by **Loève** (1963).

A comprehensive account of the central limit theorem and related problems (up to the early 1950s) has been given by Gnedenko and Kolmogorov

(1954). Multidimensional extensions of central limit theorems have been investigated by Bergstrom (1945), Esseen (1958), Sadikova (1966), and Sazanov (1967) among others.

Porter (1985), while discussing the historical details of the concepts of variation and error in Quetelet's statistics, has brought out the vital role that the normal distribution plays in the mathematics of society. Wilf (1988) has commented briefly on the general quest for normality. Read (1985) has provided a fine review of the various important developments on the normal distribution. Stigler (1982) has proposed a new standard for the normal distribution.

As one would expect, there has been a phenomenal development on various aspects of the normal distribution. Consequently several books and monographs have appeared dealing specifically with inference, characterizations, tolerance limits, prediction, goodness-of-fit, and so on. It is therefore neither feasible nor necessary to discuss all these developments in detail. Fortunately there is a handbook prepared by **Patel** and Read (1981) available on the distribution; the second edition of this book is currently under preparation. We are hopeful and confident that this volume will provide a comprehensive treatment to the distribution, and hence we have concentrated on adding only those results that are primarily of distributional nature (rather than specific inferential aspect). We refer the readers to the **above-**mentioned handbook and other **books/monographs** (listed for specific topics) for an elaborate discussion.

3   MOMENTS AND OTHER PROPERTIES

If U has the unit normal distribution, then, since the distribution is symmetrical about $U = 0$,

$$E[U] = 0, \qquad\qquad (13.9)$$

and so

$$\mu_r = \mu'_r = E(U^r) = \left(\sqrt{2\pi}\right)^{-1} \int_{-\infty}^{\infty} x^r e^{-x^2/2}\, dx. \qquad (13.10)$$

If r is odd,

$$\mu_r = 0.$$

If r is even,

$$\mu_r = \left(\sqrt{2/\pi}\right)\int_0^\infty x^r e^{-x^2/2}\,dx$$

$$= \left(\sqrt{2/\pi}\right)2^{(r+1)/2}\int_0^\infty t^{(r-1)/2}e^{-t}\,dt$$

$$= 2^{r/2}\Gamma\left(\tfrac{1}{2}(r+1)\right)/\sqrt{\pi}$$

$$= (r-1)(r-3)\ldots 3.1. \tag{13.11}$$

Hence

$$\text{Var}(U) = \mu_2 = 1,$$

$$\alpha_3(U) = 0,$$

$$\beta_2(U) = \alpha_4(U) = 3.$$

This, as pointed out in Section 1, reveals that the unit normal is also the standardized normal distribution. **If** X has the general normal distribution (13.1), then

$$X = \xi + \sigma U, \tag{13.12}$$

where U is a unit normal variable.

Some normal probability density functions are shown in Figure 13.1. The nine curves shown correspond to all possible combinations of $\xi = -1, 0, 1$ and $a = \tfrac{1}{2}, 1, 2$. The curve in the center represents the unit normal distribution ($\xi = 0$, $a = 1$). The distribution is symmetrical about $X = \xi$; the probability density function has points of inflexion at $X = \xi \pm a$. The distribution is unimodal with mode at $X = \xi$ (which is also the median of the distribution). The modal value of the probability density function is $(\sqrt{2\pi})^{-1} = 0.3979$.

The moment generating function of $X (= \xi + \sigma U)$ is

$$E[e^{tX}] = e^{t\xi + (t^2\sigma^2/2)} \tag{13.13}$$

and the characteristic function is $e^{it\xi - (t^2\sigma^2/2)}$. For all $r > 2$, the cumulants κ_r are zero. This property characterizes normal distributions.
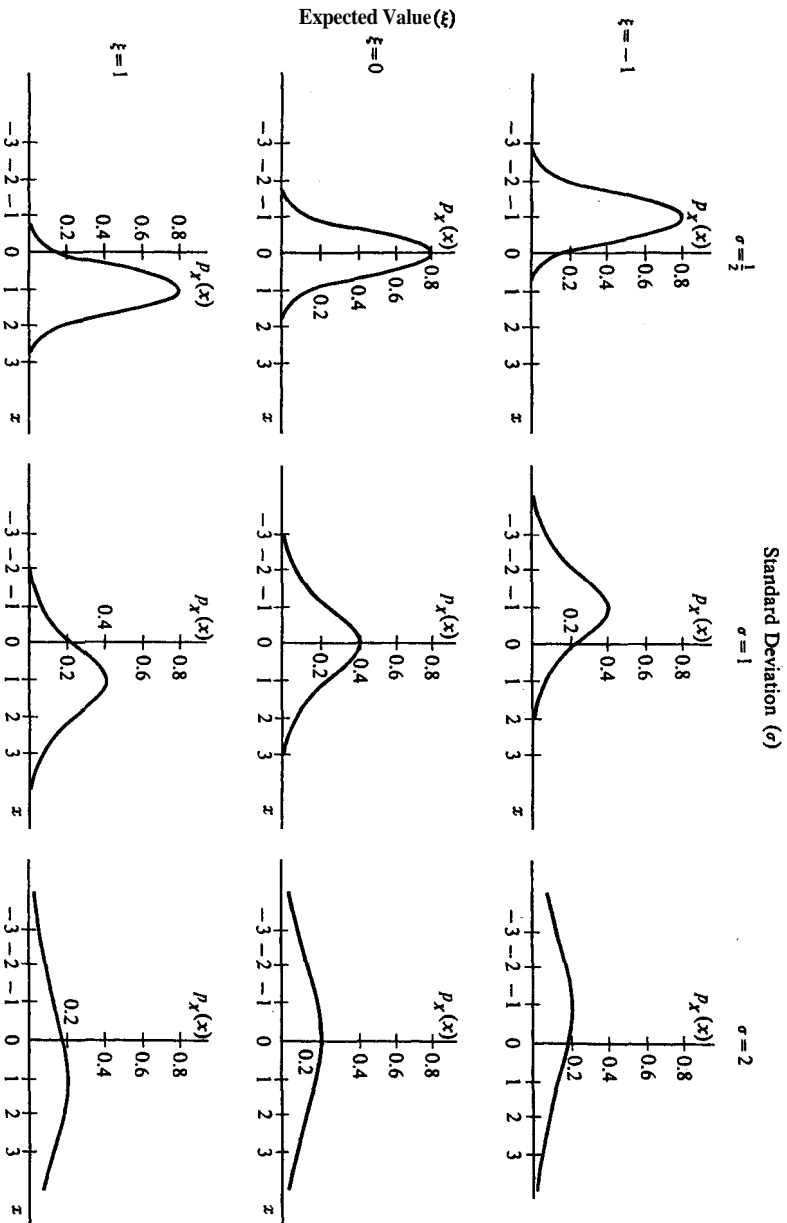
Expected Value (ξ)

Standard Deviation (σ)

$\xi = -1$

$\sigma = \frac{1}{2}$

$\sigma = 1$

$\sigma = 2$

$\xi = 0$

$\xi = 1$

Figure 13.1   Normal Density Functions

The mean deviation of X is $\sigma\sqrt{2/\pi} = 0.798\sigma$. For all normal distributions

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} = \sqrt{\frac{2}{\pi}} = 0.798. \tag{13.14}$$

The information-generating function of X is

$$\left(\sqrt{2\pi}\,\sigma\right)^{-u}\left(\frac{\sqrt{2\pi}\,\sigma}{\sqrt{u}}\right) = \left(\sqrt{2\pi}\,\sigma\right)^{-(u-1)}u^{-1/2}. \tag{13.15}$$

The entropy is

$$\log(\sqrt{2\pi}\,\sigma) + \tfrac{1}{2}. \tag{13.16}$$

It is of some interest to note that the probability density function (13.1) can be expressed in the numerical form

$$0.3979(0.6065)^{[(x-\xi)/\sigma]^2}. \tag{13.17}$$

The derivatives of the function $Z(\cdot)$ are also of some interest. They are used, for example, in the Gram-Charlier expansion (see Chapter 12). We have already discussed them in Chapter 1, and Section 1 of Chapter 12 contains some references to tables of their numerical values.

If $X_1, X_2, \ldots, X_n$ are independent, normally distributed random variables, then any linear function of these variables is also normally distributed. It is of interest to note that if X_1 and X_2 are independent, and each is normally distributed with zero expected value, then $X_1 X_2 (X_1^2 + X_2^2)^{-1/2}$ is also normally distributed. If further $\mathrm{var}(X_1) = \mathrm{var}(X_2)$, then $(X_1^2 - X_2^2)/(X_1^2 + X_2^2)$ is also normally distributed [Shepp (1964)].

If $X_1, X_2, \ldots, X_n$ are independent random variables each distributed as (13.1), then by applying the transformation

$$\begin{cases} X_1 = \bar{X} + (1 \cdot 2)^{-1/2}U_2\sigma + (2 \cdot 3)^{-1/2}U_3\sigma + \cdots + [(n-1)n]^{-1/2}U_n\sigma, \\ X_2 = \bar{X} - (1 \cdot 2)^{-1/2}U_2\sigma + (2 \cdot 3)^{-1/2}U_3\sigma + \cdots + [(n-1)n]^{-1/2}U_n\sigma, \\ X_3 = \bar{X} \qquad\qquad\qquad - 2(2 \cdot 3)^{-1/2}U_3\sigma + \cdots + [(n-1)n]^{-1/2}U_n\sigma, \\ \vdots \\ X_n = \bar{X} \qquad\qquad\qquad\qquad\qquad\qquad - (n-1)[(n-1)n]^{-1/2}U_n\sigma, \end{cases} \tag{13.18}$$

it can be shown that

1. $\bar{X}\ (= n^{-1}\sum_{j=1}^{n}X_j)$ has a normal distribution with expected value ξ and standard deviation $\sigma/\sqrt{n}$.

2, Each U_j $(j = 2,\dots,n)$ is a unit normal variable.

3, $\bar{X}, U_2,\dots, U_n$ are a mutually independent set of variables, and hence

4. $\sum_{j=1}^{n}(X_j - \bar{X})^2 = \sigma^2 \sum_{j=2}^{n} U_j^2$ is distributed as σ^2 (χ^2 with $(n - 1)$ degrees of freedom).

This last result was obtained by Helmert in 1875–76. The transformation (13.18) is called Helmert's transformation.

Since any function

$$g\left(X_1 - \text{x}, \dots, X_n - \bar{X} \right)$$

of the deviations $\{X_j - \bar{X}\}$ alone is a function of $\{U_j\}$ alone, we further note that by (13.18) and by property 3,

5. $\bar{X}$ and any function $g(X_1 - \bar{X},\dots, X_n - \bar{X})$ are mutually independent.

This result is helpful in calculating moments and distributions of statistics such as $\bar{X}[\text{Range}(X_1,\dots, X_n)]^{-1}$; $\bar{X}[n^{-1}\sum_{j=1}^{n}|X_j - \bar{X}|]^{-1}$. It can also be shown that

6. $\sum_{j=1}^{n}(X_j - \bar{X})^2$ and any function of the ratios

$$\left\{ (X_i - \bar{X})\left[\sum_{j=1}^{n} \left(X_j - \bar{X} \right)^2 \right]^{-1/2} \right\}$$

are mutually independent.

Zehna (1991) has recently given a simple proof for the result that $\bar{X}$ and S^2 are statistically independent. Bondesson (1981) discussed a normal sample with given sample mean and variance. Szekely (1985) established the multiplicative infinite divisibility of a standard normal distribution, while Chernoff (1981) presented an inequality involving the normal distribution function. Berg (1988) showed that the distribution of the cube (or, indeed, any odd power) of a normal random variable is not determined by its moments.

Hawkins (1975) has made a comment on the computation of noncentral t and normal integrals. Hawkins and Wixley (1986) have made some observations on the transformation of chi-squared variables to the normal distribution (see also Chapter 18). Aroian, Taneja, and Cornwell (1978) have derived the mathematical forms of the distribution of the product of two normal variables, while Cornwell, Aroian, and Taneja (1977) have discussed the numerical evaluation of that distribution. Hayya, Armstrong, and Gressis (1975) have considered the distribution of the ratio of two normal variables. Karhunen and Narula (1989) have similarly derived the distribution of the ratio of the absolute values of two normal variables. Anscombe and Glynn

and when $n = 4$,

$$E[X_1'X_2'] = E[X_3'X_4'] = \frac{\sqrt{3}}{\pi}, \qquad E[X_2'X_3'] = \frac{\sqrt{3}}{\pi}(2 - \sqrt{3}),$$

$$E[X_1'X_3'] = E[X_2'X_4'] = -\frac{\sqrt{3}}{\pi}(2 - \sqrt{3}), \qquad E[X_1'X_4'] = -\frac{3}{\pi}.$$

A general approach, given by **Godwin (1949a),** is to express the product moments in terms of integrals of the form

$$J_n = \int_0^\infty \cdots \int_0^\infty e^{-Q(x_1,\ldots,x_n)} \, dx_1 \ldots dx_n, \qquad (13.21)$$

where $Q(x_1,\ldots,x_n)$ is a quadratic form in the x_i's. For $n = 1, 2, 3, J_n$ can be expressed explicitly in terms of elementary functions as follows:

$$n = 1, \qquad Q(x_1) = a_{11}x_1^2, \qquad J_1 = \frac{\sqrt{\text{in-}}}{2a_{11}},$$

$$n = 2, \qquad Q(x_1, x_2) = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2,$$

$$J_2 = \frac{1}{\sqrt{\Delta_2}} \left\{ \frac{\pi}{2} - \tan^{-1}\left(\frac{a_{12}}{\sqrt{\Delta_2}}\right) \right\},$$

where

$$\Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}^2;$$

$$n = 3, \qquad Q(x_1, x_2, x_3)$$

$$= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3,$$

$$J_3 = \frac{\sqrt{\pi}}{4\sqrt{\Delta_3}} \left\{ \frac{\pi}{2} + \tan^{-1}\left(\frac{a_{12}a_{13} - a_{11}a_{23}}{\sqrt{a_{11}\Delta_3}}\right) + \tan^{-1}\left(\frac{a_{12}a_{23} - a_{13}a_{22}}{\sqrt{a_{22}\Delta_3}}\right) \right.$$

$$\left. + \tan^{-1}\left(\frac{a_{13}a_{23} - a_{12}a_{33}}{\sqrt{a_{33}\Delta_3}}\right) \right\},$$

where

$$\Delta_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix}$$

$$= a_{11}a_{22}a_{33} - a_{11}a_{23}^2 - a_{22}a_{13}^2 - a_{33}a_{12}^2 + 2a_{12}a_{13}a_{23}.$$

The values of means of order statistics have been tabulated to five decimal places by Harter (1961a) for $n = 2(1)100(25)250(50)400$, and also by Harter (1970) for some more choices of n. The mean and variance of the ith quasi-range have been tabulated by Harter (1959) for n up to 100. Tippett (1925) has computed the expected value of the sample range for n I 1000, while Harter (1960) presented tables of the mean, variance, and the coefficients of skewness and kurtosis for $n \leq 100$. Teichroew (1956) has presented tables of means and product moments of order statistics for sample sizes up to 20. By making use of Teichroew's tables, Sarhan and Greenberg (1962) have tabulated the variances and covariances of order statistics (to 10 decimal places) for $n \leq 20$. These tables have been extended by Tietjen et al. (1977) for sample sizes up to 50. The values of the mean and standard deviation of order statistics prepared by Yamauti (1972) for sample sizes up to 50 are contained in the tables of Tietjen et al. (1977). For the largest order statistic X'_n, Ruben (1954) has tabulated the first ten moments for $n \leq 50$ and Borenius (1965) has presented the first two moments for $n \leq 120$. Parrish (1992a, b) has presented tables of means, variances and covariances of order statistics (up to 25 decimal places) for some selected sample sizes up to 50. Miyakawa, Tamiya, and Kotani (1985a) have discussed the numerical evaluation of moments of order statistics through an orthogonal inverse expansion. Öztürk and Aly (1991) have proposed some simple approximations for the moments of order statistics.

Royston (1982) has given an algorithm for computing the expected values of normal order statistics. This algorithm will compute and present the exact values for sample sizes up to 1000 and will also present an approximate value for the quantity for larger sample sizes [see Koeniger (1983) for an additional remark on this algorithm]. Balakrishnan (1984) has presented an algorithm, based on an orthogonal inverse expansion, to approximate the sum of squares of normal scores, namely $S = \sum_{i=1}^{n}\{E[X'_i]\}^2$. This quantity arises often in nonparametric statistics. Dansie (1986) described the normal order statistics as permutation probability models, while Nelson (1983) discussed the usefulness of normal scores as a transformation.

For the standard normal distribution, an interesting property satisfied by order statistics is that

$$\sum_{j=1}^{n} E[X'_i X'_j] = \sum_{j=1}^{n} \text{Cov}(X'_i, X'_j) = 1, \qquad i \qquad n \qquad . \qquad (13.22)$$

That is, every row (or column) of the product-moment matrix or the variance-covariance matrix adds up to 1. This follows easily from the fact that $\bar{X}$ and $X_i' - \bar{X}$ are statistically independent. Some relationships between moments of order statistics have also been established by various authors, and almost all of them have been derived by exploiting the characterizing differential equation $Z^{(1)}(x) = -xZ(x)$. For example, Govindarajulu (1963) has shown that for $1 \leq i \text{ I } n$,

$$E\left[X_i'^2\right] = 1 + n\binom{n-1}{i-1}\sum_{j=0}^{n-i}(-1)^j\binom{n-i}{j}\frac{1}{i+j}E\left[X_{1:i+j}'X_{2:i+j}'\right]. \quad (13.23)$$

If we set $i = n$ in (13.23), we obtain the relation

$$E\left[X_i'^2\right] = E\left[X_n'^2\right] = 1 + E\left[X_1'X_2'\right]. \quad (13.24)$$

The results in (13.22), (13.23), and (13.24) have been used for checking the computation of the product moments. Furthermore Davis and Stephens (1977, 1978) have applied (13.22) and (13.24) to improve the David-Johnson approximation of the variance-covariance matrix (see Chapter 12). Reference may also be made to Shea and Scallon (1988) for further remarks in this regard.

By noting that the condition $Z^{(1)}(x) = -xZ(x)$ is satisfied by both the standard normal and the half-normal (see Section 10) distributions, Joshi and Balakrishnan (1981) established the following results satisfied by order statistics from both these distributions:

$$\sum_{j=i}^{n}E\left[X_i'X_j'\right] = 1 + \sum_{j=i}^{n}E\left[X_{i-1}'X_j'\right] \qquad \text{for } 1 \leq i \text{ I } n, \quad (13.25)$$

$$\sum_{j=i+1}^{n}E\left[X_i'X_j'\right] = \sum_{j=i+1}^{n}E\left[X_j'^2\right] - (n-i) \qquad \text{for } 1 \leq i \leq n-1, \quad (13.26)$$

$$\sum_{j=1}^{n}E\left[X_i'X_j'\right] = 1 + nE\left[X_{1:1}'\right]E\left[X_{i-1:n-1}'\right] \qquad \text{for } 1 \leq i \leq n, \quad (13.27)$$

and

$$\sum_{j=1}^{n}\text{Cov}\left(X_i', X_j'\right) = 1 - (n-i+1)E\left[X_{1:1}'\right]\{E\left[X_i'\right] - E\left[X_{i-1}'\right]\}$$

$$\text{for } 1 \leq i \leq n. \quad (13.28)$$

Joshi and Balakrishnan (1981) have used these relations to derive a convenient expression for the variance of the selection differential or reach statistic, defined as A, $= \bar{X}_k - \bar{X}$, where $\bar{X}_k$ is the average of the k largest order statistics, $X'_{n-k+1}, \ldots, X'_n$. For example, Joshi and Balakrishnan (1981) have shown that

$$E\left[\bar{X}_k^2\right] = \frac{1}{k^2} \sum_{i=n-k+1}^{n} (2i - 2n + 2k - 1) E\left[X_i'^2\right] - \frac{k-1}{k}. \quad (13.29)$$

Consequently the mean and variance of $\bar{X}_k$ (and hence of A,) can be determined from the first two raw moments of X_i' alone. These quantities have been tabulated for sample sizes up to 50 by Joshi and Balakrishnan (1981). Some properties of $\bar{X}_k$ have also been discussed by Schaeffer, van Vleck, and Velasco (1970) and Burrows (1972, 1975). In particular, they observed that k $\text{Var}(\bar{X}_k)$ remains almost constant for the selected fraction k/n. While Schaeffer, van Vleck, and Velasco (1970) tabulated the values of k $\text{Var}(\bar{X}_k)$ for $n \le 20$ and all choices of k, Burrows (1972, 1975) provided approximations to $E[\bar{X}_k]$ and k $\text{Var}(\bar{X}_k)$ for large values of n. But, Joshi and Balakrishnan (1981) have pointed out that Burrows' approximation for $k \, \text{Var}(\bar{X}_k)$ is not satisfactory for small values of k even when n is 50, and that the approximation improves when k increases. The statistic A, is related to Murphy's test statistic for outliers; for details, see Hawkins (1980) and Barnett and Lewis (1994).

The cumulative distribution function of the extreme X'_n was tabulated by Tippett (1925) for $n = 3, 5, 10, 20, 30, 50, 100(100)1000$. Percentage points for $n \le 30$ were given by Pearson and Hartley (1970), and the cumulative distribution function of X'_n was tabulated by Pearson and Hartley (1972), to 7 decimal places, for $n = 3(1)25(5)60, 100(100)1000$ and for x in steps of 0.1. Gupta (1961) and Govindarajulu and Hubacker (1964) have presented percentage points of all order statistics for $n < 10$ and $n \le 30$, respectively. Eisenhart, Deming, and Martin (1963) have tabulated percentage points of the sample median. Pearson and Hartley (1942, 1970) have provided tables for the distribution function of the sample range $X'_n - X'_1$. Harter and Clemm (1959) have given extensive tables of the cumulative distribution function (to 8 decimal places) and also of the percentage points (to 6 decimal places) of the range for $n = 2(1)20(2)40(10)100$ and the argument in steps of 0.01, and 23 different percentage points for each n, respectively. Besides reproducing these tables, Harter (1970) has presented tables of the probability density function of the sample range (to 8 decimal places) for n up to 16 and the argument in steps of 0.01. The cumulative distribution function of the ith quasi-range, $X'_{n-i+1} - X'_i$, is presented by Harter (1970) to 8 decimal places for $n = 2(1)20(2)40(10)100, i = 1, 2, \ldots, 9$, and the argument in steps of 0.05. Harter (1970) has also presented tables of corresponding percentage

points to 6 decimal places. Currie (1980) has discussed the distribution of the studentized range. An algorithm for calculating the probability integral of the sample range has been presented by **Barnard** (1978); see El Lozy (1982) for some additional remarks on this computational algorithm.

David, Kennedy, and Knight (1977) have provided tables of means, variances, and covariances of order statistics arising from a standard normal sample with one outlier. These values are presented for all n up to 20, for the two cases: (1) location-outlier, the outlier being $N(\lambda, 1)$ for A = 0(0.5)3, 4; and (2) scale-outlier, the outlier being $N(0, \tau^2)$ for $\tau = 0.5, 2, 3, 4$. These tables have since been used in many robustness studies because they can be utilized to determine exactly the bias and the mean square error of any linear estimator when there is a single outlier present in the sample.

In addition to the numerous tables listed here, there are also several tables pertaining tests for outliers. The recently published third edition of **Barnett** and Lewis (1994) provides a complete list of all the available tables.

5 RECORD VALUES

Let $X_{U(1)}, X_{U(2)}, \ldots$ be the upper record values arising from a sequence $\{X_i\}$ of i.i.d. standard normal variables. That is, with $T_0 = 0$ and

$$T_n = \min\{j : j > T_{n-1}, X_j > X_{T_{n-1}}\}$$

denoting the upper record times, the record value sequence $\{X_{U(n)}\}_{n=1}^{\infty}$ is defined by $X_{U(n)} = X_{T_{n-1}}$, n = 1, 2, ... Then the density of $X_{U(n)}$ is

$$p_{X_{U(n)}}(x) = \frac{1}{(n-1)!}\{-\log(1 - \Phi(x))\}^{n-1}Z(x), \qquad -\infty < x < \infty,$$

(13.30)

and the joint density of $X_{U(m)}$ and $X_{U(n)}$ is given by

$$p_{X_{U(m)}, X_{U(n)}}(x, y)$$

$$= \frac{1}{(m-1)!(n-m-1)!}\{-\log(1 - \Phi(x))\}^{m-1}\frac{Z(x)}{1 - \Phi(x)}$$

$$\times\{-\log(1 - \Phi(y)) + \log(1 - \Phi(x))\}^{n-m-1}Z(y),$$

$$-\infty < x < y < \infty, 1 \le m < n. \quad (13.31)$$

From (13.30) and (13.31), Houchens (1984) and Balakrishnan and Chan (1994) have determined (by numerical methods) the values of $E[X_{U(n)}]$, $\text{Var}(X_{U(n)})$, and $\text{Cov}(X_{U(m)}, X_{U(n)})$. By making use of these values, these

authors have also derived the best linear unbiased estimators of ξ and σ based on the first n upper record values. Balakrishnan and Chan (1994) have also discussed the prediction of a future record and a test for spuriosity of a current record value. They have also established that

$$E\left[X_{U(n-1)}X_{U(n)}\right] = E\left[X_{U(n)}^2\right] - 1 \qquad \text{for } n \geq 2, \qquad (13.32)$$

and consequently that

$$\text{Cov}\left(X_{U(n-1)}, X_{U(n)}\right)$$
$$= \text{Var}\left(X_{U(n)}\right) + E\left[X_{U(n)}\right]\left\{E\left[X_{U(n)}\right] - E\left[X_{U(n-1)}\right]\right\} - 1. \quad (13.33)$$

Suppose that $X_{L(1)}, X_{L(2)}, \ldots,$ denote the lower record values arising from a sequence $\{X_i\}$ of i.i.d. standard normal variables; that is, with $T_0^* = 1$ and

$$T_n^* = \min\{j : j > T_{n-1}^*, X_j < X_{T_{n-1}^*}\}$$

denoting the lower record times, the lower record value sequence $\{X_{L(n)}\}_{n=1}^{\infty}$ is defined by $X_{L(n)} = X_{T_{n-1}^*}$, $n = 1, 2, \ldots$. Then, due to the symmetry of the standard normal distribution, it may be easily observed that

$$X_{L(n)} \stackrel{d}{=} -X_{U(n)},$$

$$\left(X_{L(m)}, X_{L(n)}\right) \stackrel{d}{=} \left(-X_{U(m)}, -X_{U(n)}\right).$$

With this property, moments of the lower record values (and inference based on the lower record values) can be easily obtained from the corresponding results for the upper record values.

6   CHARACTERIZATIONS

We first summarize normal characterizations presented in the first edition of this book [Johnson and Kotz (1970)]. In all cases $X_1, \ldots, X_n$ are i.i.d. random variables, unless explicitly stated otherwise.

1. $\bar{X} = n^{-1}\sum_{j=1}^n X_j$ has a normal distribution. [See Chapter 12, Section 6. Janson (1988) expresses this in the form $d^k\phi_X(t)/dt^k|_{t=0} = 0$ for all $k > 2$, where $\phi_X(t)$ is the characteristic function of the common distribution of the X's.]

   A more sophisticated characterization based on properties of $\bar{X}$ is due to Fieger (1971). If the distribution of the X's belongs to a location family, with $F_X(x) = g(x - \theta)$, if $E[|X_i|]$ is finite, and if $\bar{X}$ is the best translation invariant estimator of 8 for any convex loss function

$W(\theta^*, 6) = W(|\theta^* - \theta|)$ with $W(u) \geq W(0)$ for all u, then the common distribution is normal.

2. $\bar{X}$ and $g(\{X_i - X_j\}, i \neq j)$ mutually independent, given one of three conditions:

   a. $g(\cdot) = \sum_{j=1}^{n}(X_j - \bar{X})^2 \ (= nS)$. [Lukacs (1942); weakened to $\bar{X}$ and S which have joint pdf of form $h(\bar{X}, S)S^{n-2}$, with $\partial h/\partial\bar{X}$ and $\partial h/\partial S$ existing, by Kaplansky (1943).]

   b. $g(\cdot)$ is a k-statistic [i.e., $g(\cdot)$ is a polynomial in the X's and $E[g(\cdot)] = \kappa_r$, which is the rth cumulant of the common distribution of the X's for some integer $r > 2$]. [Basu and Laha (1954); Lukacs (1955). See also Geary (1936) and Kawata and Sakamoto (1949).]

   c. $g(\cdot) = 0$ if and only if $X_j = \bar{X}$ for all j; and

$$g(c\underset{\sim}{X}) = |c|g(\underset{\sim}{X})$$

   [Paskevich (1958), Rao (1958), Zinger (1958)].

3. Conditions on conditional expected values:

   a. $E[\bar{X}|\{X_i - X_j\}, i \neq j] = E[\bar{X}]$ (for $n > 2$) [Kagan, Linnik, and Rao (1965, 1973)]. Note that for $n = 3$, this condition is satisfied by any symmetric distribution with finite expected value. Rao (1967) extended this condition to

$$E[X_i|X_j - \bar{X}] = E[\bar{X}] \qquad \text{for all } i \neq j.$$

   b. The common distribution has zero mean and finite variance, and there exist $(n - 1)$ linearly independent statistics $Y_j = \sum_{i=1}^{n} a_{ji} X_i$ $(j = 1, \ldots, n - 1)$ such that

$$E\left[Y_j \sum_i b_i X_i\right] = 0 \qquad \text{for a} \quad j = 1, \ldots, n - 1 \quad (13.34)$$

   and some $\{b_i\}(\{a_{j1}, \ldots, a_{jn}\}$ not all zero). Kagan, Linnik, and Rao (1973, p. 156) improve this result by requiring only $j = 1$ in (13.34) and not finite variance, but restricted to $n \geq 3$. [Cacoullos (1967b)].

   c. $\sum_{j=1}^{n} a_j b_j = 0$, where $a_j \neq 0$ $(j = 1, \ldots, n), \{b_j\}$ are not all zero, and

$$E\left[\left(\sum_{j=1}^{n} b_j X_j\right)^2 \middle| \sum_{j=1}^{n} a_j X_j\right] = E\left[\left(\sum_{j=1}^{n} b_j X_j\right)^2\right]$$

   [Cacoullos (1967a)].

4. Conditions based on identity of distributions:

   a. $\sum_{j=1}^{n} a_j X_j$ $(a_j \neq 0,\ j = 1, \ldots, n)$ and each X_i $(i = 1, \ldots, n)$ for some (a,) [Shimizu (1962), referring to Linnik (1952)].

   b. $\sum_{j=1}^{n} a_j X_j$ and $\sum_{j=1}^{n} b_j X_j$, where $a_j b_j \neq 0$ (for all j), and $\{b_j\}$ is not just a rearrangement of (a,), provided the common distribution has finite moments of all order [Marcinkiewicz (1939)]. [Linnik (1952) showed that this result is not valid if some moments are infinite.]

   c. $(n = 2)$ of X_1 and $(X_1 + X_2)/\sqrt{2}$ [Pólya (1923), referred to by Bryc (1990)].

5. Characterizations based on order statistics [Govindarajulu (1966)]:

   a. Provided the common distribution has a finite variance, the condition

   $$E\left[X_n'^2 - X_{n-1}' X_n' \right] = \sigma^2$$

   ensure that the common distribution is either normal (with variance σ^2) or truncated (from above) normal.

   b. If the expected value of the common distribution is zero, the condition

   $$nE\left[X_j' \overline{X} \right] = \sigma^2, \qquad j = 1, \ldots, n; \quad n \geq 2, \qquad (13.35)$$

   ensures that the common distribution is normal. [Note that if it be assumed that the common distribution has zero probability for negative values, condition (13.35) ensures that it is half-normal (see Section 10).]

6. Conditions based on distributions of specific statistics:

   a. $\sum_{j=1}^{n} X_j^2$ and $\{X_i^2/(\sum_{j=1}^{n} X_j^2)\}$ $(i = 1, \ldots, n)$ are mutually independent [Tamhankar (1967)].

   b. The distribution of $\sum_{j=1}^{n}(X_j + a_j)^2$ depends on the parameters $\{a_j\}$ only through $\sum_{j=1}^{n} a_j^2$ [Kagan and Shalayevskii (1967)].

   c. Provided that $a_j b_j \neq 0$ $(j = 1, \ldots, n)$, $\sum_{j=1}^{n} a_j X_j$ and $\sum_{j=1}^{n} b_j X_j$ are mutually independent for some $\{a_,, b_j\}$. [Darrnois (1951), Skitovich (1953). The result is known as the *Darmois-Skitovich* theorem. Another proof was given by Tranquilli (1968).]

   This condition and the following one do not require assumption of finite variance—or indeed, identity of distributions of $X_1, \ldots, X_n$—though mutual independence is still a necessary property.

   d. Extending condition 1 to cases of not necessarily identically distributed variables, if $(X_1 + X_2)$ has a normal distribution, so do each

of X_1 and X_2. [Cramér (1936), extended to $n > 2$ random variables by Lukacs (1956).]

e. If each X_j is distributed symmetrically about zero, the condition that the statistics

$$Y_j = \sqrt{j-1}\, X_j \left/ \left\{ \sum_{i=1}^{j-1} X_i^2 \right\}^{1/2} \right., \qquad j = 2, \ldots, n,$$

are mutually independent and distributed as t with $j - 1$ degrees of freedom $(j = 2, \ldots, n)$ is necessary and sufficient to ensure that each X_j has the same normal distribution [Kotlarski (1966)].

Books by Kagan, Linnik, and Rao (1973) and Mathai and Pederzoli (1977), as well as a survey by Kotz (1974), provide a useful, and reasonably adequate, basis for more mathematically inclined readers desiring to understand derivations of the above results. Ghurye and Olkin (1973) include a clear exposition of the work of Pólya (1923) and Linnik (1953), and they illustrate the application of Linnik's results.

In the two decades following publication of the first edition of this book, there has been substantial growth in literature on characterizations, in which characterization of normal distributions has played a prominent part. In fact, to some extent, characterization of normal distributions has become a branch of mathematics, with emphasis on functional equations and characteristic functions but with only limited impact on applied statistics. We are unable to provide a comprehensive, or even fully representative, account due in part to space restriction but also to our feeling that a sizable proportion of the more recent results are of little value in applied work. Exploitation of earlier results has not occurred to the extent that one might have expected, or hoped for. This is true even for results of type 2 above (p. 101).

A possible exception is the following characterization initially due to Csorgo and Seshadri (1971) and Csorgo, Seshadri, and Yalovsky (1975). It was used to develop tests of normality. An early version is as follows: Given $X_1, \ldots, X_n$ $(n = 2k + 3, \; k \geq 2)$ which are i.i.d. with expected value ξ and variance σ^2, and

$$Z_j = \frac{\sum_{i=1}^{j} X_i - j X_j}{\sqrt{j(j+1)}} \qquad j = 1, \ldots, n-1,$$

$$Z_n = \sum_{i=1}^{n} X_i,$$

(13.36)

and given

$$Y_g = Z_{2g-1}^2 + Z_{2g}^2, \qquad g = 1, \ldots, k + 1 = \tfrac{1}{2}(n - 1),$$

and the statistics

$$Y_r^* = \frac{\sum_{g=1}^r Y_g}{\sum_{g=1}^{k+1} Y_g}, \qquad r = 1, 2, \ldots, k,$$

which have the joint distribution of order statistics for k mutually independent standard uniform variables, then the common distribution of $X_1, \ldots, X_n$ is normal.

A parallel result, applicable for n = 2k, if $E[X_j] = \xi$ is known, and it can be assumed that the common distribution is symmetrical, is that if the statistics

$$Y_r'^* = \frac{\sum_{g=1}^r Y_g'}{\sum_{g=1}^k Y_g'}, \qquad r = 1, 2, \ldots, k - 1,$$

where $Y_g' = (X_{2g-1} - \xi)^2 + (X_{2g} - \xi)^2$ $(g = 1, \ldots, k)$ have the joint distribution of order statistics for $k - 1$ mutually independent standard uniform variables, the common distribution of the X's is normal. Further extensions have been obtained in later work.

A typical result of this class due to Pakshirajan and Mohan (1971), which states that if X_1, X_2, and X_3 are mutually independent random variables each symmetric about zero, with cdf continuous at 0 (i.e., $\Pr[X_j = 0] = 0$), then the joint characteristic function of the ratios X_1/X_3 and X_2/X_3 is $\exp(-\sqrt{t_1^2 + t_2^2})$ if and only if the X's have a common normal distribution with zero expected value. However, as noted in the first edition (p. 53), the distribution of the ratio X_1/X_2 of two i.i.d. variables does not characterize a normal distribution [Fox (1965)].

Bryc (1990) has extended Pólya's (1923) result (4c above) as follows: If X_1 and X_2 are i.i.d. random variables with finite variance such that, for some a and $\beta (> 0)$, $(\alpha + \beta X_1)$ and $(X_1 + X_2)$ have identical distributions, then X_1 has a normal distribution.

A further generalization [Wesolowski (1990)] is this: If $X_1, \ldots, X_n$ $(n \geq 2)$ are square integrable random variables and $\{a, \ldots, a,\}$ are real numbers with a, $\neq -1$, 0, or 1 and X_1 and $\sum_{j=1}^n a_j X_j$ have identical distributions, then X_1 has a normal distribution [compare 3a; Shimizu (1962) requires a similar condition for each X,]. Arnold and Isaacson (1978) give a simpler proof. See also Lukacs and Laha (1964), who relaxed the condition of square integrability.

Kagan, Linnik, and Rao (1973) contains some extensions and refinements of results already mentioned. These include

1. If $X_1, \ldots, X_n$ are i.i.d. and

$$\sum_{j-1}^{n} a_j X_j \left(\sum_{j=1}^{n} a_j^2 = 1 \right) \quad \text{and} \quad \left(\sum_{j=1}^{n} a_j X_j \right)^2$$

are mutually independent, the X's are normally distributed (pp. 105–106) (cf. **2a**).

2. If $X_1, \ldots, X_n$ $(n \geq 3)$ are mutually independent, with $E[X_i] = 0$, if

$$L_i - \sum_{j=1}^{n} a_{ij} X_j, \qquad i = 1, \ldots, n,$$

are linearly independent, and if

$$E[L_1 | L_2, \ldots, L_n] = 0,$$

then the X's each have a normal distribution (p. 419).

3. If $X_1, \ldots, X_n$ are mutually independent and $\{a_j\}, \{b_j\}$ are nonzero real numbers satisfying the conditions

$$a_j, b_j \neq 0 \qquad \text{for all } j,$$

$$a_j b_j^{-1} + a_k b_k^{-1} \neq 0 \qquad \text{for all } j \neq k,$$

then each X_j has a normal (possibly degenerate) distribution provided that the conditional distribution of $\sum_{j=1}^{n} a_j X_j$ and $\sum_{j=1}^{n} b_j X_j$ is symmetric.

We also note the following:

4. If X_1 and X_2 are i.i.d., they are unit normally distributed if and only if, for some $a_1, a_2 \neq 0$ each of $(a_1 X_1 + a_2 X_2)^2/(a_1^2 + a_2^2)$ and $(a_1 X_1 - a_2 X_2)^2/(a_2^2 + a_1^2)$ has a χ^2 distribution with 1 degree of freedom [Geisser (1973, pp. 492–494)].

5. Kelker and Matthes (1970) considered location-scale families. These are families with cdf of form

$$F_X(x) = \frac{1}{\sigma} g\left(\frac{x - \theta}{\sigma} \right),$$

where θ, σ are the location and scale parameters, respectively.

According to Kelker and Matthes, "Within the location and scale parameter families $(\overline{X}, S)$ is a sufficient statistic [for (θ, σ)] if and only if the family is the normal distribution family." This is indeed correct, but Bondesson (1977) has pointed out that their proof required $F_X(x)$ to represent a continuous distribution. He amended the proof to avoid this assumption. Kelker and Matthes (1970) also showed that

6. If $X,,\ldots, X_n$ are independent and nondegenerate random variables with location family cdfs $F_{X_i}(x) = g_i(x - \theta)$ $(i = 1, \ldots, n)$, then a necessary and sufficient condition for $\sum_{j=1}^n b_j X_j$ $(b_1, \ldots, b_n \neq 0)$ to be a sufficient statistic for 8 is that X, have a normal distribution with variance proportional to b_j^{-1}.
7. If $X_1, \ldots, X_n$ $(n = 4)$ are i.i.d. and $(X_1 - X_2)/S$ is independent of $(\overline{X}, S)$, then the common distribution of the X's is normal.

Braverman (1985) has obtained characterizations in terms of unconditional expected values. These include the following characterizations:

1. If X_1, X,, and X_3 are i.i.d. random variables with finite moments of odd order, and there are constants C_a such that

$$E\left[\left|\sum_{j=1}^3 a_j X_j\right|^p\right] = C_a\left(\sum_{j=1}^3 a_j^2\right)^{p/2} \tag{13.37}$$

for ail a $= (a,, a,, a,)$ and all odd p, then each of the three variables has a normal distribution.
   This result is not true for even integers p.
2. If X_1 and X_2 are i.i.d., with a symmetric common distribution, with

$$E\left[\exp\left(\lambda |X_j|^2\right)\right] < \infty \qquad \text{for some A} > 0$$

and

$$E\left[|X_j|^s\right] \neq 0 \qquad \text{for all } s$$

real or complex, with Re $s > 0$, and

$$E[|a_1 X_1 + a_2 X_2|] = C\left(a_1^2 + a_2^2\right)^{1/2} \tag{13.38}$$

for suitable C, then the common distribution is normal.
   Braverman (1989) has also shown that
3. If X_1 and X_2 are i.i.d. (not necessarily symmetric) and there are two odd numbers, p_1 and p_2, such that

$$E\left[|a_1 X_1 + a_2 X_2|^{p_j}\right] = C_j\left(a_1^2 + a_2^2\right)^{p_j}, \qquad j = 1, 2, \tag{13.39}$$

for all real a,, a, and suitable C_j, then the common distribution of the X's is normal.

Next we present a few notes on "stability" (or "robustness") of characterizations—namely, how far small departures from exact characterization conditions can affect size of departure from the characterized form.

Defining the distance between two cdfs as

$$\delta(G(x), F(x)) = \sup_x |G(x) - F(x)|$$

Meshalkin (1968) described the two distributions as *ε-coincident* if $\delta(G(x), F(x)) \leq \varepsilon$, and termed a random variable X, with cdf $F_X(x)$, &-normal if

$$\delta\left[F_X(x), \Phi\left(\frac{x - \xi}{\sigma}\right)\right] \leq \varepsilon$$

for some pair (ξ, σ).

Sapogov (1956) showed that if X_1 and X_2 are independent, with

$$F_{X_1}(0) = \tfrac{1}{2},$$

$$\int_{-g(\varepsilon)}^{g(\varepsilon)} x \, dF_{X_1}(x) = \xi_1,$$

$$\int_{-g(\varepsilon)}^{g(\varepsilon)} x^2 \, dF_{X_1}(x) = \sigma_1^2 + \xi_1^2,$$

where

$$g(\varepsilon) = (-2 \log \varepsilon)^{1/2} + 1, \qquad 0 < \varepsilon < 1,$$

then, if X, $+ X_2$ is ε-normal, X_1 is $\beta(\varepsilon)$-normal with

$$\beta(\varepsilon) = C\sigma_1^{-3}(-\log \varepsilon)^{-1/2},$$

where C is some constant (not depending on ε), because

$$\sup_x \left| F_{X_1}(x) - \Phi\left(\frac{x - \xi_1}{\sigma_1}\right) \right| < C\sigma_1^{-3}(-\log \varepsilon)^{-1/2}. \qquad (13.40)$$

Using the concept of *ε-independence* of X_1 and X_2, defined by

$$\left| \int_{a_j x_1 + b_j x_2 < c_j \; (j=1,2)} \delta\big(F_{X_1, X_2}(x_1, x_2) - F_{X_1}(x_1) F_{X_2}(x_2)\big) dx_1 dx_2 \right| \le \varepsilon$$

for all (a_j, b_j, c_j), Nye (1966) showed that

1. if X_1 and X_2 are mutually independent, and $(X_1 + X_2)$ and $(X_1 - X_2)$ are ε-independent, X_1 and X_2 are both $\beta_1(\varepsilon)$-normal,
2. if $X_1, \ldots, X_n$ are i.i.d. random variables and $\bar{X}$ and S^2 are ε-independent, then the common distribution of the X_i's is $\beta_2(\varepsilon)$-normal.

Both $\beta_1(\varepsilon)$ and $\beta_2(\varepsilon)$ were of order $(-\log \varepsilon)^{-1/2}$, and so was $\beta(\varepsilon)$.

Meshalkin (1968) obtained an improved order of magnitude in the following results: If X_1 and X_2 are i.i.d., with expected value zero and variance 1, and $E[|X_i|^2]$ is bounded, then (1') ε-independence of $(X_1 + X_2)$ and $(X_1 - X_2)$ implies $\beta_3 \varepsilon^{1/2}$-normality of the common distribution, and (2') ε-coincidence of the distribution functions of $(X_1 + X_2)/\sqrt{2}$ and X_i implies $\beta_4 \varepsilon^{1/2}$-normality of X_1 and X_2. The multipliers β_3 and β_4 depend on the bound of $E[|X_i|^3]$, but not on ε.

Yanushkyavichyus (1989) in his doctoral dissertation at Vilnius University provided the following stability theorem, improving on the above results: Let $X_1, \ldots, X_n$ be independent identically distributed random variables. If $\bar{X}$ and S^2, defined by the formulas,

$$\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j,$$

$$S^2 = \frac{1}{n} \sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2, \qquad n \ge 2$$

are (δ, ε)-independent [i.e., $\delta(F_{(X,Y)}, F_X F_Y) \le \varepsilon$, where $F_{(X,Y)}(x, y) = \Pr(X < x, Y < y)$, $F_X(x) = \Pr(X < x)$], then

$$\delta(X_j, N) \le C\left(\ln \frac{1}{\varepsilon}\right)^{-1/2} \tag{13.41}$$

[i.e., there exists a normal r.v. N, such that (13.41) is valid, where C is an absolute constant].

Yanushkyavichyus (1989) also improved on Nye (1966) and Meshalkin's result by showing that if X_1 and X_2 are independent identically distributed random variables and if $S = X_1 + X_2$ and $T = X_1 - X_2$ are $(6, \varepsilon)$-independent, then there exists a normal r.v. N, such that

$$\delta(X_j, N) \le C\varepsilon^{1/3}\left(\ln\frac{1}{\varepsilon}\right)^{(2/3)\log_2 3}$$

Note that the loss of order from ε to $\varepsilon^{1/3}$ with a logarithm factor is due to passage from the functional equations, in terms of distribution functions $F_{(S, D)}(x, y) = F_S(x)F_D(y) + r(x, y)$, $(|r(x, y)| \le \varepsilon$, for all $(x, y) \in R^2)$ to a functional equation in terms of characteristic functions. Yanushkyavichyus's results do not involve any restrictions on moments or on possible degeneracy of the r.v. under consideration.

Stability in regression models was studied by Lukacs (1942) and Yanushkyavichyus (1989), among others. Yanushkyavichyus's result is as follows: Let μ be a distance defined on the space of random variables. A r.v. X (with a finite expectation $E[X]$) has (μ, ε)-*regression* on r.v. Y if

$$\mu|E[X|Y], E[X]| \le \varepsilon.$$

Yanushkyavichyus (1989) selects $\mu|Z_1, Z_2| = \mathrm{E}|Z_1 - Z_2|$ and proves the following theorem: If X,, X_2 are i.i.d. random variables with $E[X_j] = 0$, $E[X_j^2] = a^2$, $E[|X_j|^{2+\delta}] \le M < \infty$ for some $6 > 0$ and $L_1 = a_1X_1 + a_2X_2$ has (μ, ε)-regression on $L_2 = b_1X_1 + b_2X_2$ (and $a_1b_1 + a_2b_2 \ne 0$; $|b_2|$, $|b_1| < 1$), then there is a normal distribution $[G(x)]$ with parameter $(0, \sigma^2)$ such that

$$\sup_x |F(x) - G(x)| \le C\sqrt{-\log\varepsilon},$$

where $F(x) = \Pr[X_j \mathbf{I} x]$ and C depends on M and (a,, $a,$, b_1, b_2) only.

Numerous other characterizations of normal are available in the literature. Gabovich (1974) has discussed the stability of certain characterization results. Khatri (1975a, b) has characterized the normal distribution by the constancy of regression; see Gyires (1975). Sampson (1975) has characterized the general exponential family of distributions by moment generating functions. Ruben (1974, 1975) has presented some characterizations through the distribution of the sample variance (see also Chapter 18). A characterization of normal distribution, through the general linear model, has been given by Ruben (1976). Parthasarathy (1976) characterized the normal law through the local independence of some statistics. Several weak sense analogues of characteristic properties were given by Kagan (1976). Zinger (1977) presented

a characterization through identically distributed linear statistics. Many of these characterizations have been reviewed in the book on this topic by Mathai and Pederzoli (1977).

Some characterization results were also developed based on some properties of estimators. For example, Klebanov and **Melamed** (1976) provided a characterization through properties of Bayesian estimators. Klebanov and **Melamed** (1978) also characterized the normal distribution through properties of Fisher information amount. Fieger (1976) characterized the distribution based on a homogeneous **Pitman** estimator. Bondesson (1976), by starting with the question when the sample mean is BLUE, established a characterization result. Bondesson (1974, 1975, 1978) also proved some characterizations of the normal law through properties of distributions of some statistics. In particular, in the 1978 paper Bondesson has shown that the sample variance, when properly normalized, is distributed as chi-square only for the normal distribution (also see Chapter 18). In an interesting note, **Goel** and **DeGroot** (1980) established that only normal distributions have linear posterior expectations in linear regression. **Ahmad (1978)** and Wesolowski (1987) have provided further characterizations based on regression. Bischoff, Cremers, and Fieger (1987) used the sufficiency of the least-squares estimation for characterization. Eberl (1986) has characterized the normal distribution in translation classes through properties of Bayes estimators. Stadje (1988) provided a characterization through maximum likelihood estimation. Klebanov and Neupokoeva (1990) recently proved an interesting characterization by a property of the expected values of order statistics. A characterization of the normal law in the Gauss-Markov model has been given by Stepniak (1991).

Fieger (1977) discussed transformations that characterize the normal distribution. Arnold and Isaacson (1978) presented some characterizations through the distribution of linear forms (assuming finite variance). Prakasa Rao (1979) used some identities for characterizations. Lukacs (1976, 1977, 1980) presented stability theorems for many characterizations. Some more assorted characterization results for the normal distribution are due to Talwalker (1980), Lajko (1980), Joshi (1982), Borovkov and Utev (1983), Ramasubramanian (1985), Viskov (1985), Ahsanullah and Hamedani (1988), and Ahsanullah (1990). Letac (1981) proved some interesting characterizations via the concepts of isotropy and sphericity. Findeisen (1982) has discussed Gauss's characterization of the normal distribution. Hombas (1985) has characterized the normal density function as the solution of a differential equation. Ahsanullah (1989) has used properties of linear statistics as well as chi-squared in order to characterize the normal distribution. It is important to mention that the finiteness of the variance is a critical assumption in many of the above mentioned characterization results; see Lancaster (1987) for some comments in this regard. Quite recently, Cacoullos, Papathanasiou, and Utev (1993) discussed a characterization of the normal distribution and also presented a proof of the central limit theorem connected with it.

7   APPROXIMATIONS AND ALGORITHMS

The most common use of the normal distribution is as an approximation where either normality is ascribed to a distribution in the construction of a model or a known distribution is replaced by a normal distribution with the same expected value and standard deviation. Examples of such replacement are the Fisher and Wilson-Hilferty approximations to the χ^2-distribution (Chapter **18**), the normal approximation to the (central) t-distribution (Chapter **28**), and the use of normal distribution to approximate the distribution of the arithmetic mean of a number (often not very large, around 8 or more) of independent and identically distributed random variables. But now we are concerned with approximations to the normal distribution. It is possible to regard the distributions that are approximated by the normal distribution as being, themselves, approximations to normal distributions. However, they are usually more complex than the normal distribution, and we would like to study approximations that are simpler than the normal distribution.

From the point of view of replacement of a normal distribution by another distribution we note that:

1. A lognormal distribution can give a good representation of a normal distribution that has a small absolute value (say, less than 0.25) of the coefficient of variation.
2. A particular form of logistic distribution is very close to a normal distribution (see Chapter 23).
3. A form of the Weibull distribution with the shape parameter ≈ 3.25 is almost identical with the unit normal distribution (see Chapter 21).
4. Raab and Green (1961) have suggested that the distribution with probability density function

$$(2\pi)^{-1}(1 + \cos x), \qquad -\pi < x < \pi, \qquad (13.42)$$

can be used to replace a normal distribution. The correspondence is not very precise (see Table 13.2 comparing standardized percentile deviates of the two distributions) but will sometimes give useful analytical results. The replacement would only be used if substantial simplification in analysis were effected thereby.

The expected value and standard deviation of a random variable with distribution (13.42) are zero and $(\frac{1}{3}\pi^2 - 2)^{1/2} = 1.14$. The standardized range of the distribution $(-\pi, \pi)$ is thus from -2.77 to $+2.77$ standard deviations, and obviously the replacement gives a poor fit in the tails.
5. Bell (1962) has described even simpler approximations, using triangular distributions (Chapter 26). He pointed out that such approximations can be regarded as the second stage in a series of approximations

Table 13.2   Standardized Percentile Points of Distribution (13.42) and the Normal Distribution

Cumulative Probability	Standardized Value	
	Normal	Distribution (13.42)
0.5	0.000	0.000
0.6	0.253	0.279
0.75	0.674	0.732
0.9	1.282	1.334
0.95	1.645	1.649
0.975	1.960	1.888
0.99	2.326	2.124
β_2	3.000	2.406

by distributions of means of increasing numbers of independent rectangularly distributed variables (see the method of construction of "random normal deviates" used by Buslenko et al. (1966) described in Section 1).

Chew (1968) includes 2, 4, and 5 in a list of five possible replacements for normal distributions. The two other distributions he suggests are uniform (Chapter 26) and Laplace (Chapter 24). These also will be very crude approximations.

6. Hoyt (1968) has suggested using the distribution of the sum of three mutually independent random variables each uniformly distributed over the interval -1 to $+1$ as an approximation to the unit normal distribution. The density function is

$$\tfrac{1}{8}(3 - x^2) \qquad \text{for } |x| \leq 1,$$

$$\tfrac{1}{16}(3 - |x|)^2 \qquad \text{for } 1 \mathbf{I} |x| 13.$$

This gives an error not exceeding 0.01 in the cumulative distribution function.

7. Steffensen (1937) has suggested the use of the distribution of a multiple of a chi random variable (i.e., $c\chi_\nu$), with v sufficiently large. He called this a "semi-normal" distribution.

8. A different kind of approximation has been developed in connection with calculation of the functions $\Phi(\cdot), Z(\cdot)$ in computers. These approximations usually employ polynomial expressions. They give quite high accuracy, sometimes only within definite limits on the values of the variable. Outside these limits they may give quite poor approximation.

Zelen and Severo (1964) quote, among other things, the **following** formulas, which are based on formulas given by **Hastings (1955)**:

$$\Phi(x) \doteq 1 - \left(a_1 t + a_2 t^2 + a_3 t^3\right) Z(x), \qquad (13.43)$$

with $t = (1 + 0.33267x)^{-1}$, $a_1 = 0.4361836$, $a_2 = -0.1201676$, and $a_3 = 0.9372980$. The error in $\Phi(x)$, for $x \geq 0$, is less than 1×10^{-5}.

$$\Phi(x) \doteq 1 - \tfrac{1}{2}\left(1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4\right)^{-4}, \quad (13.44)$$

with $a_1 = 0.196854$, $a_2 = 0.115194$, $a_3 = 0.000344$, and $a_4 = 0.019527$. The error in $\Phi(x)$, for $x \geq 0$, is less than 2.5×10^{-4}.

$$Z(x) \doteq \left(a_0 + a_2 x^2 + a_4 x^4 + a_6 x^6\right)^{-1}, \qquad (13.45)$$

with $a_0 = 2.490895$, $a_2 = 1.466003$, $a_4 = -0.024393$, and $a_6 = 0.178257$. The error in $Z(x)$ is less than 2.7×10^{-3}.

Very accurate results can be obtained with the formula [Hart (1966)]

$$1 - \Phi(x) \doteq \left(x\sqrt{2\pi}\right)^{-1}\left[\exp(-x^2/2)\right]$$

$$\times \left[1 - \frac{(1 + bx^2)^{1/2}}{(1 + ax^2)} \left\{ x\sqrt{\frac{\pi}{2}} \right. \right.$$

$$\left. \left. + \left[\frac{1}{2}\pi x^2 + \frac{(1 + bx^2)^{1/2}}{(1 + ax^2)} e^{-x^2/2} \right]^{1/2} \right\}^{-1} \right] \quad (13.46)$$

with

$$a = \frac{1}{2\pi}\left[1 + (1 + 6\pi - 2\pi^2)^{1/2}\right] = 0.212024,$$

$$b = \frac{1}{2\pi}\left[1 + (1 + 6\pi - 2\pi^2)^{1/2}\right]^2 = 0.282455.$$

For $x > 2$, Schucany and Gray **(1968)** have constructed the simpler formula

$$1 - \Phi(x) \doteq \left[(x^2 + 2)\sqrt{2\pi}\right]^{-1}$$

$$\times x\left[\exp\left(\frac{-x^2}{2}\right)\right]\frac{x^6 + 6x^4 + 14x^2 - 28}{x^6 + 5x^4 - 20x^2 - 4}, \quad (13.47)$$

which is even better than (13.46) for $x > 3$. [The proportionate error of (13.46) for $5 \leq x \leq 10$ is about 0.5×10^{-5}; that of (13.47) decreases from 0.39×10^{-5} for $x = 5$ to 0.42×10^{-7} for $x = 10.1$

By use of rather elaborate formulas, quite remarkable accuracy can be attained. **Strecock** (1968) gives formulas for values of $\mathrm{erf}(x)$ [see (13.5)] and of the inverse function [$\mathrm{inverf}(y)$ where $\mathrm{erf}(\mathrm{inverf}(y)) = y$], correct to 22 decimal places for $|x|$ (or $|\mathrm{inverf}(y)|$) less than 7.85.

9. Burr (1967) has considered approximations to $\Phi(x)$ of form

$$G(x) = 1 - \left[1 + (\alpha + \beta x)^c \right]^{-k}.$$

He suggests taking $a = 0.644693$, $\beta = 0.161984$, $c = 4.874$, and $k = -6.158$. An even better approximation is obtained by using

$$H(x) = \tfrac{1}{2}[G(x) + 1 - G(-x)],$$

which is a symmetrical function of x. The discrepancy $|H(x) - \Phi(x)|$ reaches its maximum value of about 0.00046 when $x \doteq \pm 0.6$.

10. **McGillivray** and Kaller (1966) have considered the discrepancy between $\Phi(x)$ and $\Phi(x) + a_{2r} Z(x) H_{2r-1}(x)$, where $H_{2r-1}(x)$ is the **Hermite** polynomial of order $2r - 1$ and a,, is a constant chosen so that $1 + a_{2r} H_{2r}(x)$ cannot be negative. This means that a,, must be between zero and

$$A,, = \left| \inf_x H_{2r}(x) \right|^{-1}.$$

The second function $\Phi(x) + a_{2r} Z(x) H_{2r-1}(x)$ is the cumulative distribution function of a symmetrical distribution having the first r even central moments (and of course all odd moments), the same as those for a unit normal distribution. The discrepancy cannot exceed

$$A_{2r} \sup_x \{ Z(x) | H_{2r-1}(x) | \}.$$

The values of this quantity, for $r = 2$, 3, and 4 are 0.10, 0.03, and 0.005, respectively. [Of course other distributions with the same (zero) odd central moments and first r even central moments *might* have greater discrepancies, but these results do give a useful idea of the accuracy obtained by equating moments.]

11. Riffenburgh (1967) has suggested that a symmetrical truncated unit normal distribution be approximated by the density function

$$\frac{\frac{1}{2}(Z(x) - Z(c))}{\Phi(c) - \frac{1}{2} - cZ(c)}, \qquad -c \le x \le c,$$

where $-c, c$ are the points of truncation. Use of this approximation is recommended only when c exceeds 1 (preferably c $\ge$ 1.5). Tables of the variance (to 3 decimal places) of the approximate distribution are given by Riffenburgh (1967) for c = 0.8(0.1)1.2(0.05)4.00, and also of $\Pr[X \le x] - \frac{1}{2}$ (to 4 decimal places) for c = 1.2(0.1)3.0 and x at intervals of 0.05. (Riffenburgh has also developed test procedures based on this distribution.)

We now discuss some bound on the value of $\Phi(x)$. Various inequalities for Mills's ratio can also be interpreted as bounds for $\Phi(x)$ or $Z(x)$. Using a simple geometrical argument (based on the joint distribution of two independent unit normal variables) it can be shown that

$$\frac{1}{2}\left[1 + (1 - e^{-x^2/2})^{1/2}\right] \le \Phi(x) \le \frac{1}{2}\left[1 + (1 - e^{-x^2})^{1/2}\right] \quad (13.48)$$

[e.g., see D'Ortenzio (1965)]. By a refinement of the argument, on the left-hand side of (13.48), $(1 - e^{-x^2/2})$ can be replaced by

$$1 - e^{-x^2/2} + \left(2\pi^{-1} - \frac{1}{2}\right)^2 e^{-x^2},$$

and on the right-hand side, $(1 - e^{-x^2})$ can be replaced by

$$1 - e^{-x^2} - (1 - 2\pi^{-1})^2 e^{-x^2}.$$

The approximation

$$\Phi(x) \doteq \frac{1}{2}\left[1 + \left\{1 - \exp\left(\frac{-2x^2}{\pi}\right)\right\}^{1/2}\right] \quad (13.49)$$

was obtained by Pólya (1945). This has a maximum error of 0.003, when $x = 1.6$. Cadwell (1951) modified (13.49) to

$$\Phi(x) \doteq \frac{1}{2}\left[1 + \left\{1 - \exp(-2\pi^{-1}x^2 - \frac{2}{3}\pi^{-2}(\pi - 3)x^4))\right\}^{1/2}\right]. \quad (13.50)$$

Over the range $0 < x < 3.5$, the maximum error of (13.50) is 0.0007, when $x = 2.5$. Formula (13.50) should not be used for large values of x. **Cadwell** suggested, on empirical grounds, the addition of the terms

$$-0.0005x^6 + 0.00002x^8$$

to the exponent in (13.50). This reduces the maximum error to 0.00005.

Carta (1975) developed approximations, similar to that of **Hastings** in (**13.44**), of the form

$$\Phi(x) \doteq 1 - \tfrac{1}{2}(a_1 + a_2 x + \cdots + a_n x^{n-1})^{-2q}, \quad \text{for } x \ge 0. \quad (13.51)$$

Note the added flexibility of a variable first coefficient, as opposed to the fixed value of 1 in (13.44). For different choices of n and q, Carta (1975) has presented the coefficients a_i's that yield the minimum absolute error. For example, from **Carta's** table, we have the coefficients corresponding to $n = 6$ and $q = 4$ as

0.9999998582,   0.0487385796,   0.02109811045,

0.003372948927, -0.00005172897742,   0.00008569579420.

The absolute error for this approximation (for all $x \ge 0$) is less than 1.2×10^{-6}. Note that in this case (and also for some other choices of n, q), the leading coefficient is very nearly 1 [as in (13.44)]. Carta also has presented similar approximations for x restricted in the intervals $[0, 3.091, [0, 4.001,$ and $[0, 5.20]$.

Badhe (1976) presented the following approximation which is easily **implementable** on a hand calculator:

$$\Phi(x) \doteq 1 - \frac{Z(x)}{x}\left[1 - \frac{1}{Y}\left(1 + \frac{1}{Y}\left\{7 + \frac{1}{Y}\left[55 + \frac{1}{Y}\left(445 + \frac{3745Q_1(x)}{Y}\right)\right]\right\}\right)\right],$$

$$(13.52)$$

where $Y = x^2 + 10$ and

$$Q_1(x) = 8.5(x^2 - 0.4284639753x^{-2} + 1.240964109)^{-1} + 1.$$

Badhe (1976) has pointed out that the approximation in (**13.52**) is good when $x > 4$, but certainly not suitable when x I 2. For the case where x I 2, **Badhe** has presented a seventh degree polynomial approximation (obtained

by Chebyshev economization) given by

$$\Phi(x) \doteq \tfrac{1}{2} + x\{a + Y[b + Y(c + Y\{d + Y[e + Y(f + Y(g + hY))]\})]\},$$
(13.53)

where

$Y = x^2/32,$
$a = 0.3989422784,$
$b = -2.127690079,$
$c = 10.2125662121,$
$d = -38.8830314909,$
$e = 120.2836370787,$
$f = -303.2973153419,$
$g = 575.073131917,$
$h = -603.9068092058.$

The maximum absolute error for this approximation, for $x \in [0,2]$, is 0.2×10^{-8}.

By making use of the **Hermite** expansion

$$\sqrt{2\pi} \int_0^x Z(t)\, dt = 2e^{-x^2/8} \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} (x/2)^{2n+1} H_{2n}(x/2), \quad (13.54)$$

where $H_n(x)$ is the nth **Hermite** polynomial, and the known recurrence relation

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x) \qquad \text{for n} = 1, 2, \ldots,$$

with $H_0(x) \equiv 1$ and $H_1(x) = x$, Kerridge and Cook (1976) suggested using the series

$$\sqrt{2\pi} \int_0^x Z(t)\, dt = xe^{-x^2/8} \sum_{n=0}^{\infty} \frac{1}{2n+1} \theta_{2n}(x/2) \qquad (13.55)$$

for computing $\Phi(x)$ on a computer. In (13.55), $\theta_n(x) = x^n H_n(x)/n!$, which is easily computed using the recurrence relation

$$\theta_{n+1} = \frac{x^2(\theta_n - \theta_{n-1})}{n+1}, \qquad n = 1, 2, \ldots.$$

Beasley and Springer (1977) have provided an algorithm for computing the percentage point x_p, for a specified value of p, such that

$$p = \int_{-\infty}^{x_p} Z(t)\, dt.$$

Their Fortran subroutine replaces p by $q = p - \frac{1}{2}$ and then compares $|q|$ with 0.42; if $|q| \le 0.42$, x_p is determined by a rational approximation

$$x_p = \frac{qA(q^2)}{B(q^2)},\qquad\qquad (13.56)$$

where A and B are polynomials of degrees 3 and 4, respectively, while if $|q| > 0.42$ an auxiliary variable, $r = \{\ln(\frac{1}{2} - |q|)\}^{1/2}$ is first formed and then x_p as

$$x_p = \pm \frac{C(r)}{D(r)},\qquad\qquad (13.57)$$

where C and D are polynomials of degrees 3 and 2, respectively, the sign being taken that of q. See also an earlier algorithm given by Odeh and Evans (1974).

Page (1977) considered simple approximations of the form

$$\Phi(x) \doteq \frac{e^{2y}}{1 + e^{2y}},\qquad\qquad (13.58)$$

where $y = a_1 x(1 + a_2 x^2)$, and determined a, = 0.7988 and a, = 0 04417 to provide an approximation with maximum error 0.14×10^{-3}. Page has also presented a similar simple approximation for the percentage point that gives two decimal accuracy.

Derenzo (1977) provided an approximation to the unit normal cdf as

$$\Phi(x) = 1 - \frac{1}{2}\exp\left\{-\frac{(83x + 351)x + 562}{(703/x) + 165}\right\},\qquad x > 0,\quad (13.59)$$

with a **maximum** absolute error of 0.042% for $x \in (0, 5.5]$. Another approximation that Derenzo presented is

$$\Phi(x) \doteq 1 - \frac{1}{\sqrt{2\pi}\, x}\exp\left(-\frac{x^2}{2} - \frac{0.94}{x^2}\right)\qquad\qquad (13.60)$$

with a maximum absolute error of 0.040% for $x \ge 5.5$. Derenzo (1977) also provided an approximation for the percentage point x_p (for a given p) as

$$x_p \doteq \left\{\frac{[(4y + 100)y + 205]y^2}{[(2y + 56)y + 192]y + 131}\right\}^{1/2},\qquad\qquad (13.61)$$

where $y = -\log(1 - p/2)$, with a maximum absolute error of 1.3×10^{-4} for $x \in (0, 5.2)$. For $x \in [5.2, 22.6)$, Derenzo has given the approximation

$$x_p \doteq \left\{ \frac{[(2y + 280)y + 572]y}{(y + 144)y + 603} \right\}^{1/2} \tag{13.62}$$

with a maximum absolute error of 4×10^{-4}.

Some more simple (but not necessarily more accurate for all x) approximations for the cumulative distribution function $\Phi(x)$ have been given by Parsonson (1978), Easingwood (1979), Heard (1979), Martynov (1981), Monahan (1981), Allasia and Giordano (1982), Hawkes (1982), Fearn (1983), Edgeman (1988), and Abernathy (1988). Pugh (1989) has made a survey of many of the algorithms available for computing the cdf $\Phi(x)$.

Moran (1980), by slightly modifying a formula of Strecock (1968) given for the error function, came up with the approximation

$$\Phi(x) \doteq \frac{1}{2} + \frac{1}{\pi} \sum_{n=0}^{12} \left(n + \frac{1}{2}\right)^{-1} e^{-(n+(1/2))^2/9} \sin\left\{\frac{\sqrt{2}}{3}\left(n + \frac{1}{2}\right)x\right\}. \tag{13.63}$$

Compared to the 38-term approximation given by Strecock, this approximation is simpler to use and is also accurate to nine decimal places for $|x| \le 7$.

Shore (1982), by using a "betalike" cumulative distribution $G(x)$ [with $G(-\infty) = 0$, $G(\infty) = 1$, $G(0) = \frac{1}{2}$, and

$$\frac{dG}{dx} = \begin{cases} cG^{k_1}(1 - G)^{k_2}, & 0 \le G \le \frac{1}{2}, \\ c(1 - G)^{k_1}G^{k_2}, & \frac{1}{2} \le G \le 1, \end{cases}$$

where $k_1 > 0$] to approximate $\Phi(x)$, derived the following three approximations for the percentage point x_p:

$$x_p \doteq -5.5310\left\{\left(\frac{1 - p}{p}\right)^{0.1193} - 1\right\}, \qquad p > \frac{1}{2}, \tag{13.64}$$

$$x_p \doteq -0.4115\left\{\frac{1 - p}{p} + \ln\left(\frac{1 - p}{p}\right) - 1\right\}, \qquad p \ge \frac{1}{2}, \tag{13.65}$$

$$x_p \doteq -a\ln\left(\frac{1 - p}{p}\right) + b, \qquad p \ge \frac{1}{2}, \tag{13.66}$$

where a and b should satisfy

$$b = [1 - 1.3682a^2]^{1/2} - 1.3862a$$

or its approximate simpler form b = *1.3086* − 2.3735*a* (which is easier to use when a value of a is sought that gives the best fit for any desired range of *x*).

The approximation in (13.64) is the most accurate of the three with a maximum absolute difference of *0.0073 (0.5%)* for $0 \le x \ne 2.3$. It is interesting to note that for b = 0, we have *a* = *0.5513* = $\sqrt{3}/\pi$, in which case the simple approximation in *(13.66)* becomes the logistic approximation (with location parameter 0 and shape parameter $\sqrt{3}/\pi$); see Chapter *23.* For *−0.5* ≤ *x* ≤ *2.2,* the best approximation of (13.66) is with *a* = *0.495* and *b* = *0.1337;* for *x* > *2.2,* the best approximation is obtained with *a* = *0.4506* and b = *0.2252.*

Shore (1982) has also obtained a good approximation for the standard normal pdf as

$$Z(x) \doteq 1.4184(1-p)^{0.8632}p,$$
$$p = \Phi(x) \ge \tfrac{1}{2}. \tag{13.67}$$

This approximation immediately yields an approximation for the Mills's ratio (hazard rate−see Chapter 33, Section *2)* as

$$\frac{Z(x)}{1-\Phi(x)} \doteq \begin{cases} \left[1.4184p^{0.8632}\right]^{-1}, & p \le \tfrac{1}{2}, \\ \left[1.4184(1-p)^{-0.1368}p\right]^{-1}, & p \ge \tfrac{1}{2}. \end{cases} \tag{13.68}$$

Shore has also discussed the accuracy of this approximation.

Shah *(1985)* suggested the following approximation for $\Phi(x) - \tfrac{1}{2}, x \ge 0$:

$$\begin{cases} \dfrac{x(4.4-x)}{10}, & 0 \le x \le 2.2 \\ 0.49, & 2.2 < x < 2.6 \\ 0.50, & x \ge 2.6. \end{cases} \tag{13.69}$$

Even though this approximation is simple to use, it is clear from *(13.69)* that it is not designed to approximate right tail areas for *x* ≥ *2.6.* For this reason Norton (1989) proposed the approximations for *1* − $\Phi(x)$ as

$$\tfrac{1}{2}e^{-(x^2+x)/2}, \qquad 0 \le x \le 2.6, \tag{13.70}$$

$$\tfrac{1}{2}e^{-(x^2+1.2x^{0.8})/2}, \qquad 0 \le x \le 2.7, \tag{13.71}$$

$$Z(x)/x, \qquad x > 2.7. \tag{13.72}$$

A number of comments and criticisms have been made on these approximations given by Norton; see Hoaglin (1989), Pratt (1989), Cox (1989), Shore

(1990), Sharp (1990), Revfeim (1990), and McConnell (1990), and also the replies to these comments by Norton (1990a, b).

Schonfelder (1978) discussed further on Chebyshev expansions for the error and related functions, including $\Phi(x)$. Hamaker (1978) presented simple approximations for the cdf $\Phi(x)$ as well as for the percentage point x_p. Lin (1988) discussed alternatives to Hamaker's approximations. Schmeiser (1979) gave easy approximations for the percentage point x_p that could be used on hand calculators. Bailey (1981) proposed alternatives to Hastings's approximation to x_p. Lew (1981) presented an approximation to $\Phi(x)$ with simple coefficients. Wichura (1988) and Guirguis (1991) have discussed approximations to the percentage point x_p. While Heffernan (1988) gave a simple formula for the cdf $\Phi(x)$, Lin (1989) presented approximations for the normal tail probability as well as its inverse which are easy to use even on a hand calculator. Lin (1990) has also proposed a simpler logistic approximation to $\Phi(x)$ and to x_p [also see the comment after (13.66)]. Even though all these approximations are quick and easy to use, their relative accuracies may vary dramatically, and therefore their use should be adjudged for the particular situation one intends to use. The modern personal computer with its great power and memory does make some of the work (if not all) unnecessary, as the following remarks indicate.

Fleming (1989) has advocated the use of numerical integration for approximating $\Phi(x)$, for it can be easily demonstrated using a spreadsheet program and a personal computer. For example, $\Phi(x)$ can be computed using Lotus 1-2-3 by approximating the area under the standard normal pdf $Z(x)$. Fleming (1989) has noted that a rectangular integration gives results accurate to four decimal places for $0 \le x \le 3$, with the interval width being taken as 0.01. As he has pointed out, other methods of integration (like the trapezoidal rule, Simpson's rule, Newton's three-eighths rule) are also easily adaptable and will be even more accurate in evaluating $\Phi(x)$. When a powerful personal computer with a spreadsheet program is available, it should be put to use rather than relying on simple-to-use approximations.

A mechanical method of drawing a normal probability density curve has been described by Edwards (1963). *Normal probability paper* is graph paper with a natural scale in the horizontal (abscissa) direction, while the distances on the vertical (ordinate) scale are proportional to the corresponding normal deviates. The vertical scale is usually marked in percentages. Thus 50% correspond to the horizontal axis, 25% and 75% are at distances 0.6745 below and above this line, 5% and 95% are at distances 1.9600 below and above this line, and so on (see Figure 13.2). Barnett (1976) has discussed convenient probability plotting positions, while Nelson (1976) has elaborated the construction of normal probability paper. Recently Nelson (1989) put forward a stabilized normal probability plotting technique; Rouncefield (1990), among many others, explained how one could use the normal probability paper to assess the validity of the assumption of normal distribution for a sample.
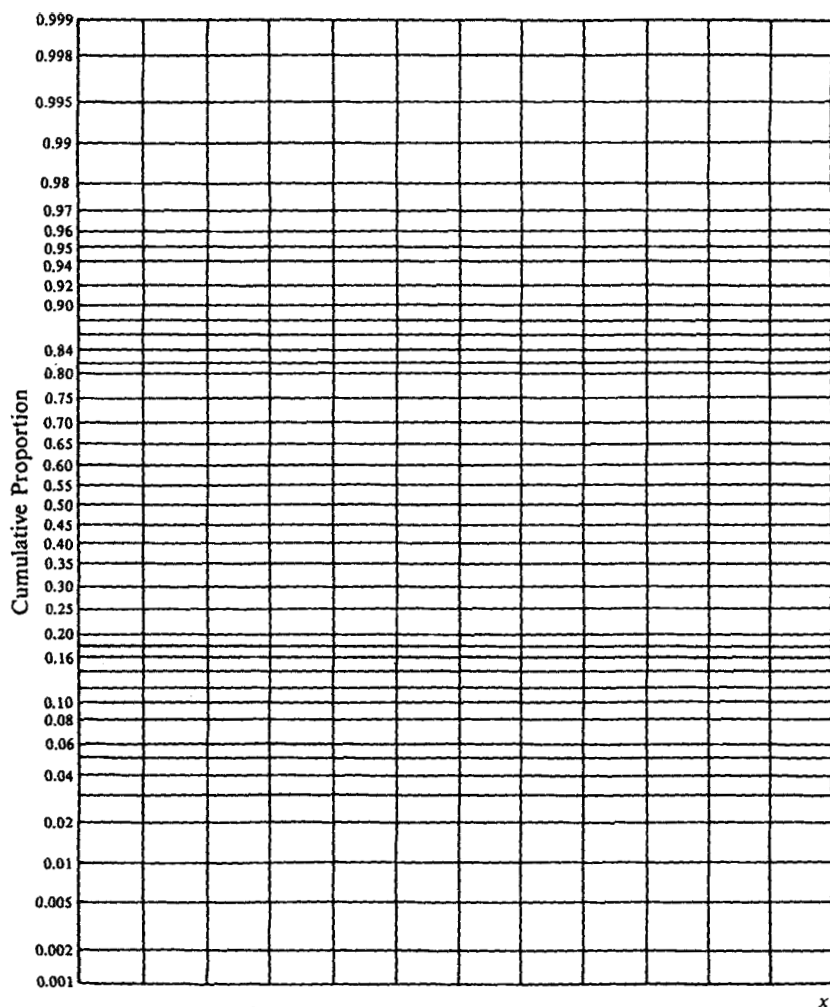
Figure 13.2 Normal Probability Paper

If X has the distribution (13.1) and $\Pr[X \leq x]$ is plotted (as the ordinate) against x (as the abscissa), then a straight line is obtained. The slope of this line is σ^{-1} and its intercept on the horizontal axis is at $x = \xi$. If observed frequencies of the events $(X \leq x)$ are used in place of the actual probabilities, an approximately straight-line plot may be expected. A straight line fitted to these observed points gives estimates of σ and ξ. Such graphical methods of estimation can give good practical accuracy,

If the horizontal scale is logarithmic, we have *lognormal probability paper* (see Chapter 14). *Half-normal probability paper* is simply normal probability paper with negative abscissas omitted. It is used in connection with analysis of variance techniques developed by Daniel (1959).

8 ESTIMATION

The theory of estimation of ξ and σ has been fully worked out. To facilitate comprehension, this section is divided into four subsections. Subsections 8.1 through 8.3 describe techniques primarily appropriate to a complete sample (though some apply also to censored data) corresponding to values of n independent random variables each having distribution (13.1). Subsection 8.4 describes techniques suitable when the data have been censored by omission of certain order statistics. (Truncated normal distributions will be discussed in Section 10.) Subsection 8.5 lists various **books/monographs** available on specific topics relating to inference for the normal distribution.

The variety of applicable methods can be bewildering. To judge rapidly between them, it is necessary to bear in mind accuracy, sensitivity to variations from normality, and ease of calculation. The relative importance of these factors varies with circumstance, but they should always be taken into account. As mentioned earlier in Section 2, there are numerous volumes available on specific topics relating to inference for the normal distribution (as listed in Section 8.5). This is understandable of course, due to the volume of literature. Because of the availability of all these volumes and the upcoming revised edition of "Handbook of the Normal Distribution" by **Patel** and Read, we only present briefly the recent developments, referring the readers to appropriate volumes for a comprehensive treatment of the topic of interest.

8.1 Estimation of ξ

The arithmetic mean $\bar{X} = n^{-1}\sum_{j=1}^{n} X_j$ and the mean square deviation $S^2 = n^{-1}\sum_{j=1}^{n}(X_j - \bar{X})^2$ are jointly sufficient for ξ and σ, and $\bar{X}$ is sufficient for ξ alone. For most practical purposes $\bar{X}$ is the best estimator for ξ, whether or not σ is known. It is the maximum likelihood estimator and is unbiased.

The only circumstances under which this estimator would not be used are (1) when not all observations are available (as will be discussed more fully in Section 8.4), or (2) when the accuracy of some values (e.g., outlying values) is doubtful. In either case estimation may be based on a central block of order statistics. As an extreme case (when n is odd), a single order statistic, the median, may be used to estimate ξ. This is an unbiased estimator of ξ, and has standard deviation approximately equal to

$$\tfrac{1}{2}\pi\sigma^2 n^{-1} = 1.5708\sigma^2 n^{-1}$$

compared with

$$\mathrm{Var}(\bar{X}) = \sigma^2 n^{-1}.$$

The efficiency of the median, relative to $\bar{X}$, is thus approximately

$$100 \times (1.5708)^{-1}\% = 63.7\%.$$

Alternatively, the jth *Winsorized* mean

$$\tilde{\xi}_{(j)} = n^{-1}\left[jX'_{j+1} + \sum_{i=j+1}^{n-j} X'_i + jX'_{n-j} \right] \tag{13.73}$$

may be used $(j < [(n-1)/2])$. It can be seen that $\tilde{\xi}_{(j)}$ is obtained by replacing each of $X'_1, X'_2, \ldots, X'_j$ by X'_{j+1} and $X'_{n-j+1}, \ldots, X'_n$ by X'_{n-j}. This is also an unbiased estimator of ξ. It is interesting to note that Chernoff, **Gastwirth,** and Johns (1967) obtain a formula of the type (13.73), if only $X'_{j+1}, \ldots, X'_{n-j}$ are available, with the multipliers (j) of X'_{j+1} and X'_{n-j} replaced by

$$a = \frac{[Z(U_\varepsilon)]^2 \varepsilon^{-1} + U_\varepsilon Z(U_\varepsilon)}{1 - 2\varepsilon + 2U_\varepsilon Z(U_\varepsilon) + 2[Z(U_\varepsilon)]^2 \varepsilon^{-1}},$$

where $\varepsilon = j/(n+1)$, and n^{-1} replaced by $(n - 2j + 2a)^{-1}$. In fact a is slightly less than j.

Rather than replace the values of extreme observations by more central values, we can simply omit them. The resulting unbiased estimator of ξ is the jth trimmed mean

$$\tilde{\xi}'_{(j)} = (n - 2j)^{-1} \sum_{i=j+1}^{n-j} X'_i. \tag{13.74}$$

Some relative efficiencies of $\tilde{\xi}'_{(j)}$, compared with $\bar{X}$, are shown in Table 13.3. (For efficiencies of $\tilde{\xi}_{(j)}$, see Table 13.9.)

Table 13.3   Efficiency of Trimmed Means, Relative to $\bar{X}$

n	j	Relative Efficiency of $\tilde{\xi}'_{(j)}$ (%)
5	2	69
10	2	89
10	3	81
15	2	92
15	4	83
15	6	73

It is apparent that the Winsorized mean $\tilde{\xi}_{(j)}$ is more efficient than $\tilde{\xi}'_{(j)}$. In fact, compared with the best linear unbiased estimator using the same order statistics, its efficiency never falls below 99.9% for $n \leq 20$ [Sarhan and Greenberg (1962)]. Meeden (1976) has pointed out a special property of linear estimates of ξ, while Mehran (1975) has derived relationships between the UMVUEs of the mean and median of a function of a normal distribution.

Knowledge of σ is of no help in calculating point estimators of ξ. It is, however, used in calculating the standard deviations of such estimators and in constructing confidence intervals for 5. If σ is known, $100(1 - a)\%$ confidence limits for ξ are

$$\bar{X} \pm \frac{U_{1-\alpha/2}\sigma}{\sqrt{n}}. \tag{13.75}$$

Although the similar formulas

$$\tilde{\xi}_{(j)} \pm U_{1-\alpha/2} \sqrt{\mathrm{Var}\!\left(\tilde{\xi}_{(j)}\right)},$$

$$\tilde{\xi}'_{(j)} \pm U_{1-\alpha/2} \sqrt{\mathrm{Var}\!\left(\tilde{\xi}'_{(j)}\right)} \tag{13.76}$$

do not give exact limits (since $\tilde{\xi}_{(j)}, \tilde{\xi}'_{(j)}$ do not have normal distributions), they give limits that are useful provided that n is not too small (e.g., $n \geq 15$).

If σ is not known, the above formulas cannot be used. It is natural to replace σ by an estimator of σ. If the sample size is large and a good (efficient) estimator of σ is used, this can be done with little serious effect on the confidence coefficient. The "estimator" of σ most often employed is

$$(1 - n^{-1})^{-1/2}S = \left[(n - 1)^{-1} \sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2\right]^{1/2},$$

although this is not an unbiased estimator of σ. If this estimator is used, then $U_{1-\alpha/2}$ in (13.75) should be replaced by $t_{n-1, 1-\alpha/2}$, the upper $50\alpha\%$ point of the t distribution with $(n - 1)$ degrees of freedom (see Chapter 28). The effect of replacement of σ by $(1 - n^{-1})^{-1/2}S$ in formulas (13.76) is not so clear, but there should be comparable increase in the multiplying factor $U_{1-\alpha/2}$.

It can be shown (see Section 3) that $\bar{X}$ and any function of the deviations $\{X_j - \bar{X}\}$, only, are mutually independent. This facilitates computation of percentage points of distributions of statistics of form

$$\frac{\sqrt{n}\,(\bar{X} - \xi)/\sigma}{f\left(X_1 - \bar{X}, \ldots, X_n - \bar{X}\right)} = T_{\{f\}},$$

say, with various functions $f(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ in the denominator, subject to the restrictions that $f(\cdot)$ is positive with probability one and that

$$f(ay_1, ay_2, \ldots, ay_n) = af(y_1, y_2, \ldots, y_n)$$

for any $a \geq 0$. For example, we might have

$$f(X_1 - 8, \ldots, X_n - \bar{X}) = \text{range}$$

$$= \max(X_1 - \bar{X}, \ldots, X_n - \bar{X})$$

$$- \min(X_1 - \bar{X}, \ldots, X_n - \bar{X})$$

$$= W$$

or

$$f(X_1 - \bar{X}, \ldots, X_n - \bar{X}) = \text{mean deviation}$$

$$= n^{-1} \sum_{j=1}^{n} |X_j - \bar{X}|$$

$$= M.$$

Indeed, any of the estimators of σ to be described in Section 8.2 might be used as $f(\cdot)$.

Under the conditions stated, the distribution of $\sigma^{-1}f(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ does not depend on σ. The distribution of $T_{(f)}$ therefore does not depend on ξ or σ. As a result it is possible, in principle, to construct tables of percentage points $T_{(f), a}$ of this distribution, defined by $f(.)$ and a alone. The relation

$$\Pr\left[T_{(f), \alpha_1} < \frac{\sqrt{n}\,(\bar{X} - \xi)/\sigma}{f(X_1 - \bar{X}, \ldots, X_n - \bar{X})} < T_{(f), 1-\alpha_2} \Big| \xi, \sigma \right] = 1 - (\alpha_1 + \alpha_2)$$

can be arranged to show that the limits

$$\bar{X} - \frac{\sigma T_{(f), \alpha_1} f(X_1 - \bar{X}, \ldots, X_n - \bar{X})}{\sqrt{n}},$$

$$\bar{X} - \frac{\sigma T_{(f), 1-\alpha_2} f(X_1 - \bar{X}, \ldots, X_n - \bar{X})}{\sqrt{n}},$$

$$(13.77)$$

form a $100(1 - \alpha_1 - \alpha_2)\%$ confidence interval for ξ.

To calculate such limits it is necessary to have tables of percentage points of the distribution of $T_{(f)}$, among which the following are available:

For $f(\cdot) = (1 - n^{-1})^{-1/2}S$, as described in Chapter 28 (tables of distribution).

For $f(\cdot) = W$ [in Lord (1947)].

For $f(\cdot) = M$ [in Herrey (1965)].

If such tables are not available, approximations may be used, for example, the approximations to the distributions of M and W to be described in Section 8.2; alternatively, they may be determined through Monte Carlo simulations.

8.2 Estimation of σ

The maximum likelihood estimator of a (ξ not being known) is

$$S = \left[n^{-1} \sum_{j=1}^{n} \left(X_j - \bar{X} \right)^2 \right]^{1/2}. \tag{13.78}$$

If ξ is known, the maximum likelihood estimator is

$$\left[n^{-1} \sum_{j=1}^{n} (X_j - \xi)^2 \right]^{1/2} = \left[S^2 + (\bar{X} - \xi)^2 \right]^{1/2}. \tag{13.79}$$

It is, however, very unusual to know ξ exactly, and we will not discuss this estimator further, except to note that neither (13.78) nor (13.79) is an unbiased estimator of a. In fact

$$E[S] = \frac{\sigma(2/n)^{1/2}\Gamma(n/2)}{\Gamma[(n-1)/2]}$$

$$= \frac{\sigma}{a_n} \tag{13.80}$$

and

$$E\left[\left\{ S^2 + (\bar{X} - \xi)^2 \right\}^{1/2} \right] = \frac{\sigma(2/n)^{1/2}\Gamma[(n+1)/2]}{\Gamma(n/2)}.$$

To obtain an unbiased estimator of σ, we must multiply S by a,. A few values of a, are shown in Table 13.4. Values of $a'_n = a_n\sqrt{1 - n^{-1}}$, such that $a'_n E[\{(n-1)^{-1}\Sigma(X_j - \bar{X})^2\}^{1/2}] = \sigma$, are also shown.

Table 13.4   Multipliers a,, a'_n Such That $E[a_n S] =$ $\sigma = E[a'_n \sqrt{V}]$

n	a_n	a'_n
2	1.77245	1.25331
3	1.38198	1.12838
4	1.25331	1.08540
5	1.18942	1.06385
6	1.15124	1.05094
7	1.12587	1.04235
8	1.10778	1.03624
9	1.09424	1.03166
10	1.08372	1.02811

For n greater than 10, the formulas

$$a_n \doteq 1 + \tfrac{3}{4}(n - 1)^{-1},$$

$$a'_n \doteq 1 + \tfrac{1}{4}(n - 1)^{-1}$$

give quite good approximations.

$$V = n(n - 1)^{-1}S^2 = (n - 1)^{-1} \sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2 \qquad (13.81)$$

is an unbiased estimator of u^2. Jarrett (1968) has given an interesting historical account of tables of these multiplying factors. [See also Cureton (1968) and Bolch (1968).] Note that the value of a minimizing the *mean square error* of aS^2 as an estimator of σ^2 is $(n^{-1} + 1)^{-1}$. The value of b minimizing the mean square error of bS as an estimator of σ is a'_{n+1} [Markowitz (1968)]. Iliescu and Voda (1974) have discussed the estimation of σ in detail.

The variance of the unbiased estimator, $a_n S$, of σ is

$$\text{Var}(a_n S) = \left[a_n^2(1 - n^{-1}) - 1\right]\sigma^2. \qquad (13.82)$$

The variance of V is

$$\text{Var}(V) = 2(n - 1)^{-1}\sigma^4. \qquad (13.83)$$

Unbiased estimator of σ can also be obtained by multiplying the mean deviation (M) and the range (W) by appropriate factors (depending of course on n). The resulting unbiased estimators $b_n M, c_n W$ are identical with $a_n S$ for $n = 2$, and have greater variances than $a_n S$ for $n > 2$. Values of b_n

can be calculated from the simple formula $b_n = \sqrt{(\pi/2)n(n-1)^{-1}}$. Values of c, for n = 2(1)20 are given in **Pearson** and **Hartley** (1948).

Relative efficiencies (inverse ratio of variances) of $b_n M, c_n W$ and other unbiased estimators of σ compared with $a_n S$, are shown in Table 13.7 on page 136. From this table it can be seen that the estimator based on range is (slightly) more efficient than that based on mean deviation for n = **4, 5,** but less efficient for n ≥ 6. (For n = 2,3 the two estimators give identical estimators of σ.)

The formula for b_n, quoted above, follows from

$$E[M] = \sigma\sqrt{\left(\frac{2}{\pi}\right)(1 - n^{-1})} . \qquad (13.84)$$

The variance of M is

$$\text{Var}(M) = \frac{2\sigma^2}{n\pi}\left(1 - \frac{1}{n}\right)\left\{\frac{1}{2}\pi + \sqrt{n(n-2)} - n + \sin^{-1}\frac{1}{n-1}\right\}. \qquad (13.85)$$

For n ≥ 5, a very good approximation (error no more than about 0.00001) is

$$\sigma^2\,\text{Var}(M) \doteq n^{-1}(1 - 2\pi^{-1})(1 - 0.12n^{-1}). \qquad (13.86)$$

Approximate formulas for the moment ratios of M are

$$\beta_1(M) \doteq 1.05n^{-1},$$
$$\beta_2(M) = 3 + 0.91n^{-1}.$$

Godwin and **Hartley (1945)** calculated a table of the probability integral of the distribution of M, giving $\Pr[M \le mu]$ to 5 decimal places for m = 0.00(0.02)3.00 and n = 2(1)10, and also multipliers M_α for percentile points such that

$$\Pr[M \le M_\alpha \sigma] = a$$

to 3 decimal places, for n = 2(1)10 and a = 0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.9, 0.95, 0.975, 0.99, 0.995, and 0.999. For n = 10 there are also given approximate values calculated from the formula

$$M_\alpha \doteq E\left[\frac{M}{\sigma}\right] + U_\alpha\left(\sqrt{\text{Var}\left(\frac{M}{\sigma}\right)}\right). \qquad (13.87)$$

Although the upper and lower **2.5%,** 5% and 10% values are not too inaccurate, approximation is poor for the more extreme values. A better

approximation was obtained by **Cadwell (1953),** by regarding $(M/\sigma)^{1.8}$ as having (approximately) a $c\chi_\nu^2$ distribution with

$$\log c = -\log 2 - 1.8\left[\log\left(\frac{5}{9} + \frac{\nu}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right)\right.$$

$$\left. - \frac{1}{2}\log\left\{\left(\frac{2}{\pi}\right)(1 - n^{-1})\right\}\right], \quad (13.88)$$

$$\nu = \nu_0 + 0.196 - 0.159\nu_0^{-1},$$

$$\nu_0 = \frac{0.617[E[M]]^2}{\mathrm{Var}(M)}.$$

The approximation was obtained by first finding values of A, c, and ν to make $(M/\sigma)^\lambda$ and $c\chi_\nu^2$ have the same first three moments. The values of course depend on n; they are shown in Table 13.5 [based on Cadwell (1953)]. This table also gives the results of similar calculations for the range W. It can be seen that for M, the values of **A** do not vary much, once n exceeds 5, say. An "average" value of 1.8 was chosen; the values of ν and c were then adjusted to make the first **two** moments of $(M/\sigma)^{1.8}$ and $c\chi_\nu^2$ agree. It might

Table 13.5   Values of ν, A, and log c Such That First Three Moments of $(T/\sigma)^\lambda$ and $c\chi_\nu^2$ Agree

n	T = Range			T = Mean Deviation		
	v	A	$\log_{10} c$	v	A	$\log_{10} c$
2	1.00	2.00	0.3010	1.00	2.00	$\bar{1}.6990$
3	2.05	1.96	0.2368	2.05	1.96	$\bar{1}.4234$
4	3.20	1.90	0.1547	3.35	1.84	$\bar{1}.2370$
5	4.5	1.83	0.0607	4.6	1.80	$\bar{1}.1149$
6	6.0	1.75	$\bar{1}.9588$	5.9	1.77	$\bar{2}.0177$
7	7.7	1.67	$\bar{1}.8574$	7.2	1.74	$\bar{2}.9388$
8	9.5	1.60	$\bar{1}.7685$	8.4	1.74	$\bar{2}.8765$
9	12.0	1.51	$\bar{1}.6457$	9.6	1.74	$\bar{2}.8222$
10	14.0	1.46	$\bar{1}.5785$	11.0	1.72	$\bar{2}.7674$
12	19.0	1.36	$\bar{1}.4252$	13.0	1.74	$\bar{2}.6964$
14	26.0	1.24	$\bar{1}.2479$	16.0	1.70	$\bar{2}.6126$
16	34.0	1.14	$\bar{1}.0964$	18.0	1.72	$\bar{2}.5617$
18	46.0	1.03	$\bar{2}.9142$	21.0	1.70	$\bar{2}.4988$
20	60.0	0.94	$\bar{2}.7569$	23.0	1.71	$\bar{2}.5491$

be thought that $A = 1.7$ would have been a better choice, but the chosen value of A ($= 1.8$) does give closer approximation for smaller values of n, without affecting the accuracy too severely for larger values of n.

In the same paper **Cadwell** discussed approximations to the distributions of the arithmetic mean of a number (k) of (independent) (M/σ)'s or (W/σ)'s. He also considered the distributions of the ratios

$$\frac{\max_j(M_j)}{\min_j(M_j)} \quad \text{and} \quad \frac{\max_j(W_j)}{\min_j(W_j)},$$

which are analogous to the ratios of maximum to minimum of a number of independent mean square deviations $(S_1^2, S_2^2, \ldots, S_k^2)$, each based on the same number of observed values. **Cadwell** gives tables of approximate upper 5% and 1% points of these statistics for $n = 3(1)10$ and $k = 2(1)12$ (for the 5% points, $n = 12, 15, 20, 30, 60$ are also included for the mean deviation ratio, and $n = 12, 15, 20$ for the range ratio). Similar tables for the statistics

$$\frac{\max_j(S_j^2)}{\min_{,}(S_j^2)}$$

are given by **Pearson** and **Hartley** (1948). They also provide some useful tables of values connected with the distribution of (M/σ), including the expected value, variance, β_1, and β_2 for $n = 2(1)20$, 30, 60, and the upper and lower 0.1, 0.5, 2.5, 5, and 10 percentage points for $n = 2(1)10$.

Among the few simple exact results concerning the distribution of range, we note the following:

For $n = 2$, $E[W] = 2\sigma/\sqrt{\pi}$; $\text{Var}(W) = 2\sigma^2(1 - 2\pi^{-1})$.
For $n = 3$, $E[W] = 3\sigma/\sqrt{\pi}$; $\text{Var}(W) = \sigma^2[2 - (9 - 3\sqrt{3})\pi^{-1}]$.
For $n = 4$, $E[W] = (3\sigma/\sqrt{\pi})(1 + 2\pi^{-1}\sin^{-1}(\tfrac{1}{3}))$.

Godwin (1949a) gives a number of other exact values of first and second moments.

Subsequently, quite extensive tables of the distribution and moments of (W/σ) have become available. A historical survey (up to 1960) of these tables has been given by Harter (1960), who also provided tables of percentage point multipliers, W_α, to 6 decimal places for $n = 2(1)20(2)40(10)100$ and $\alpha = 0.0001$, 0.0005, 0.001, 0.005, 0.01, 0.025, 0.05, 0.1(0.1)0.9, 0.95, 0.975, 0.99, 0.995, 0.999, 0.9995, and 0.9999. There are also tables of the expected value, variance, and β_1 and β_2 of (W/σ) to 8 (or more) significant figures for $n = 2(1)100$ in Harter (1960). **Pearson** and **Hartley** (1948) give tables of $\Pr[W \le w\sigma]$ to 4 decimal places for $n = 2(1)20$ and $w = 0.00(0.05)7.25$. They also give the upper and lower 0.1, 0.5, 2.5, 5, and 10 percentage points of the

distribution of (W/σ) to 2 decimal places, and expected value and variance (5 decimal places), β_1 (4 decimal places), and β_2 (3 decimal places) for $n = 2(1)20$, 30, and 60. More elaborate tables are also available in this context, and some numerical algorithms have also been developed; see Section 4 for details.

From Table 13.5 it can be seen that a single value of A might not be found such that $(W/\sigma)^\lambda$ is well approximated by a distribution of $c\chi_\nu^2$ (for suitably chosen c and ν) for a range of values of $n \leq 20$. Pearson (1952) and Cox (1949) have investigated this kind of approximation in some detail. From their investigations it appears that for smaller values of n (e.g., $n \leq 8$), an approximation of the form $c\chi_{\nu_2}$ is preferable (indeed, it is exact for $n = 2$); an approximation of the form $c\chi_\nu^2$ is better for larger values of n.

Using tables of percentage points of (M/σ), (W/σ), and χ^2 (see Chapter 18), it is possible to construct confidence intervals for σ by rewriting the equation

$$\Pr\left[T_{\alpha_1}\sigma < T < T_{1-\alpha_2}\sigma | \sigma\right] = 1 - \alpha_1 - \alpha_2$$

in the form

$$\Pr\left[\frac{T}{T_{1-\alpha_2}} < \sigma < \frac{T}{T_{\alpha_1}} | \sigma\right] = 1 - \alpha_1 - \alpha_2, \qquad (13.89)$$

which shows that $(T/T_{1-\alpha_2}, T/T_{\alpha_1})$ is a $100(1 - a_1 - \alpha_2)\%$ confidence interval for a. Here T can be replaced by M, W, or $\sqrt{n}\,S$, and T_α by $(M/\sigma)_\alpha, (W/\sigma)_\alpha$, or $\sqrt{n}\,\chi_{n-1,\alpha}$, respectively.

The maximum likelihood estimator S of σ (like its unbiased counterpart $a_n S$) is not a linear function of the observed values of $X_1, X_2, \ldots, X_n$. It is, however, possible to construct a best linear unbiased estimator of a, using the order statistics $X'_1, X'_2, \ldots, X'_n$. Such estimators (using all the sample values) are of form

$$D = \sum_{j=1}^{[n/2]} \alpha_j\left(X'_{n-j+1} - X'_j\right). \qquad (13.90)$$

Values of α_j have been calculated for $n = 2(1)20$; they have been published in Sarhan and Greenberg (1962). Balakrishnan (1990) has presented similar tables for $n = 21(1)30(5)40$. The efficiency of D relative to $a_n S$ is always greater than 98%. Although this is very satisfactory, these estimators are not used often because it is just as easy to calculate $a_n S$ if all values are to be used, and this does not require such extensive auxiliary tables. If a linear estimator is desired (e.g., to reduce effects of inaccurate outlying observations), there are other linear estimators, nearly as efficient as D, but with simpler formulas for the coefficients.

We take especial note of *Gini's* mean difference

$$G = \binom{n}{2}^{-1} \sum_{\substack{i<j}}^{n} \sum^{n} |X_i - X_j|$$

$$= \frac{4}{n(n-1)} \sum_{j=1}^{[n/2]} \left\{ \frac{1}{2}(n+1) - j \right\} (X'_{n-j+1} - X'_j). \qquad (13.91)$$

We have

$$E[G] = \left(\frac{2}{\sqrt{\pi}} \right) \sigma,$$

$$\text{Var}[G] = 4[n(n-1)]^{-1} \left[\frac{1}{3}(n+1) + 2\pi^{-1}(n-2)\sqrt{3} \right.$$

$$\left. - 2\pi^{-1}(2n-3) \right] \sigma^2.$$

The statistic $\frac{1}{2}\sqrt{\pi}\,G$ is an unbiased estimator of σ. The first three lines of Table 13.7 [taken from Nair (1949)] show the efficiencies of $M\sqrt{\pi/2}$, $\frac{1}{2}\sqrt{\pi}\,G$ and D, the best linear unbiased estimator of σ, for n = 2(1)10. It can be seen that $\frac{1}{2}\sqrt{\pi}\,G$ is very nearly as efficient as D. As n tends to infinity, the efficiency of $\frac{1}{2}\sqrt{\pi}\,G$, relative to $a_n S$, tends to

$$\left(\tfrac{2}{3} + 4\sqrt{3} - 8 \right)^{-1} = 97.8\%.$$

The asymptotically efficient estimator obtained by the method of Chernoff, Gastwirth, and Johns (1967) (see Section 2 of Chapter 12) is obtained by putting α_j in (13.90) equal to $U_{j/(n+1)}$ for all j.

We also mention an estimator of σ suggested by Gupta (1952) as an approximation to D. This is obtained by replacing the coefficients α_j in (13.90) by

$$a_j = \frac{E[U'_{n-j+1}]}{\sum_{j=1}^{n} \{E[U'_j]\}^2}, \qquad (13.92)$$

where $U'_1 \leq U'_2 \leq \cdots \leq U'_n$ are order statistics corresponding to n indepen-

dent unit normal variables so that the estimator is

$$D' = \frac{\sum_{j=1}^{n} X_j' E\left[U_j'\right]}{\sum_{j=1}^{n} \left\{E\left[U_j'\right]\right\}^2}. \tag{13.93}$$

For large n this estimator is very nearly the same as the asymptotically efficient estimator just described. Shapiro and Wilk (1965) described estimators similar to D', except that the ratio of a', to the remaining α''s is modified.

In view of the accuracy attainable with $\frac{1}{2}\sqrt{\pi}\,G$, it does not seem necessary to consider the use of D'. However, we note an estimator proposed by Mead (1966) which is based on the same general idea as Gupta's estimator and may be convenient to apply in special circumstances (e.g., using certain kinds of rapid measuring apparatus). Suppose that the number of observations (n) is a multiple of m, say, km. The data are then sorted into k groups so that the m least values are in the first group, the next m least values in the second group, and so on; the last group consists of the m greatest values. If the unit normal distribution is truncated between u_{i-1} and u_i $(> u_{i-1})$, then (see Section 10.1) the expected value of the truncated distribution is

$$\lambda_{(i)} = \frac{Z(u_i) - Z(u_{i-1})}{\Phi(u_i) - \Phi(u_{i-1})}.$$

Mead's estimator is then

$$\frac{\sum_{i=1}^{n} \lambda_{(i)}[\text{Mean of the ith group}]}{\sum_{i=1}^{k} \lambda_{(i)}^2}, \tag{13.94}$$

with

$$u_0 = -\infty,$$

$$u_i = U_{i/k}, \qquad i = 1, \ldots, (k-1),$$

$$u_k = \infty.$$

(Note that $\Phi(u_i) - \Phi(u_{i-1}) = k^{-1}$ for all i.) Denoting the mean of the ith group by $\bar{Y}_i$, the estimators are as follows: for k = 2, 0.62666 $(\bar{Y}_2 - \bar{Y}_1)$; for k = 3, 0.45838 $(\bar{Y}_3 - \bar{Y}_1)$; for k = 4, 0.36927 $(\bar{Y}_4 - \bar{Y}_1) + 0.09431 (\bar{Y}_3 - \bar{Y}_2)$; for k = 5, 0.31213 $(\bar{Y}_5 - \bar{Y}_1) + 0.11860 (\bar{Y}_4 - \bar{Y}_2)$; and for k = 6, 0.27177 $(\bar{Y}_6 - \bar{Y}_1) + 0.12373 (\bar{Y}_5 - \bar{Y}_2) + 0.03844 (\bar{Y}_4 - \bar{Y}_3)$. Mead obtained these

values for the asymptotic efficiency of this estimator (n large) relative to $a_n S$:

k	2	3	4	5	6
Asymptotic Efficiency (%)	87.6	93.3	95.6	96.8	97.8

Yet another class of linear estimators has been studied by Oderfeld and **Pleszczyńska** (1961) and **Pleszczyńska** (1965). These are linear functions of the order statistics $Y_1', \ldots, Y_n'$ corresponding to the random variables $Y_j = |X_j - \bar{X}|$. These authors estimated the values of the coefficients in

$$\sum_{j=1}^{n} \alpha_j Y_j'$$

on an empirical basis, using result of sampling experiments. In the first paper the only value used for n was 5, and the suggested estimator was

$$-0.065 Y_1' + 0.150 Y_2' + 0.175 Y_3' + 0.312 Y_4' + 0.405 Y_5'.$$

In the second paper the smallest absolute deviate was omitted (i.e., a, taken equal to zero). Coefficients (α_j) were estimated for n = 3(1)10. In all cases the largest coefficient was α_n, indicating the relative importance of extreme observations in estimating σ. These estimators appear (from empirical evidence) to have quite high (over 95%) efficiencies relative to $a_n S$, though they are no higher than, for example, estimators based on Gini's mean difference.

For $n \leq 10$ (at least) there are even simple unbiased estimators of σ that are not much less efficient than $\frac{1}{2}\sqrt{\pi}\, G$ or D. These are based on the thickened ranges

$$J_{(r)} = \sum_{j=1}^{r} \left(X_{n-j+1}' - X_j' \right) \tag{13.95}$$

[Jones (1946)]. Values of the multiplying factor to be applied to $J_{(2)}$ to make it an unbiased estimator of σ are shown in Table 13.6 [taken from Jones (1946)]. Relative efficiencies of these estimators (compared with $a_n S$) are also shown in Table 13.7. It can be seen that if $J_{(1)}$ ($\equiv W$) be used for $n \leq 5$ and $J_{(2)}$ for $6 \leq n \leq 10$, the relative efficiency never falls below 98%.

For large n Prescott (1968) has given the approximate formulas

$$E[J_{(r)}] = \frac{2r\sigma Z(U_p)}{p},$$

$$\mathrm{Var}(J_{(r)}) = \frac{2r^2\sigma^2}{np^2}\left[p + \{pU_p - Z(U_p)\}\{(1 - 2p)U_p + 2Z(U_p)\} \right],$$

Table 13.6   Multiplying Factor for $J_{(2)}$

n	Factor
4	0.37696
5	0.30157
6	0.26192
7	0.23702
8	0.21970
9	0.20684
10	0.19684

where $p = r/n$ is not too small. For n large the maximum efficiency (96.65%) is attained with $p = 0.225$. Prescott suggests using $p = \frac{1}{6}$, since the efficiency is still over 90% and the easily remembered quantity

$$\frac{1}{3r} J_{(r)}$$

is very nearly an unbiased estimator of a.

   Dixon (1960) has considered estimators of the form

$$k' \sum_j W_{(j)} = k' \sum_j (X'_{n+j+1} - X'_j), \tag{13.96}$$

where the summation $\sum_j$ is over some set of values of j. The statistic $W_{(j)}$ is

Table 13.7   Relative Efficiencies of Unbiased Estimators of σ

Estimator Based on		n							
	2	3	4	5	6	7	8	9	10
Mean deviation	100.0	99.19	96.39	94.60	93.39	92.54	91.90	91.4	91.0
Gini mean difference	100.0	99.19	98.75	98.50	98.34	98.24	98.16	98.1	98.1
Best linear	100.0	99.19	98.92	98.84	98.83	98.86	98.90	98.9	99.0
Range $(J_{(1)} \equiv W)$	100.0	99.19	97.52	95.48	93.30	91.12	89.00	86.9	85.0
Thickened range $\quad J_{(2)}$	—	—	91.25	93.84	95.71	96.67	96.97	96.8	96.4
$\quad J_{(3)}$	—	—	—	—	90.25	91.78	93.56	95.0	95.9
$\quad J_{(4)}$	—	—	—	—	—	—	89.76	90.7	92.2
$\quad J_{(5)}$	—	—	—	—	—	—	—	—	89.4

Table 13.8   Unbiased Estimators of a

n	Estimator	Efficiency (%) Relative to $a_n S$
11	$0.1608(W + W_{(2)} + W_{(4)})$	96.7
12	$0.1524(W + W_{(2)} + W_{(4)})$	97.2
13	$0.1456(W + W_{(2)} + W_{(4)})$	97.5
14	$0.1399(W + W_{(2)} + W_{(4)})$	97.7
15	$0.1352(W + W_{(2)} + W_{(4)})$	97.7
16	$0.1311(W + W_{(2)} + W_{(4)})$	97.5
17	$0.1050(W + W_{(2)} + W_{(3)} + W_{(5)})$	97.8
18	$0.1020(W + W_{(2)} + W_{(3)} + W_{(5)})$	97.8
19	$0.09939(W + W_{(2)} + W_{(3)} + W_{(5)})$	97.9
20	$0.10446(W + W_{(2)} + W_{(4)} + W_{(6)})$	98.0

sometimes called the jth quasi-range, occasionally the jth subrange. Evidently

$$J_{(r)} = \sum_{j=1}^{r} W_{(j)} \qquad (\text{and } W_{(1)} \equiv W). \tag{13.97}$$

Dixon found that for n ≤ 10 the most efficient estimators of form (13.96) are those based on the range ($W \equiv W_{(1)}$), or thickened range $J_{(2)}$ just described. For n = 11 − 20 he obtained the most efficient unbiased estimators [in the class (13.96)] given in Table 13.8. The efficiencies compare quite well with those of D, though they are not generally as high as those of Gini's mean difference.

Note that those unbiased linear estimators which are optimal (in various restricted classes) give large relative weight to the extreme observations. [Even S, which appears to be symmetrical, can be regarded (very roughly) as using weights proportional to the absolute magnitude of deviation from the sample mean.] Although we have obtained high efficiencies with these estimators, the calculations are all based on the complete validity of the normal distribution as applied to the data. Distributions of extreme order statistics are likely to be especially sensitive to departures from normality, and it is sometimes more important to guard against this possibility than to squeeze the last drop of formal "efficiency" from the data.

The simplest method of this kind uses only a single pair of symmetrically placed order statistics, in the form of a quasi-range, giving an estimator

$$\beta_r W_{(r)} = \beta_r (X'_{n-r+1} - X'_r).$$

For n large it is best to take $r \doteq 0.069n$. The efficiency (relative to $a_n S$) of the corresponding unbiased estimator of a is about 65% [Pearson (1920)]. [For estimating ξ by a statistic of form

$$\beta_{r/2}(X'_{n-r+1} + X'_r),$$

the best choice is $r \doteq 0.270n$, and the efficiency is about 67%.]

Kulldorff (1963, 1964) has studied the construction of estimators of this kind in some detail. He found that if estimators of form

$$\sum_{i=1}^{k} \beta_{r_i} W_{(r_i)} \qquad (13.98)$$

were considered, quite good results could be obtained by taking β_{r_i} proportional to i, giving estimators of form

$$\gamma \sum_{i=1}^{k} i W_{(r_i)}. \qquad (13.99)$$

For $k = 2$ Kulldorff found that the best values of r_1 and r, [subject to an estimator of form (13.99) being used] to take are $0.0235n$ and $0.1279n$, respectively, with $\gamma = 0.1174$; for $k = 3$ optimal values are $r_1 = 0.0115n$, $r, = 0.0567n$, and $r_3 \doteq 0.1704n$, with $\gamma = 0.0603$. The corresponding relative efficiencies are approximately 82% for $k = 2$ and 89% for $k = 3$. Note that these results apply to "large" samples (large values of n) and cannot be expected to apply when $n \leq 20$, as in the discussion of other estimators.

Using the large-sample approximations, Eisenberger and Posner (1965) have constructed "best linear unbiased estimators" of mean and standard deviation using only a fixed number (k) of quantiles, and excluding quantiles more extreme than 1 and 99%, or 2.5 and 97.5%, for $k = 2(2)20$. They also give (for the same values of k) pairs of linear estimators minimizing

(Variance of estimator of mean)

$+ \lambda($Variance of estimator of standard deviation)

for $\lambda = 1$, 2, and 3. Formulas appropriate for censored samples (described in Section 8.4) can also be used for complete samples if certain observed values must be ignored.

Ogawa (1976) has commented on the optimal spacing of the systematic statistics for the estimation of normal parameters. Ogawa (1977) later presented optimal spacings for the simultaneous estimation of ξ and a based on selected two sample quantiles. Cheng (1980) discussed the asymptotic best linear unbiased estimator of ξ from a censored sample. Fattorini (1978) gave percentile estimators for ξ and a, while Cheng and Ling (1983) discussed the

best linear unbiased estimators based on incomplete samples. Miyakawa, Tamiya, and Kotani (1985b) addressed the question of whether optimal spacings (in the case of small samples) for the estimation of ξ and σ should be symmetric. Balakrishnan and Cohen (1991) have presented a detailed account of the optimal linear estimation problem.

For the linear estimators based on order statistics (apart from those based on W alone) there are no easily available tables of percentage points. Such tables would be needed to construct confidence intervals for σ by rearranging the relation

$$\Pr\left[T_{\alpha_1}\sigma < T < T_{1-\alpha_2}\sigma|\sigma\right] = 1 - \alpha_1 - \alpha_2$$

in the form

$$\Pr\left[\frac{T}{T_{1-\alpha_2}} < \sigma < \frac{T}{T_{\alpha_1}}\middle|\sigma\right] = 1 - \alpha_1 - \alpha_2$$

[see (13.89)]. Even for those cases where such tables are available (M and W) only symmetrical intervals (with $a_r = a_r$) are practically useful. Except in connection with intervals based on S, no attempt has been made to construct shortest confidence intervals for σ. Nair (1947) suggested calculation of upper and lower 1% and 5% points of the distribution of the second thickened range $J_{(2)}$, but such tables have not been published. In many cases an approximation using a distribution of the form of that of a multiple of a chi random variable $(c\chi_\nu)$ (Chapter 18) may give results that are not seriously inaccurate for practical purposes.

8.3 Estimation of Functions of ξ and a

Certain functions of both ξ and σ are sometimes the primary target of statistical estimation. Among these we particularly note the $100\alpha\%$ percentile point $((+ U_\alpha\sigma)$ and the proportion (of population) less than a fixed number x,

$$P_x = \left(\sqrt{2\pi}\,\sigma\right)^{-1}\int_{-\infty}^{(x-\xi/\sigma}e^{-u^2/2}\,du. \tag{13.100}$$

These quantities may be estimated by general methods, not using the special form of distribution. However, when the validity of the assumption of normality has been clearly established, it is to be expected that more accurate estimates can be obtained by using this knowledge of the form of distribution.

Evidently, if $\hat{\xi}', \hat{\sigma}'$ are any unbiased estimators of ξ, σ, respectively, then $(\hat{\xi}' + U_\alpha\hat{\sigma}')$ is an unbiased estimator of $\xi + U_\alpha\sigma$. If $\overline{X}$ is used as the estimator of ξ and if $\hat{\sigma}'$ is any of the unbiased estimators of σ described in

Section 8.2, then, since $\overline{X}$ and $\hat{\sigma}'$ are independent,

$$\text{Var}\left(\overline{X} + U_\alpha \hat{\sigma}'\right) = \sigma^2 n^{-1} + U_\alpha^2 \text{Var}(\hat{\sigma}'). \qquad (13.101)$$

Combination of $\overline{X}$ with the best linear unbiased estimator of σ (or one of the other, nearly as efficient, linear estimators of σ) will give a good linear estimator of $\xi + U_\alpha \sigma$.

If $a_n S$ is used as an estimator of σ, the distribution of the estimator $(\overline{X} + U_\alpha a_n S)$ may be evaluated in the following way:

$$\Pr\left[\overline{X} + U_\alpha a_n S \leq K\right]$$

$$= \Pr\left[\left(\xi + \frac{U\sigma}{\sqrt{n}}\right) + \left(\frac{U_\alpha a_n \sigma}{\sqrt{n}}\right)\chi_{n-1} \leq K\right]$$

$$= \Pr\left[\left\{U + \frac{\sqrt{n}\,(\xi - K)}{\sigma}\right\}\left(\frac{\chi_{n-1}}{\sqrt{n-1}}\right)^{-1} \leq -U_\alpha a_n \sqrt{n-1}\right]$$

$$= \Pr\left[t'_{n-1}\left(\frac{\sqrt{n}\,(\xi - K)}{\sigma}\right) \leq -U_\alpha a_n \sqrt{n-1}\right], \qquad (13.102)$$

where $t'_{n-1}(\lambda)$ denotes a noncentral t variable (see Chapter 31) with n − 1 degrees of freedom and a noncentrality parameter A.

If other estimators of σ are used, approximate results of similar form can be obtained by approximating the distribution of $\hat{\sigma}'$ by that of $c\chi_\nu$, with suitable values of c and ν. It will usually be troublesome to assess the accuracy of these approximations. If effects of unreliable outlying observations are to be specially avoided, then estimators of ξ and σ not using such observations may be used. However, if the reason for this precaution is that lack of normality is suspected, it is doubtful that $\xi + U_\alpha \sigma$ should be estimated at all.

Coming now to the estimation of quantities like

$$\Pr[X \leq x] = \left(\sqrt{2\pi}\right)^{-1} \int_{-\infty}^{(x-\xi)/\sigma} e^{-u^2/2}\,du,$$

it is clear that the maximum likelihood estimator is obtained by replacing ξ by $\overline{X}$, and σ by S. The resulting estimator is, in general, biased. (It is unbiased if it so happens that $x = \xi$.)

To obtain the minimum variance unbiased estimator, the Blackwell-Rao theorem may be used. The estimator

$$T = \begin{cases} 1 & \text{if } X_1 \le x, \\ 0 & \text{if } X_1 > x \end{cases} \tag{13.103}$$

is an unbiased estimator of $\Pr[X \le x]$ and $\bar{X}$ and S are jointly complete sufficient statistics for ξ and a. Hence the minimum variance unbiased estimator of $\Pr[X \le x]$ is

$$E[T|\bar{X}, S] = \Pr[X_1 \le x|\bar{X}, S]$$
$$= \Pr\left[\frac{X_1 - \bar{X}}{S} \lessgtr \frac{x - \bar{X}}{S}\bigg|\bar{X}, S\right].$$

Since the conditional distribution of $(X_1 - \bar{X})/S$ is independent of both $\bar{X}$ and S, it is the same as the unconditional distribution of $(X_1 - \bar{X})/S$. Making an orthogonal transformation with one new variable $\sqrt{n/(n-1)}\,(X_1 - \bar{X})$, and one equal to $\sqrt{n}\,\bar{X}$, it can be seen that $(X_i - \bar{X})/S$ is distributed symmetrically about zero as the signed square root of $(n-1)$ times a beta variable with parameters $\frac{1}{2}, \frac{n}{2} - 1$ (see Chapter 25). Hence the minimum variance unbiased estimator of $\Pr[X \le x]$ is

$$\begin{cases} \left[B\left(\frac{1}{2}, \frac{n}{2} - 1\right)\right]^{-1} \int_{-1}^{(x-\bar{X})/(S\sqrt{n-1})} (1 - v^2)^{(n-4)/2}\, dv & \text{for } |x - \bar{X}| \le S\sqrt{n-1}, \\ 0 & \text{for } x < \bar{X} - S\sqrt{n-1}, \\ 1 & \text{for } x > \bar{X} + S\sqrt{n-1}. \end{cases} \tag{13.104}$$

(Numerical evaluation can be effected using tables of the incomplete beta function, as described in Chapter 25.)

At this point we note that if X_{n+1} is independent of, and has the same distribution as each X_j, $(X_{n+1} - \bar{X})/S$ is distributed as $[(n+1)/(n-1)]^{1/2}$ times t with $n - 1$ degrees of freedom. Hence the interval

$$\left(\bar{X} + t_{n-1,\alpha_1}\left[\frac{n+1}{n-1}\right]^{1/2} S, \ \bar{X} + t_{n-1,1-\alpha_2}\left[\frac{n+1}{n-1}\right]^{1/2} S\right)$$

contains on average a proportion $1 - \alpha_1 - a$, of the population values. It is thus a form of tolerance interval for the normal distribution. Unlike the tolerance intervals described in Chapter 12, the construction of this interval

makes use of knowledge of the form of population distribution. It cannot be used for other populations without the possibility of introducing bias.

Wald and Wolfowitz (1946) have shown that a good approximation to tolerance limits, such that there is a probability equal to $1 - a$ that the limits include at least a specified proportion γ of the population, is

$$\overline{X} \pm \frac{\lambda_\gamma \sqrt{n}\, S}{\chi_{n-1,\alpha}}, \tag{13.105}$$

where λ_γ satisfies the equation

$$\Phi\left(n^{-1/2} + \lambda_\gamma\right) - \Phi\left(n^{-1/2} - \lambda_\gamma\right) = y.$$

The construction of exact one-sided tolerance limits can be simply effected, using the noncentral t distribution (see Chapter 31). We note that the population proportion less than $(\overline{X} + kS)$ is

$$\Phi\left(\frac{\overline{X} + kS - \xi}{\sigma}\right),$$

and this is at least γ if

$$\frac{\overline{X} + kS - \xi}{\sigma} \geq U_\gamma.$$

This inequality can be rearranged in the form

$$\frac{\sqrt{n}\left\{(\overline{X} - \xi)/\sigma\right\} - \sqrt{n}\, U_\gamma}{\sqrt{n}\,(S/\sigma)/\sqrt{n-1}} \geq -k\sqrt{n-1}. \tag{13.106}$$

The statistic on the left-hand side of (13.106) has a noncentral t distribution with $(n - 1)$ degrees of freedom and noncentrality parameter $(-\sqrt{n}\, U_\gamma)$. In order that the probability that at least a proportion γ of the population is less than $(\overline{X} + kS)$, should be equal to $(1 - a)$, we make $(-k\sqrt{n - 1})$ equal to the lower $100\alpha\%$ point of the noncentral t distribution, i.e.

$$k = \frac{-t'_{n-1,\alpha}\left(-\sqrt{n}\, U_\gamma\right)}{\sqrt{n-1}} \tag{13.107}$$

Nelson (1977) has discussed tolerance factors for the normal distribution. Odeh and Owen (1980) have presented elaborate tables of normal tolerance limits. Gerisch, Struck, and Wilke (1987) determined one-sided tolerance limit factors in the case of censored samples through Monte Carlo simula-

tions. Mee (1988) discussed the estimation of the percentage of a normal distribution lying outside a specified interval. Eberhardt, Mee, and Reeve (1989) determined factors for exact two-sided tolerance limits for the normal distribution.

Sometimes it is desired to estimate the mean square error $(\xi - \xi_0)^2 + \sigma^2$, where ξ_0 is a specified number. The mean square

$$n^{-1} \sum_{j=1}^{n} (X_j - \xi_0)^2$$

is an unbiased estimator of this quantity. It is distributed as

$$n^{-1}\sigma^2 \times \left\{ \begin{array}{l} \text{Noncentral } \chi^2 \text{ with} \\ \text{n degrees of freedom and} \\ \textbf{noncentrality} \text{ parameter} \\ n(\xi - \xi_0)^2/\sigma^2 \end{array} \right\}$$

(see Chapter 29) and has variance

$$2n^{-2}\sigma^4\left[n + \frac{2n(\xi - \xi_0)^2}{\sigma^2}\right] = 2n^{-1}\sigma^4\left[1 + \frac{2(\xi - \xi_0)^2}{\sigma^2}\right].$$

A natural estimate of the coefficient of variation (σ/ξ) is the ratio $a_n S/\overline{X}$, or more generally $a'_n S/\overline{X}$, with a', being suitably chosen. Since the expected value of $S/\overline{X}$ is infinite, it is not possible to obtain an unbiased estimator of this form. We can, however, construct an approximate confidence interval for σ/ξ. We will suppose that $\Pr[\overline{X} < 0]$ can be neglected (i.e., σ/ξ sufficiently small—less than $\frac{1}{4}$, say). Then, since $\overline{X}/S$ is distributed as $(n - 1)^{-1/2}$ times noncentral t with $n - 1$ degrees of freedom and noncentrality parameter $\sqrt{n}\,\xi/\sigma$, it follows that (in the notation of Chapter 31)

$$\Pr\left[t'_{n-1,\alpha_1}\left(\frac{\sqrt{n}\,\xi}{\sigma}\right) \le \frac{\sqrt{n-1}\,\overline{X}}{S} \le t'_{n-1,1-\alpha_2}\left(\frac{\sqrt{n}\,\xi}{\sigma}\right)\right] = 1 - \alpha_1 - \alpha_2$$

or

$$\Pr\left[n^{-1/2}g_{1-\alpha_2}\left(\frac{\sqrt{n-1}\,\overline{X}}{S}\right) \le \frac{\xi}{\sigma} \le n^{-1/2}g_{\alpha_1}\left(\frac{\sqrt{n-1}\,\overline{X}}{S}\right)\right] \doteq 1 - \alpha_1 - \alpha_2,$$

$$(13.108)$$

where $g_\alpha(z)$ is the solution (for g) of the equation

$$t'_{n-1,\alpha}(g) = z \tag{13.109}$$

(assuming that $\overline{X}$ is not too small). Assuming now that $\xi > 0$, (13.108) can be rewritten

$$\Pr\left[\sqrt{n}\big/g_{\alpha_1}\left(\frac{\sqrt{n-1}\,\overline{X}}{S}\right) < \frac{\sigma}{\xi} < \sqrt{n}\big/g_{1-\alpha_2}\left(\frac{\sqrt{n-1}\,\overline{X}}{S}\right)\right] \doteq 1 - \alpha_1 - \alpha_2.$$

It is necessary to use tables of the noncentral t distribution (see Chapter 31, Section 7) to calculate even these approximate limits.

More easily calculable, but rather rough, approximate limits are obtained from the formula

$$
\begin{aligned}
\textbf{Lower limit} &= V\left[1 - n^{-1/2}U_{\alpha_2}\sqrt{\tfrac{1}{2} + V^2}\right]^{-1}, \\
\text{Upper limit} &= V\left[1 - n^{-1/2}U_{1-\alpha_1}\sqrt{\tfrac{1}{2} + V^2}\right]^{-1},
\end{aligned}
\tag{13.110}
$$

where $V = S/\overline{X}$. These are based on the assumption that $(S - k\overline{X})$ is approximately normally distributed with expected value $(a - k\xi)$ and variance $n^{-1}\sigma^2(1 + \tfrac{1}{2}k^2)$ so that (since $\xi \gg a$)

$$\Pr\left[\frac{S}{\overline{X}} < k\right] \doteq \Phi\left(\sqrt{\frac{n}{1 + \tfrac{1}{2}k^2}}\left\{1 - \frac{k\xi}{\sigma}\right\}\right),$$

that is, $\sqrt{n/\left(1 + \tfrac{1}{2}V^Z\right)}\,(1 - V\xi/\sigma)$ has approximately a unit normal distribution. A similar argument indicates that if X_1, X_2 are independent normal random variables and $E[X_j] = \xi_j$, $\mathrm{Var}(X_j) = \sigma_j^2$ $(j = 1, 2)$ with $\xi_2 \gg \sigma_2$, then putting $X_1/X_2 = R$, the distribution of

$$\frac{R\xi_2 - \xi_1}{\left(R^2\sigma_2^2 + \sigma_1^2\right)^{1/2}}$$

is approximately unit normal.

Koopmans, Owen, and Rosenblatt (1964) have pointed out that if the distribution of each of the independent variables is *lognormal* (see Chapter 14), then construction of exact confidence intervals for the coefficient of variation is straightforward. Since it is possible to approximate a normal distribution quite closely by a lognormal distribution (see Chapter 14, Section 3), it is likely that the same formulas will give good results for normal variables (though they will not of course give exactly specified values for

confidence coefficients). The (approximate) confidence limits, in terms of the original variables $X_1, \ldots, X_n$ obtained by this method, are

$$\left[\exp\left\{ \frac{\sum_{j=1}^{n}\left(\log X_j - \overline{\log X}\right)^2}{\chi^2_{n-1, 1-\alpha_1}} \right\} - 1 \right]^{1/2} \qquad (13.111)$$

and

$$\left[\exp\left\{ \frac{\sum_{j=1}^{n}\left(\log X_j - \overline{\log X}\right)^2}{\chi^2_{n-1, \alpha_2}} \right\} - 1 \right]^{1/2},$$

where $\overline{\log X} = n^{-1}\sum_{j=1}^{n} \log X_j$.

The cumulative distribution function of the rth quasi-range for random samples from a unit normal distribution is [Jones et al. (1969)]

$$F_{W_{(r)}}(w) = \sum_{i=0}^{r} \frac{n^{(2r-i+1)}}{r!(r-i)!} \sum_{j=0}^{r-i} \sum_{k=0}^{n-2r+i-1} (-1)^{n-2r+i-1-j+k} P, \qquad w > 0,$$

where

$$P = \binom{r-i}{j}\binom{n-2r+i-1}{k}\Pr\left[\bigcap_{l=1}^{n-r+i+j-1} \left(Y_l \le \frac{\delta_{j+h-l}w}{\sqrt{2}} \right) \right]$$

and $6, = 0, 1$ for $h < 0, \ge 0$, respectively; the Y's are standardized multi-normal variables (Chapter 35) with all correlations equal to $\frac{1}{2}$.

Estimates for the reliability function of the normal distribution has been considered by Hurt (1980). Sinha (1985) has discussed Bayes's estimations of the reliability function.

Owen and Hua (1977) have presented tables of confidence limits on the tail area of the normal distribution. Fertig and Mann (1977) have discussed one-sided prediction intervals for at least p out of m future observations. Nelson and Schmee (1981) have given prediction limits for the last failure time of a normal sample from early failures. Lingappaiah (1983) has discussed prediction problems in normal samples. Along the lines of Fertig and Mann (1977), Chou and Owen (1986) have studied one-sided simultaneous lower prediction intervals for 1 future samples from a normal distribution. Whitmore (1986) has made some interesting comments on the prediction limits for a single normal observation. Odeh (1989a) has developed simultaneous two-sided prediction intervals to contain at least 1 out of k future means. Similarly Odeh (1989b) has also discussed simultaneous one-sided prediction intervals to contain all of k future means.