RANDOM MEASUREMENT ERRORS WITH CONTINUOUS PROBABILITY DISTRIBUTIONS

1. Statistical Basis of Measurement.

If we consider a sequence of n values $X_1, X_2, \ldots, X_n$ obtained from an instrument by repeated observations of the same constant physical quantity, the values generally will not be identical. This scatter or dispersion in successive observed values will be assumed due to independent random errors which are added to the constant physical quantity being measured. By independent we mean that the error in any observed value does not depend in any way on the errors in other observed values. (In some measuring instruments the successive observed values are not independent.)

If we take the mean of a large number n of observed values

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1.1}$$

and repeat this for other groups of n observed values it becomes evident that the dispersion of the values of $\bar{X}$ is considerably smaller than the dispersion of the values of X. This leads us to wonder whether the dispersion in $\bar{X}$ can be made as small as we wish by taking n sufficiently large and if so whether the limit

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = q \tag{1.2}$$

exists. We shall see that in many random processes the limit q does exist in a certain sense and that q can be equated to the true value of the constant physical quantity before addition of random errors.

The problem before the engineer is that of forming an estimate of the true value q by making a finite and reasonable number of measurements instead of the infinite number implied by (1.2). Such an estimate (for example, $\bar{X}$) will have a random error and will be of little value unless its dispersion is known. In the remainder of these notes we shall be concerned primarily with answering these two questions about a set of n measurements of a constant quantity where the errors are additive, independent, and random:

1. What is the best estimate of the true value that can be formed?
2. How good is this estimate?

2. Probability Distribution of Random Errors.

Any set of n measurements can be thought of as a small sample chosen at random from an arbitrarily large set of measurements that could have been made. This hypothetical large set of measurements is sometimes called the parent population from which the sample of n measurements was obtained. This parent population has a probability distribution which is obtained as follows. Choose a set of rectangular coordinate axes and represent any deviation x as a point on the x-axis in the usual manner. Divide the range of the deviations into a number of equal intervals of width Δx as shown in Figure 2.1. Then assign a y-value to each interval which is equal to the fraction of the measurements from the parent population having deviations that fall to the left of the interval.

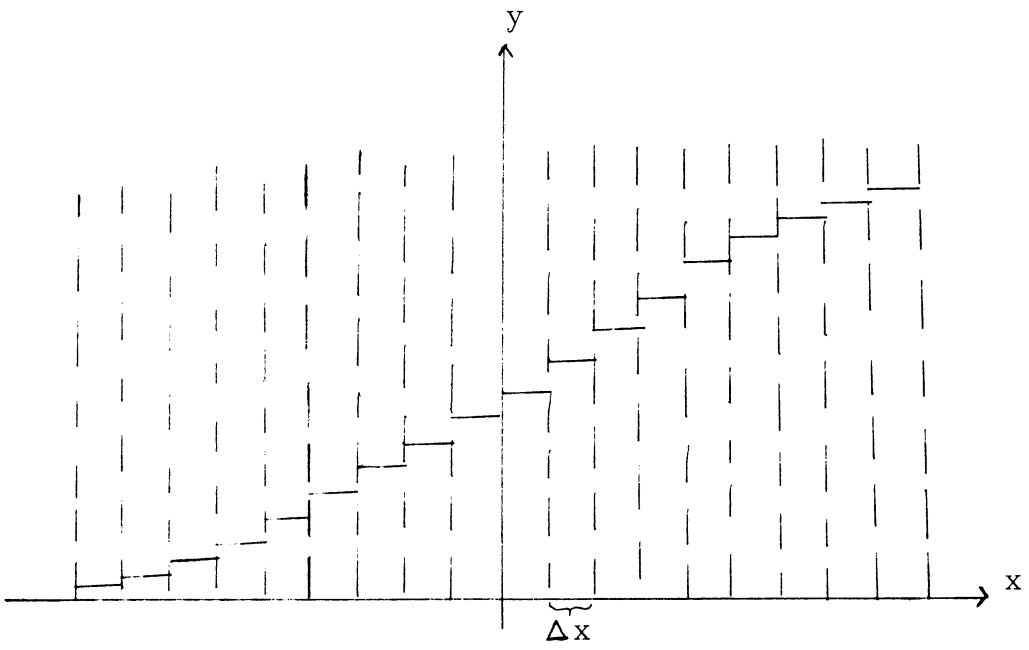


Figure 2.1

If we let Δx approach zero, the number of steps will become larger and approach, as a limit, the continuous curve indicated in Figure 2.2. This curve $y = F(x)$ is called the distribution function or cumulative distribution function (CDF). $F(x)$ is the probability that an observed random value X* is less than or equal to a prescribed x. This is written

*We use the symbol X to denote both an observed element of the parent population and a general element of the parent population.

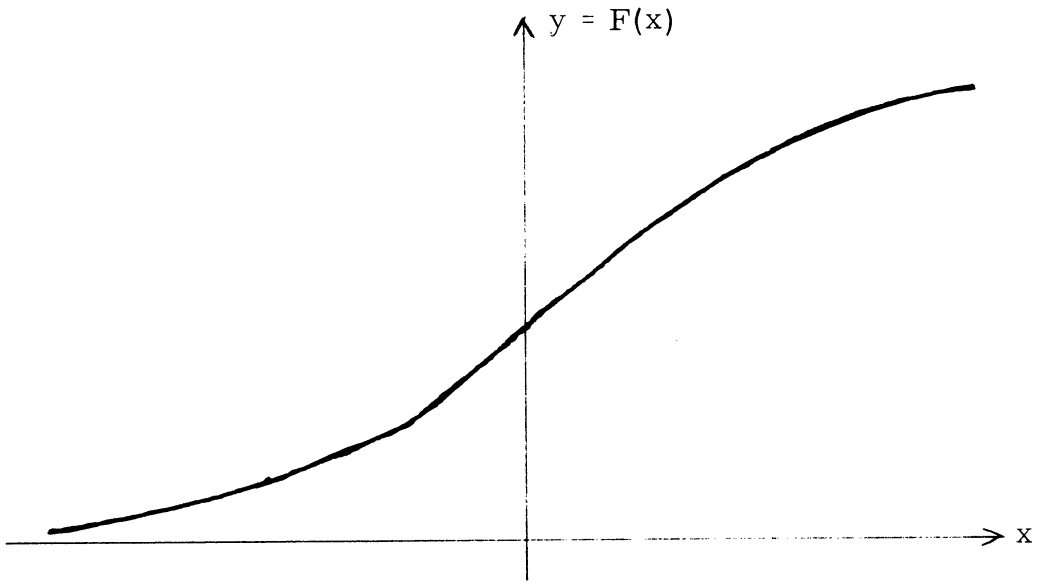


Figure 2.2

$$P(X \leq x) = F(x) \tag{2.1}$$

and for an interval

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1). \tag{2.2}$$

If we define

$$f(x) = \frac{d}{dx} F(x) \tag{2.3}$$

then

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x)\, dx \tag{2.4}$$

and

$$dP = P(x < X \leq x + dx) = f(x)\, dx. \tag{2.5}$$

f(x) is called the frequency function or probability density function (PDF). In the following we shall limit our consideration to random errors described by a PDF which exists and is finite. The PDF f(x) corresponding to the CDF F(x) of Figure 2.2 is indicated in Figure 2.3.

A little consideration of (2.1) will show that

$$\lim_{x \to 0} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1 \tag{2.6}$$

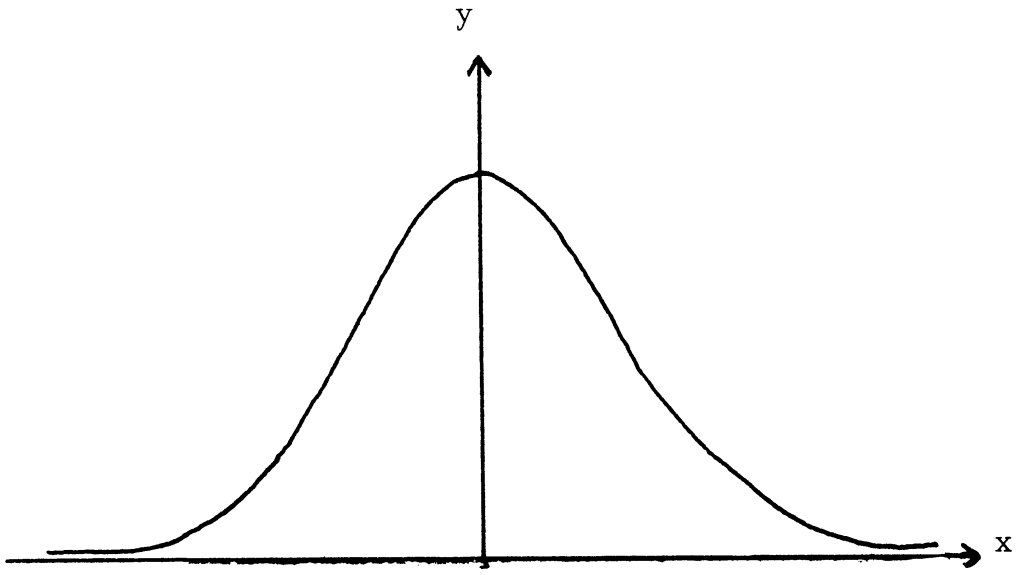


Figure 2.3

and therefore

$$\int_{-\infty}^{\infty} f(x)\,dx = 1. \qquad (2.7)$$

A PDF satisfying (2.7) is said to be normalized. Any function having a finite area over the infinite interval can be normalized by dividing its values by its area. Thus

$$f(x) = \frac{g(x)}{\int_{-\infty}^{\infty} g(x)\,dx}$$

is a normalized PDF if the denominator is finite.

3. Expected Values and Distribution Parameters.

The average or expected value **E**(X) of a measurement X with random error is found by taking the weighted sum of each possible value multiplied by its probability of occurrence. By (2.5)

$$dE(X) = x\,f(x)\,dx \qquad (3.1)$$

and

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx\ . \tag{3.2}$$

Similarly the average of any function $\phi(X)$ of X is given by the expected value

$$E[\phi(X)] = \int_{-\infty}^{\infty} \phi(x)\, f(x)\, dx. \tag{3.3}$$

E(X) given by (3.2) is called the mean of X and is usually designated by m_X. It can be recognized as the x-component of the centroid of the area under the PDF of Figure 2.3. A maximum of the PDF has a value of x called a mode and the value of x equally dividing the area under the PDF is called the median. The mean-square value of X with with respect to its mean value m_X is called the variance, denoted by σ_X^2.

$$\sigma_X^2 = E(X - m_X)^2 = \int_{-\infty}^{\infty} (x - m_X)^2\, f(x)\, dx \tag{3.4}$$

The square root of the variance σ_X is called the standard deviation and it is the root-mean-square value of X with respect to its mean. The standard deviation has the same physical dimensions as X and is the most common measure of the dispersion of X.

4. The Normal or Gaussian Distribution.

As an example of the definitions introduced in the preceding section we introduce a particular PDF describing the normal or Gaussian distribution. Many measurement problems with random errors involve this distribution which we shall return to in a later section.

The normal PDF is of the form

$$k\ e^{-h^2(x - c)^2} \tag{4.1}$$

where k, h, and c are parameters. The parameter k is determined by the normalization condition

$$k \int_{-\infty}^{\infty} e^{-h^2(x-c)^2}\ dx = 1.$$

The integral is evaluated by standard methods to give the result $k = h/\sqrt{\pi}$. Thus

$$f(x) = \frac{h}{\sqrt{\pi}} \; e^{-h^2(x-c)^2} \; . \qquad (4.2)$$

It can be seen that h is connected with the dispersion of the measurements, for when h is large the probability density is very small except close to $x = c$.

From the symmetry of $f(x)$ about $x = c$, it is clear that $m_X = c$ and this is verified by

$$m_X = E(X) = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} x \, e^{-h^2(x-c)^2} \, dx \qquad (4.3)$$

$$= \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} (y + c) \, e^{-h^2 y^2} \, dy = c.$$

The variance of the normal distribution is given by

$$\sigma^2 = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} (x - m)^2 \, e^{-h^2(x-m)^2} \, dx \qquad (4.4)$$

$$= \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} y^2 \, e^{-h^2 y^2} \, dy.$$

The integral is evaluated by standard methods to give

$$\sigma^2 = \frac{1}{2h^2} \; . \qquad (4.5)$$

Substituting for h in terms of σ, the normal PDF can be expressed as

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \; e^{-\frac{(x-m)^2}{2\sigma}} \; . \qquad (4.6)$$

The probable error p is the x-interval on each side of the mean such that the total interval $-p < x \leq p$ includes an area of 0.5 under the PDF; that is, the probability of occurence of an error with magnitude less than or equal to the probable irror is 0.5. Thus

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-p}^{p} e^{-\frac{x^2}{2\sigma^2}} dx = 0.5 \tag{4.7}$$

and the solution is approximately

$$p = 0.6745\sigma. \tag{4.8}$$

The probable error can be generalized to other probability distributions, but its value in terms of σ will differ from that of (4.8).

Figures 4.1 and 4.2 show respectively the normal CDF and PDF for mean zero and positive values of the argument. In effect, the function plotted in Figure 4.1 is

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \tag{4.9}$$

where

$$F(x) = \Phi\left(\frac{x-m}{\sigma}\right) . \tag{4.10}$$

Similarly, the function ploted in Figure 4.2 is

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{4.11}$$

where

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x-m}{\sigma}\right) . \tag{4.12}$$

The magnitudes of the precision indices $1/h$, σ, and p are indicated by dashed vertical lines in both figures.

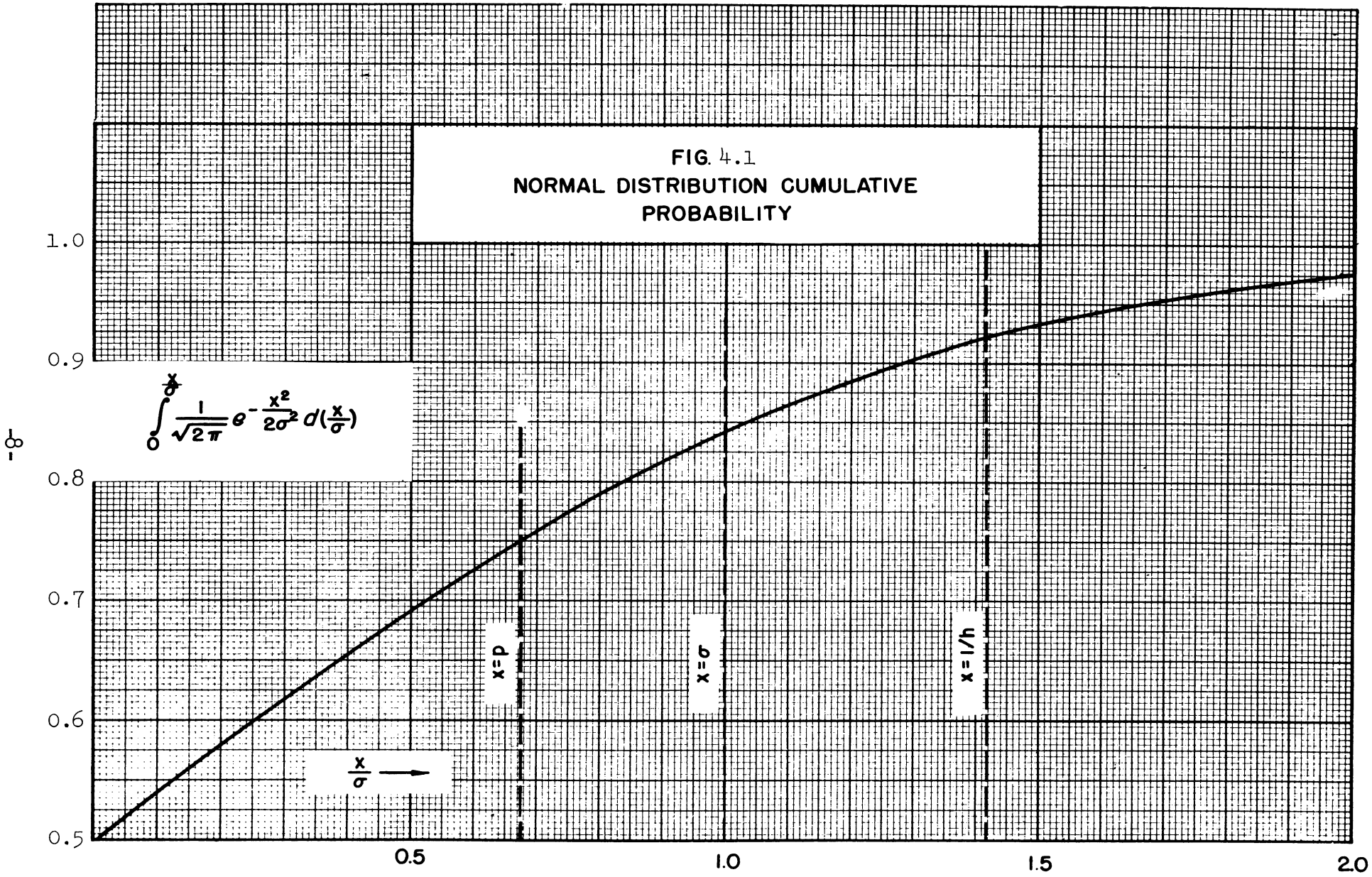


FIG. 4.1
NORMAL DISTRIBUTION CUMULATIVE PROBABILITY
1.0
0.9
0.8
0.7
0.6
0.5
0.5
1.0
1.5
2.0
$\int_0^{\frac{x}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} d(\frac{x}{\sigma})$
x=p
x=b
x=1/h
$\frac{x}{\sigma}$ ⟶

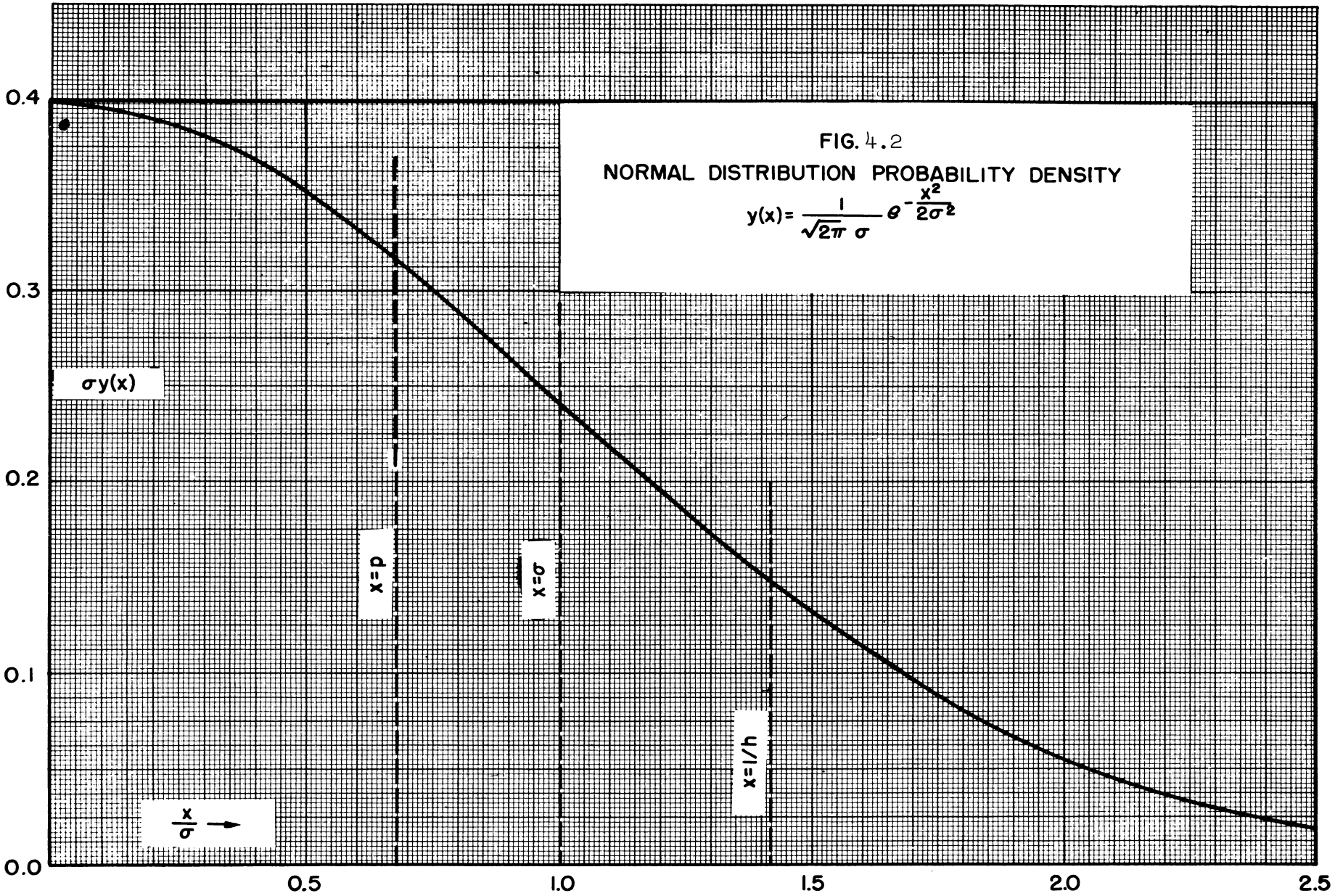


FIG. 4.2
NORMAL DISTRIBUTION PROBABILITY DENSITY
$y(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}}$
0.4
0.3
0.2
0.1
0.0
σy(x)
x = p
x = σ
x = 1/h
x/σ →
0.5
1.0
1.5
2.0
2.5

5. Joint Probability Distribution of Several Measurements

Let X and Y be the values from different measurement processes with respective PDF's $f_1(x)$ and $f_2(y)$. These two distribution functions cannot in general describe the relative behavior of X and Y; that is, the parent population consisting of all possible pairs of values (X, Y). To describe this parent population requires the joint CDF

$$P(X \leq x,\ Y \leq y) = F(x,y) \tag{5.1}$$

and the corresponding joint PDF f(x,y)

$$P(x < X \leq x + dx,\ y < Y \leq y + dy) = f(x,y)\, dxdy. \tag{5.2}$$

Clearly

$$\lim_{x,y \to \infty} F(x,y) = 1 \tag{5.3}$$

and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dxdy = 1 \tag{5.4}$$

is the normalization condition.

The necessary and sufficient condition for the independence of the errors in the measurement processes X and Y is

$$f(x,y) = f_1(x)\, f_2(y) \tag{5.5}$$

where $f_1(x)$ and $f_2(y)$ are the respective marginal distributions (ordinary PDF's) given by

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy \tag{5.6}$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)dy. \tag{5.7}$$

6. Expected Values of Several Measurements

The expected value of any function $\varphi(X,Y)$ is given by

$$E[\varphi(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x,y)\, f(x,y)dxdy \ . \tag{6.1}$$

Let

$$Z = aX + bY \tag{6.2}$$

where a and b are real constants. Then

$$E(Z) = E(aX + bY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax + by)f(x,y)dxdy$$

$$= a\int_{-\infty}^{\infty} xf_1(x)\, dx + b\int_{-\infty}^{\infty} yf_2(y)dy = aE(X) + bE(Y). \tag{6.3}$$

This distributive result for expected values immediately generalizes to linear combinations of n measurement processes

$$E(a_1 X_1 + \ldots + a_n X_n) = a_1 E(X_1) + \ldots + a_n E(X_n). \tag{6.4}$$

Let

$$Z = XY \ . \tag{6.5}$$

Then

$$E(Z) = E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)dxdy \ . \tag{6.6}$$

If X and Y are independent, them by (5.5)

$$E(XY) = \int_{-\infty}^{\infty} xf_1(x)dx \int yf_2(y)dy = E(X)\, E(Y). \tag{6.7}$$

For n independent measurements this generalizes to

$$E(X_1 X_2 \ldots X_n) = E(X_1)\, E(X_2) \ldots E(X_n). \tag{6.8}$$

7. Mean and Standard Deviation of Several Measurements

It follows directly from (6.4) that if

$$Z = X_1 + \ldots + X_n \tag{7.1}$$

Then

$$m_Z = m_{X_1} + \ldots + m_{X_n} . \tag{7.2}$$

Consider

$$W = X + Y . \tag{7.3}$$

$$\sigma_W^2 = E(W-m_W)^2 = E(W^2) - m_W^2 = E(X-m_X)^2 + E(Y-m_Y)^2$$

$$+ 2E(XY) - 2m_X m_Y . \tag{7.4}$$

If X and Y are independent, then by (6.8)

$$E(XY) = m_X m_Y \tag{7.5}$$

and

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 . \tag{7.6}$$

For $X_1, \ldots, X_n$ independent this generalizes to

$$\sigma_{X_1+\ldots+X_n}^2 = \sigma_{X_1}^2 + \ldots + \sigma_{X_n}^2 . \tag{7.7}$$

Now consider the mean of n independent measurement values from the same distribution

$$\bar{X} = \frac{X_1}{n} + \frac{X_2}{n} + \ldots + \frac{X_n}{n} \tag{7.8}$$

Then

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2} (\sigma_{X_1}^2 + \sigma_{X_2}^2 + \ldots + \sigma_{X_n}^2) . \tag{7.9}$$

Since

$$\sigma_{X_1}^2 = \sigma_{X_2}^2 = \ldots = \sigma_{X_n}^2 = \sigma_X^2 \tag{7.10}$$

$$\sigma^2_{\bar{X}} = \frac{\sigma^2_X}{n} \tag{7.11}$$

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} . \tag{7.12}$$

We have thus established the conjecture made in Section 1; namely, the dispersion of $\bar{X}$ can be made as small as we wish by taking n sufficiently large.

8. Estimates of Distribution Parameters from Observed Values

We shall assume that measurement devices are calibrated so that the mean of the distribution is the true value of the physical quantity; that is, the mean of the additive error is zero. Now, on the basis of n observed values how can the mean of the distribution be estimated. Clearly, one such estimate is the mean X of the observed values.

We shall call an estimate α_e of a distribution parameter α an unbiased estimate if and only if

$$E(\alpha_e) = E(\alpha) . \tag{8.1}$$

This means the average of a large number of unbiased estimates, based on a finite n, approaches the estimated parameter. Now the estimate $\bar{X}$ for m_X is unbiased, for

$$E(\bar{X}) = \frac{1}{n} [E(X_1) + \ldots + E(X_n)] = E(X) = m_X . \tag{8.2}$$

By (7.12) the standard deviation of $\bar{X}$ is the standard deviation of the distribution of X divided by $\sqrt{n}$. So we must estimate σ_X in order to estimate $\sigma_{\bar{X}}$. An estimate for σ^2_X would appear to be

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2, \tag{8.3}$$

the mean square deviation of the observed values from their mean. However s^2 is a biased estimate for σ_X^2. This is seen as follows,

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (X_i - m)^2 - (\bar{X} - m)^2 \tag{8.4}$$

and

$$E(s^2) = \frac{1}{n} \sum_{i=1}^{n} E(X_i - m)^2 - E(\bar{X} - m)^2 = \sigma_X^2 - \frac{1}{n} \sigma_X^2$$

$$= \frac{n-1}{n} \sigma_X^2 . \tag{8.5}$$

Thus an unbiased estimate for σ_X^2 is

$$\frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2. \tag{8.6}$$

Just as $\bar{X}$ has a dispersion described by $\sigma_{\bar{X}} = \sigma_X/\sqrt{n}$, the estimate (8.6) for σ_X^2 has a standard deviation which can be calculated when the distribution of X is known.

In the next section we shall consider the question of whether the above estimates are best estimates. To do this we must specify the distribution and a criterion for "best".

9. Maximum Likelihood Estimates

In order to estimate a distribution parameter from observed values in accordance with a specific criterion it is necessary to have the CDF or PDF of the distribution in terms of the parameters.

For the maximum likelihood estimate of a parameter α we require that the estimate α_e be chosen to maximize the probability of obtaining the observed values.

We apply this to the normal distribution by considering a set of n independent observed values $X_1, \ldots, X_n$. On the basis of an estimated mean m_e the probability that the first measurement is in the small interval of width ΔX at X_1 is, from (2.5) and (4.6),

$$P(m_e, X_1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_1-m_e)^2}{2\sigma^2}} \Delta X_1 . \qquad (9.1)$$

The probability of obtaining the set of n independent measurements in the order given is the product of the probabilities for obtaining each measurement,

$$P(m_e, X_1,\ldots,X_n) = \frac{\Delta X_1,..,\Delta X_n}{(2\pi)^{n/2}\sigma^n} e^{-\frac{\sum_{i=1}^{n}(X_i-m_e)^2}{2\sigma^2}} . \qquad (9.2)$$

Now m_e is chosen to make P a maximum. To find the maximum we solve $\partial P/\partial m_e = 0$ for m_e and leave it to the student to verify that for this value $\partial^2 P/\partial m_e^2$ 0.

The term before the exponential is a positive constant which we shall call C. Thus

$$\frac{\partial P}{\partial m_e} = Ce^{-\frac{\sum_{i=1}^{n}(X_i-m_e)^2}{2\sigma^2}} \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-m_e) \qquad (9.3)$$

The terms preceeding the final summation are all non-zero so $\partial P/\partial m_e = 0$ requires

$$\sum_{i=1}^{n}(X_i-m_e) = 0 \qquad (9.4)$$

or

$$m_e = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X} \ . \qquad (9.5)$$

Thus the maximum likelihood estimate of the distribution mean m (the true quantity measured) is the mean of the n observed measurements. Although the mean of the n measurements is the estimate of the true value, it is of course in general not equal to the true value which can only be obtained as the limiting mean of the infinite parent distribution.

The maximum likelihood estimate σ_e of the standard deviation σ is found in a similar manner by differentiating (9.6) with respect to σ_e.

$$P(\sigma_e, X_1, \ldots, X_n) = \frac{\Delta X_1 \ldots \Delta X_n}{(2\pi)^{n/2}} \cdot \frac{1}{\sigma_e^n} e^{-\frac{\sum_{i=1}^{n} (X_i - m)^2}{2\sigma_e^2}} \qquad (9.6)$$

$$\frac{\partial P}{\partial \sigma_e} = C \frac{1}{\sigma^{n+1}} e^{-\frac{\sum_{i=1}^{n} (X_i - m)^2}{2\sigma_e^2}} \left[\frac{\sum_{i=1}^{n} (X_i - m)^2}{\sigma^2} - n \right]. \qquad (9.7)$$

The terms preceeding the bracket are all non-zero so $\partial P / \partial \sigma_e = 0$ requires

$$\frac{\sum_{i=1}^{n} (X_i - m)^2}{\sigma_e^2} = n \qquad (9.8)$$

or

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^{n} (X_i - m)^2}{n}} \ . \qquad (9.9)$$

Unfortunatly this estimate cannot be formed from the $X_1, \ldots, X_n$ because it involves the mean of the distribution for which we have only the estimate $\bar{X}$. If m is replaced by $\bar{X}$ in (9.9) the result is the biased estimate of (8.3). The bias can be removed, as in (8.6), to

give

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}} \quad . \tag{9.10}$$

10. Calculation of Means and Standard Deviations

It is always possible to apply Equations (9.5) and (9.10) directly to the data to obtain estimates for the mean and standard deviation. However, it is often possible to save time by applying the following procedure.

Choose a trial mean $\bar{\bar{X}}$ which gives a set of trial $(X_i - \bar{\bar{X}})$. Since $\bar{\bar{X}}$ can be chosen judiciously, these differences are quickly formed and are smaller in magnitude that the original readings X_i. Now the mean $\bar{X}$ is obtained by

$$\bar{X} = \frac{\sum_{i=1}^{n} (X_i - \bar{\bar{X}})}{n} + \bar{\bar{X}} \quad . \tag{10.1}$$

It is easily verified that the estimate for the standard deviation is given by

$$\sigma_e^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{\bar{X}})^2}{n-1} - \frac{n(\bar{X} - \bar{\bar{X}})^2}{n-1} \quad . \tag{10.2}$$

A sample calculation for n = 10 follows, with the choice $\bar{\bar{X}} = 23.0$:

i	X_i	$(X_i - \bar{\bar{X}})$	$(X_i - \bar{\bar{X}})^2$
1	23.2	0.2	0.04
2	22.7	- 0.3	0.09
3	22.8	- 0.2	0.04
4	23.5	0.5	0.25

5	23.0	0.0	0.00
6	23.1	0.1	0.01
7	23.4	0.4	0.16
8	22.6	- 0.4	0.16
9	23.1	0.1	0.01
10	23.8	0.8	0.64
		1.2	1.40

$$\bar{X} = \frac{1.2}{10} + 23.0 = 23.12 \qquad (10.3)$$

$$\sigma^2 = \frac{1.4}{9} - \frac{10 \times 0.12^2}{9} = 0.1395 \qquad (10.4)$$

$$\sigma = 0.37 \ . \qquad (10.5)$$

The trial deviations $(X_i - \bar{\bar{X}})$ need contain no more decimal places than the readings X_i while the deviations $(X_i - \bar{X})$ must contain as many decimal places as $\bar{X}$. Thus $\Sigma (X_i - \bar{\bar{X}})^2$ is easier to calculate than $\Sigma (X_i - \bar{X})^2$.

11. A Test for the Normal Probability Distribution from Observed Values

Among the various tests available the easiest to apply directly to measurement data is perhaps the curve fitting type. It is easy to see how the data could be plotted as in Figure 2.1 and compared with the normal curve of Figure 4.1, with σ adjusted for best fit. It is important to understand that a finite number of measurements from a perfectly normal parent distribution will not give a perfect fit by the very nature of things. The fit will become better as the number of measurements becomes larger. Thus, approximate fits are a good indication of a normal parent distribution.

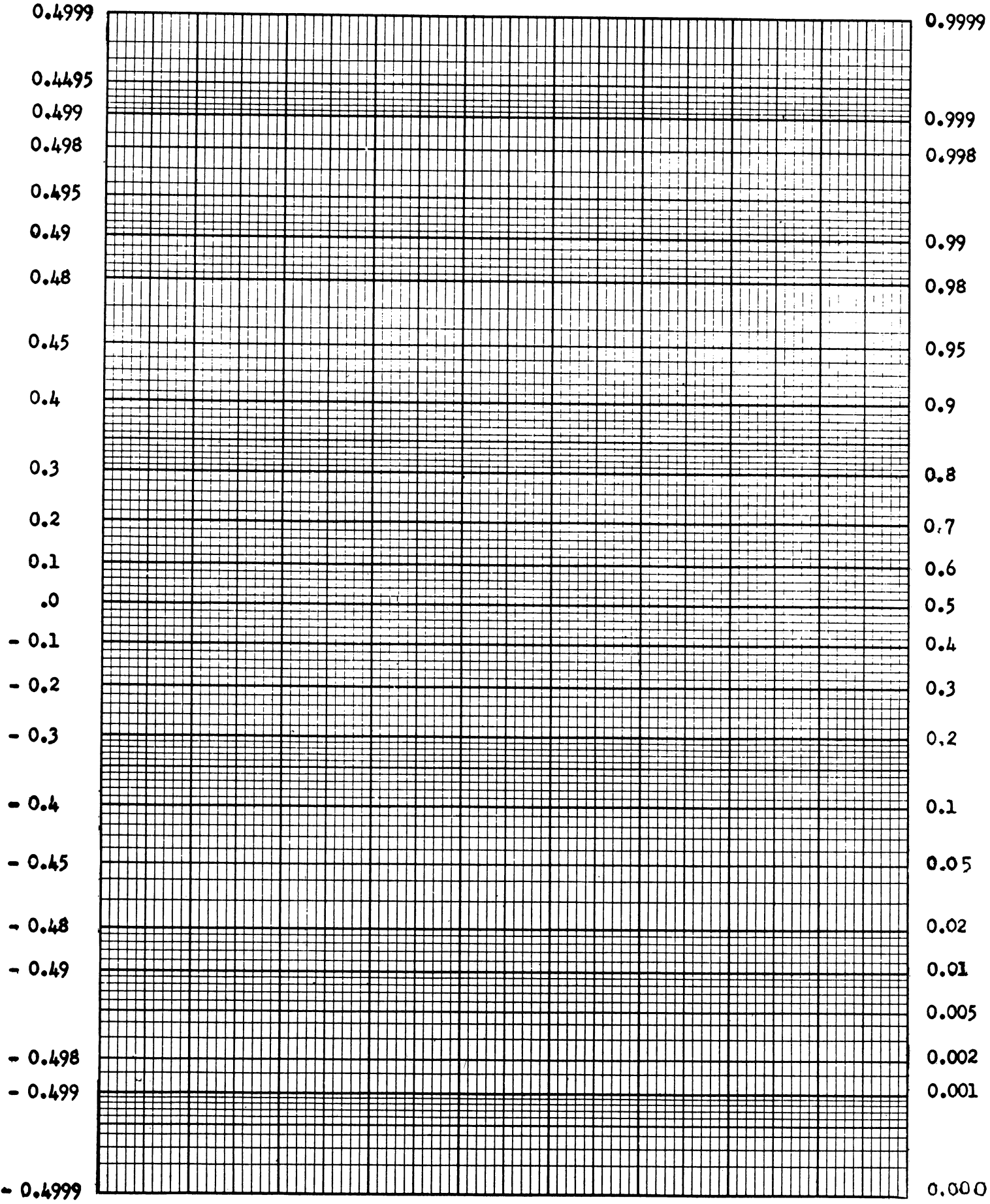
The curve fitting process with best choice of σ is tedious and time consuming and can be eliminated by using the sample of "Cumulative Probability Paper" on the next page. It is graph paper with a non-linear ordinate scale so designed that the normal CPF will always appear as a straight line. Then if steps are plotted representing the fraction of the measurements less than a series of equally spaced values (the abcissa scale represents the magnitude of the measurements) the step function approximation will lie approximately on a straight line for a normal parent distribution if a reasonable number of measurements are used and the steps are not too short. Various choices of the abcissa scale merely change the slope and intercept of the straight line passing through the center of the steps.

As an exercise the student should plot the curve of Figure 4.1 on the cumulative probability paper. Note that there is room for both branches of the curve. What is the significance of the abcissa values for which the left ordinate values are $\pm$ 0.25 and of the abcissa value for a left ordinant value of 0? In the case of discrete data explain why the straight line approximation should be fitted to the center of the step function approximation.

12. Estimation of the Mean by Measurements with Different Standard Deviations

Consider a set of n independent measurements $X_1,\ldots,X_n$ of the same quantity, but using different processes. Thus each X_i comes from a different distribution where the distributions have a common mean m, but where the standard deviation of the distribution of X_i is σ_{X_i}. Let us consider as an estimate for m, based on two measurements,

CUMULATIVE PROBABILITY PAPER



the weighted mean

$$m_e = w_1 X_1 + w_2 X_2 \tag{12.1}$$

where

$$w_1 + w_2 = 1. \tag{12.2}$$

Let us apply the minimum variance criterion by adjusting the weights to minimize $\sigma^2_{m_e}$. From (7.6)

$$\sigma^2_{m_e} = w_1^2 \sigma^2_{X_1} + w_2 \sigma^2_{X_2}$$

$$= w_1^2 \sigma^2_{X_1} + (1-w)^2 \sigma^2_{X_2}. \tag{12.3}$$

For a minimum

$$\frac{d}{dw_1} \sigma^2_{m_e} = 2w_1 \sigma^2_{X_1} + 2(w_1-1) \sigma^2_{X_2} = 0 . \tag{12.4}$$

The solution is

$$w_1 = \frac{\frac{1}{\sigma^2_{X_1}}}{\frac{1}{\sigma^2_{X_1}} + \frac{1}{\sigma^2_{X_2}}} \tag{12.5}$$

$$w_2 = \frac{\frac{1}{\sigma^2_{X_2}}}{\frac{1}{\sigma^2_{X_1}} + \frac{1}{\sigma^2_{X_2}}} . \tag{12.6}$$

Substitution of (12.5) and 12.6) into (12.3) gives

$$\sigma^2_{m_e} = \frac{1}{\frac{1}{\sigma^2_{X_1}} + \frac{1}{\sigma^2_{X_2}}} . \tag{12.7}$$

For the case of n measurements the results generalize to

$$w_i = \frac{\frac{1}{\sigma^2_{X_i}}}{\sum_{j=1}^{n} \frac{1}{\sigma^2_{X_j}}} \tag{12.8}$$

$$\sigma^2_{m_e} = \frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma^2_{X_j}}} \quad . \tag{12.9}$$

Note that σ_{m_e} is smaller than the smallest σ_{X_i}.

13. Propagation of Precision Indices

Often the desired result R of an experiment is only indirectly measurable as a function of several direct measurements $X_1, X_2, \ldots X_n$

$$R = R(X_1, X_2, \ldots X_n). \tag{13.1}$$

Each of the quantities $X_1, X_2, \ldots X_n$ is an estimate consisting of an observed value from a particular distribution whose mean is taken to be the "true value" of the quantity estimated. Let the means of the respective distributions be

$$m_{X_i} = E(X_1), \quad i = 1, 2, \ldots n \quad . \tag{13.2}$$

and the corresponding variances be

$$\sigma^2_{X_i} = E(X_1 - m_{X_i})^2, \quad i = 1, 2, \ldots n \; . \tag{13.3}$$

Define

$$dX_1 = X_1 - m_{X_i}, \quad i = 1, 2, \ldots n \; . \tag{13.4}$$

Then for sufficiently small dX_1, total differentiation gives

$$dR \approx \frac{\partial R}{\partial X_1} dX_1 + \frac{\partial R}{\partial X_2} dX_2 + \ldots + \frac{\partial R}{\partial X_n} dX_n \qquad (13.5)$$

or

$$R - m_R \approx \frac{\partial R}{\partial X_1}(X_1 - m_{X_1}) + \frac{\partial R}{\partial X_2}(X_2 - m_{X_2}) + \ldots$$

$$+ \frac{\partial R}{\partial X_n}(X_n - m_{X_n}) \qquad (13.6)$$

where we have used

$$m_R \approx R(m_{X_1}, m_{X_2}, \ldots, m_{X_n}). \qquad (13.7)$$

If $X_1, X_2, \ldots, X_n$ are independent then the coefficients of the partial derivatives are independent random variables with mean zero and by (7.7)

$$\sigma_R^2 = E(R - m_R)^2 \approx (\frac{\partial R}{\partial X_1})^2 \sigma_{X_1}^2 + (\frac{\partial R}{\partial X_2})^2 \sigma_{X_2}^2 + \ldots$$

$$+ (\frac{\partial R}{\partial X_n})^2 \sigma_{X_n}^2. \qquad (13.8)$$

It is important to note that the above result and (13.7) are approximate with errors which depend not only on the magnitude of the σ_{X_i} but also on the nature of the tails of the distributions of the X_i and on the function R. It is possible for (13.8) to fail entirely even if the σ_{X_i} are arbitrarily small. However if the X_i come from truncated distributions (a good assumption for many measurement problems), then the errors in (13.7) and (13.8) approach zero for sufficiently small σ_{X_i} Often it is not necessary to assume truncation of the distributions.

Example: Let us estimate the area of a circle by measuring the radius. Thus σ_r is known and σ_A is desired where

$$A = \pi r^2 . \tag{13.9}$$

From (13.8)

$$\sigma_A \approx (\frac{\partial A}{\partial r}) \sigma_r \tag{13.10}$$

or

$$\sigma_A^2 \approx (2\pi)^2 r^2 \sigma_r^2 . \tag{13.11}$$

The partial derivative in (13.8) must be evaluated at a value of r near m_r and if σ_r is small, the estimated value of r can be used rather than m_r.

The value of m_A, given by (13.7), is

$$m_A \approx \pi m_r^2 . \tag{13.12}$$

14. Least Squares Fitting of a Straight Line to a Set of Points in the Plane

When two physical quantities X and Y are known to be related linearly as

$$Y = aX + b \tag{14.1}$$

It is often necessary to obtain estimates for a and b from pairs of measurements of X and Y. It is clear that one pair of values X and Y, no matter how accurately known, is not enough to determine a and b. In geometrical language, one point does not determine a straight line.

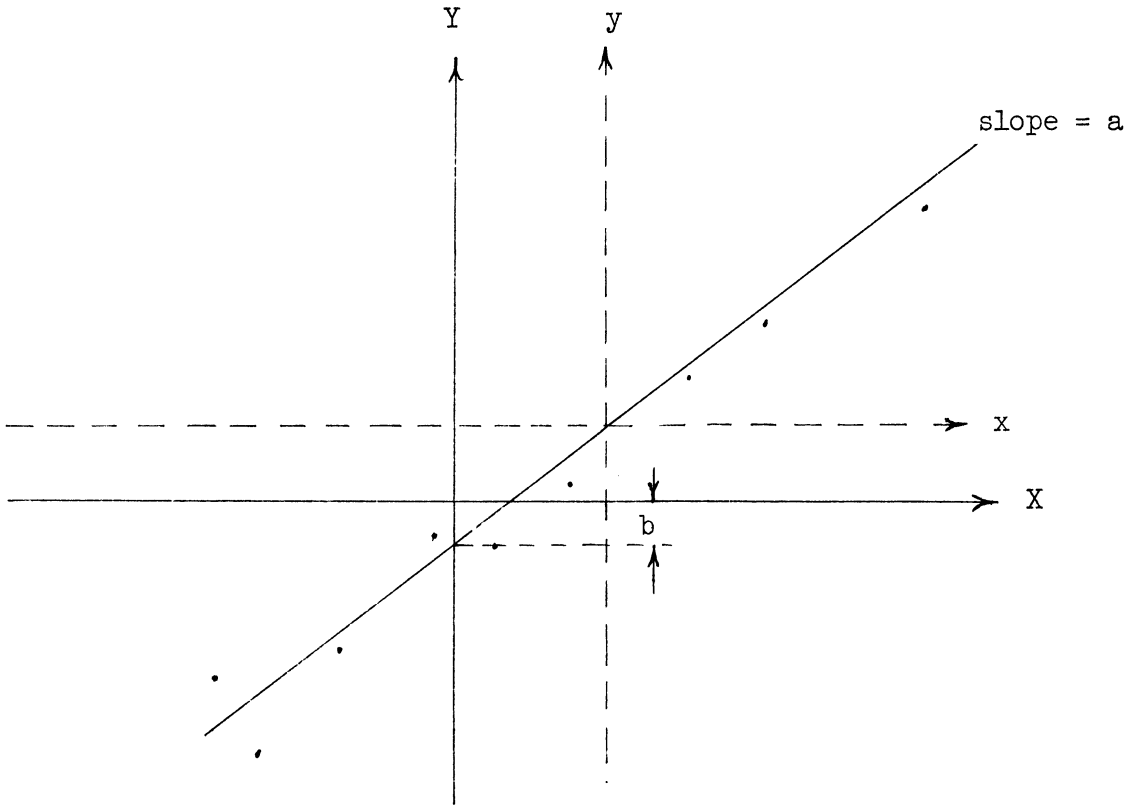


Figure 14.1

Each pair of measured values represents a point (X_i, Y_i) in Figure 14.1. By some means we wish to obtain a best fitted straight line. Assuming for the moment that the straight line represents the true relation between the quantities X and Y, the measured points may fail to lie on the line due to statistical errors in the measurement of X or Y or both. We shall simplify the situation by assuming that errors occur only in the measurements of Y.

To simplify subsequent calculations we define and use the differences from the mean

$$x_i = X_i - \frac{1}{n} \sum_{i=1}^{n} X_i \qquad (14.2)$$

$$y_i = Y_i - \frac{1}{n} \sum_{i=1}^{n} Y_i. \qquad (14.3)$$

Geometrically this amounts to choosing a new set of coordinate axes (x,y) with origin located at the centroid of the points as indicated in Figure 14.1. In terms of the new coordinate system let the equation of the straight line be

$$y = ax + d \tag{14.4}$$

The y-deviation from the straight line for a point (x_i, y_i) is then, by (14.4)

$$\Delta y_i = y_i - ax_i - d. \tag{14.5}$$

We have seen previously in Section 9 for the case of the normal distribution how the maximum likelihood estimate of the true value is obtained by minimizing the sum of the squares of the deviations. Following this procedure, we shall determine the values of a and d which minimize the sum of the squares of the y-deviations,

$$\sum_{i=1}^{n} (\Delta y_i)^2 = \sum_{i=1}^{n} (y_i - ax_i - d)^2. \tag{14.6}$$

To do this we set the partial derivatives of (14.6) with respect to a and d equal to zero and solve for a and d. The student should verify that this gives a minimum. Keeping in mind that from (14.2) and (14.3)

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = 0. \tag{14.7}$$

The solutions are easily obtained as

$$a = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \tag{14.8}$$

$$d = 0 . \tag{14.9}$$

These values minimize the radius of gyration of the points about the straight line. In terms of the original coordinates X_i, Y_i

$$a = \frac{\sum X_i Y_i - \frac{1}{n} \sum X_i \sum Y_i}{\sum X_i^2 - \frac{1}{n} (\sum X_i)^2} \tag{14.10}$$

$$b = \frac{1}{n} \sum Y_i - \frac{a}{n} \sum X_i = \frac{\frac{1}{n} \sum X_i^2 \sum Y_i - \frac{1}{n} \sum X_i \sum X_i Y_i}{\sum X_i^2 - \frac{1}{n} (\sum X_i)^2} \tag{14.11}$$

where the summations have the usual limits.

The known relation between the quantities X and Y may be such that b = 0 in (14.1). In this case (14.10) reduces to the simple relation

$$a = \frac{\sum X_i Y_i}{\sum X_i^2} \tag{14.12}$$

It is important not to confuse this result with (14.8).

15. Mean Square Value of the Least Squares Deviation

The mean square value of the deviation $\Delta Y_i = \Delta y_i$ of the points from the straight line is given by

$$MS(\Delta y) = \frac{1}{n} \sum (y_i - a x_i)^2 \tag{15.1}$$

Substituting for a from (14.8) gives

$$MS(\Delta y) = \frac{1}{n} \sum y_i^2 - \frac{(\sum x_i y_i)^2}{n \sum x_i^2} \tag{15.2}$$

16. Standard Deviation of the Least Squares Parameters

By (14.8) the parameter a is a function of the measurements Y_i and X_i. Let us assume that the Y_i have independent errors, each with mean zero and standard deviation σ_y. We shall call $\bar{y}_i$ the true value of y_i corresponding to x_i, and $\bar{a}$ the true value of a. Then we have

$$(a - \bar{a})^2 = \frac{[\Sigma\, x_i(y_i - \bar{y}_i)]^2}{[\Sigma\, x_i^2]^2} \tag{16.1}$$

The expectation of (16.1) is taken to obtain the variance of a. We note that $E(y_i - \bar{y}_i) = 0$, because the error in each y_i has mean zero. The independence of the errors in y_i gives us

$$E[(y_i - \bar{y}_i)\,(y_j - \bar{y}_j)] = 0 \tag{16.2}$$

whenever $i \neq j$. Therefore, the cross terms in the numerator of (16.1) drop out when we take the expectation. For the other terms,

$$\sigma_a = \frac{\sigma_y}{\sqrt{\Sigma\, x_i^2}} \tag{16.3}$$

Similar reasoning can be applied to obtain an expression for σ_b. The expression (14.11) is used for b, and the true value of b subtracted. The result is squared, and the expectation taken. Again most terms cancel, leaving

$$\sigma_b = \sqrt{\frac{\Sigma\, X_i^2}{n\Sigma X_i^2 - (\Sigma X_i)^2}}\,\sigma_y \tag{16.4}$$

Both (16.3) and (16.4) can also be calculated using the propagation of error approach of Section 13. The final result is the same in either case.

If b = 0 as in (14.12) then (16.3) reduces to the simple relation

$$\sigma_a = \frac{\sigma_y}{\sqrt{\sum X_i^2}} \tag{16.5}$$

It is important not to confuse this result with (16.3).

17. Linear Correlation

The pairs of values (X_i, Y_i) introduced in Section 14 can be thought of as corresponding values from the two sequences $X_1, X_2, \ldots X_n$ and $Y_1, Y_2, \ldots, Y_n$. The question occurs regarding the correlation between the two sequences. That is, with what probability can the Y_i values be determined from the corresponding X_i values. By linear correlation we mean that a linear relation like (14.1) is used to estimate the Y_i values from the corresponding X_i values.

Clearly, if the points of Figure 14.1 lie in a straight line the correlation is certainty and should be assigned a value of unity. In this case $\sigma_y = 0$ and by (15.2)

$$(\sum x_i\, y_i)^2 = \sum x_i^2 \sum y_i^2 \ . \tag{17.1}$$

Consider the expression

$$C_{xy} = \frac{\sum x_i\, y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} \tag{17.2}$$

It is called the normalized cross correlation between the zero-mean sequences x_i and y_i. It is a measure of how well the changes in the sequence Y_i can be estimated from the changes in the sequence X_i and, conversely. From (15.2) we obtain the inequality

$$-1 \leq C_{xy} \leq 1 \ . \tag{17.3}$$

When $C_{xy} = 1$ then $y_i = ax_i$ with $a > 0$ and when $C_{xy} = -1$ then $y_i = ax_i$ with $a < 0$. In either case the correlation is certainty. When $C_{xy} = 0$ the x_i and y_i sequences are completely independent of each other; that is, the prediction of values in one sequence by linear means from the corresponding values of the other sequence is completely uncertain.

A useful interpretation of C_{xy} is that it is the cosine of the angle between the two n-dimensional vectors with components x_i and y_i respectively. When the vectors are normal $C_{xy} = 0$, when they are parallel with the same direction $C_{xy} = 1$, and when they are parallel with opposite directions $C_{xy} = -1$.

If we choose

$$y_i = x_{i+j}, \quad j \geq 1$$

then $C_{xy} = C_x(j)$ is called the serial correlation or autocorrelation of the sequence x_i, $C_x(j)$ contains information regarding the interrelations between the terms of the sequence x_i, such as the periodic properties of the terms of the sequence x_i as a function of i.