THE METHOD OF LEAST SQUARES

D. E. WELLS
E. J. KRAKIWSKY

May 1971

LECTURE NOTES
18

THE METHOD OF LEAST SQUARES

D.E. Wells
E.J. Krakiwsky

Department of Geodesy and Geomatics Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B.
Canada
E3B 5A3

May 1971
Latest reprinting February 1997

PREFACE

In order to make our extensive series of lecture notes more readily available, we have scanned the old master copies and produced electronic versions in Portable Document Format. The quality of the images varies depending on the quality of the originals. The images have not been converted to searchable text.

PREFACE

These notes have been compiled in an attempt to integrate a description of the method of least squares as used in surveying with

(a) statistical concepts

(b) linear algebra using matrix notation, and

(c) the use of digital computers.

They are a culmination of concepts first described at a symposium held in 1964 at the University of New Brunswick on "The Impact of Electronic Computers on Geodetic Adjustments" (see The Canadian Surveyor, Vol. IX, No. 1, March 1965). We also owe a considerable debt to Professor Urho Uotila of The Ohio State University, Department of Geodetic Science whose lecture notes ("An Introduction to Adjustment Computations", 1967) provided a starting point for these notes. We have attempted to retain Uotila's notation with minimum changes.

We acknowledge the invaluable help of Mr. Mohammed Nassar who meticulously proofread the manuscript before the first printing, and Dr. Georges Blaha who suggested a number of corrections which have been incorporated in the second printing.

INDEX

LIST OF FIGURES

LIST OF TABLES

THE METHOD OF LEAST SQUARES

1. INTRODUCTION

The method of least squares is the standard method used to obtain unique values for physical parameters from redundant measurements of those parameters, or parameters related to them by a known mathematical model.

The first use of the method of least squares is generally attributed to Karl Friedrich Gauss in 1795 (at the age of 18), although it was concurrently and independently used by Adrien Marie Legendre. Gauss invented the method of least squares to enable him to estimate the orbital motion of planets from telescopic measurements.

Developments from three other fields are presently finding increasing application in the method of least squares, and are profoundly influencing both the theory and practice of least squares estimation. These three developments are the concepts of modern statistical estimation theory; matrix notation and the concepts of modern linear algebra; and the use of large fast digital computers.

These notes are an attempt to describe the method of least squares making full use of these three developments. A review of the concepts of statistics is given in chapters 2, 3, 4, 5 and 8 of these notes. The required background in matrix notation and linear algebra is

given in the course notes on "Matrices" [Wells 1971]. A description of digital computer programming is beyond the scope of this presentation, however, an Appendix to these notes contains listings of results obtained by computer for a specific problem discussed throughout these notes.

The remainder of this chapter briefly outlines the relationship of the method of least squares to statistics and linear algebra, and describes the current impact of digital computers on practical computing techniques.

1.1 STATISTICS AND THE METHOD OF LEAST SQUARES

Physical quantities can never be measured perfectly. There will always be a limiting measurement precision beyond which either the mathematical model describing the physical quantity, or the resolution of the measuring instrument, or both will fail. Beyond this limiting precision, redundant measurements will not agree with one another (that is they will not be consistent).

For example if we measure the length of a table several times with a meter stick and eyeball, the limiting precision is likely to be about one millimeter. If we record our measurements only to the nearest centimeter they will be consistent. If we record our measurements to the nearest tenth of a millimeter, they will be inconsistent.

The precision which we desire is often beyond the limiting precision of our measurements. In such a case we can not <u>know</u> the "true"value of our physical quantity. All we can do is make an <u>estimate</u> of the "true" value. We want this estimate to be unique

(that is determined by some standard method which will always yield the same estimate given the same measurements), and we want to have some idea of how "good" the estimate is.

The scientific method for handling inconsistent data is called statistics. The methods for determining unique estimates (together with how good they are) is called statistical estimation. The method of least squares is one such method, based on minimizing the sum of the squares of the inconsistencies.

It should be emphasized that there are other methods which will yield unique estimates, for example minimizing the sum of the absolute values of the inconsistencies, or minimizing the maximum inconsistency [Hamming 1962]. However, these other methods have at least three disadvantages compared to the method of least squares. The method of least squares can be applied to problems involving either linear or non-linear mathematical models; these other two methods are restricted to linear problems only, because of fundamental continuity and differentiability limitations. Least squares estimates are related to a statistical quantity (the arithmetic mean) which is usually more important than the statistical quantities (the median and mid-range respectively) to which these other methods are related. And finally the method of least squares is in common use in many fields, making it the standard method of obtaining unique estimates.

Statistics is sometimes called the theory of functions of a random variable. An ordinary variable is defined only by the set of permissible values which it may assume. A random variable is defined both by this set of permissible values, and by an associated frequency

(or probability density) function which expresses how often each of these values appear in the situation under discussion. The most important of these functions is the normal (or Gaussian) frequency function. Physical measurements can almost always be assumed to be random variables with a normal frequency function.

A unique statistical estimate of the value of a physical quantity (called a point estimate) together with some indication of how close it is to the "true" value can be made whether the frequency function is known or not. However, there are other kinds of estimates (called interval estimates and hypothesis tests), which cannot be made unless a particular frequency function is specified.

Chapter 2 summarizes statistical nomenclature and concepts. Chapters 3 and 4 present the properties of some particular distributions, related to the normal distribution. Chapters 6 and 7 discuss least squares point estimators, and Chapters 5 and 8 discuss interval estimation and hypothesis testing.

1.2 LINEAR ALGEBRA AND THE METHOD OF LEAST SQUARES

The system of linear equations

$$A X = L \qquad 1\text{-}1$$

where X is called the unknown vector, L is the constant vector, A the coefficient matrix, $[A \mid L]$ the augmented matrix, has a unique nontrivial solution only if

$$L \neq 0 \text{ (the system is nonhomogeneous)}, \qquad 1\text{-}2a$$

$$\text{rank of } A = \text{dimension of } X\ , \qquad 1\text{-}2b$$

$$\text{rank of } [A \mid L] = \text{rank of } A \text{ (system is consistent)}. \qquad 1\text{-}2c$$

In the case where there are no redundant equations, criterion (1-2b) will mean that A is square and nonsingular, and therefore has an inverse. The solution is then given by

$$X = A^{-1} L . \qquad 1\text{-}3$$

In the case where there are redundant equations, A will not be square, but $A^T A$ will be square and nonsingular, and the solution is given by

$$X = (A^T A)^{-1} A^T L \quad . \qquad 1\text{-}4$$

(See course notes on "Matrices" for a more detailed treatment of the above ideas).

Let us consider the case where the elements of L are the results of physical measurements, which are to be related to the elements of X by equation (1-1). If there are no redundant measurements (the number of measurements equals the number of unknowns) there will be a unique nontrivial solution for X. However, if there are redundant measurements, they will be inconsistent because physical measurements are never perfect. In that case criterion (1-2c) will not be satisfied, the system of equations will be inconsistent, and no unique solution will exist. All we are able to do is make a unique estimate of the solution. In order that a unique estimate exists, we must find some criterion to use in place of criterion (1-2c). There are a number of possible criteria, but the one commonly used is the least squares criterion; that the sum of the squares of the inconsistencies be a minimum. Before stating this criterion, let us find an expression for the inconsistencies.

Because equation (1-1) is inconsistent, let us write an equation which is consistent by adding a vector which will "cancel" the inconsistencies.

$$A X - L = V \qquad 1\text{-}5$$

where V is usually called the residual vector. The elements of V are not known and must be solved for, since we have no way of knowing what the inconsistent parts of each measurement will be. We can now replace criterion (1-2c), the consistency criterion, with the least squares criterion, which states that the "best" estimate $\hat{X}$ for X is the estimate which will minimize the sum of the squares of the residuals, that is

$$\hat{V}^T \hat{V} = (A \hat{X} - L)^T (A \hat{X} - L) = \text{minimum.} \qquad 1\text{-}6$$

The estimate $\hat{X}$ so determined is called the least squares estimate, and we will see (in Chapter 6 of these notes) that it is equal to the expression in equation 1-4, that is

$$\hat{X} = (A^T A)^{-1} A^T L \ , \qquad 1\text{-}7$$

and that the "best" estimate of the observation errors or residuals is given by

$$\hat{V} = A \hat{X} - L \ . \qquad 1\text{-}8$$

These estimates are the simple least squares estimates (also called the equally weighted least squares estimates).

Often the physical measurements which make up the elements of L do not all have the same precision (some could have been made using different instruments or under different conditions). This fact should be reflected in our least squares estimation process, so we assign to each measurement a known "weight" and call P the matrix whose elements are these weights, the weight matrix. We modify the least squares criterion to state that the best estimate is that which minimizes the sum of the squares of the weighted residuals, that is

$$\hat{V}^T P \hat{V} = \text{minimum.} \qquad 1\text{-}9$$

And as we will see in Chapter 6 the estimate is given by

$$\hat{X} = (A^T PA)^{-1} A^T PL \qquad 1\text{-}10$$

and is called the weighted least squares estimate.

In Chapter 6, we will see that if the weight matrix P is chosen to be the inverse of the estimated covariance matrix of the observations, then the least squares estimate is the minimum variance estimate, and that if the observation errors have a normal (Gaussian) distribution, then the least squares minimum variance estimate is the maximum likelihood estimate.

In this short introduction to least squares estimates we have considered only the linear mathematical model of equation 1-5. In Chapter 7 we will consider the more general case in which

i) the observations are related to nonlinear functions of the unknown parameters, that is

$$F(X) - L = V \qquad 1\text{-}11$$

and ii) the observations are nonlinearly related to functions of the unknown parameters, that is

$$F(X, L + V) = 0 \qquad 1\text{-}12$$

In Chapter 9 we will consider further complications of the mathematical models.

1.3 DIGITAL COMPUTERS AND THE METHOD OF LEAST SQUARES

So far we have described the method of least squares from a purely theoretical point of view, discussing expressions for least squares estimates. However, from a practical point of view, the

inversion of large matrices (and even the multiplication of matrices) requires a considerable number of computation steps.

Until the advent of large fast digital computers the solution of large systems of equations was a formidable and incredibly tedious task, attempted only when absolutely necessary. One application for which it was absolutely necessary was to obtain least squares estimates for survey net coordinates. Consequently the last 150 years have seen considerable ingenuity and effort directed by geodesists towards finding shortcuts, simplifications, and procedures to reduce the number of computation steps required.

Now that digital computers are in widespread use, this work does not hold the importance it once did. However, even the largest fastest digital computer is incapable of simultaneously solving systems which may incorporate, for example, several thousand equations. Therefore, ingenuity and effort are currently directed towards developing algorithms for chopping a large system into smaller pieces and solving it piece by piece, but in a manner such that the final solution is identical to that which would have been obtained by solving simultaneously. We will discuss some of these algorithms in Chapter 10.

1.4 GAUSS AND THE METHOD OF LEAST SQUARES

To dispel the notion that the concepts discussed in this Chapter are all new, we will analyze the following quotation from Gauss' book "Theoria Motus Corporum Coelestium" published in 1809 (see Gauss [1963] for an English translation).

> "If the astronomical observations and other quantities on which the computation of orbits is based were absolutely correct, the elements also, whether deduced from three or four observations, would be strictly accurate (so far indeed as the motion is supposed to take place exactly according to the laws of Kepler) and, therefore, if other observations were used, they might be confirmed but not corrected. But since all our measurements and observations are nothing more than approximations to the truth, the same must be true of all calculations resting upon them, and the highest aim of all computations made concerning concrete phenomena must be to approximate, as nearly as practicable, to the truth. But this can be accomplished in no other way than by a suitable combination of more observations than the number absolutely requisite for the determination of the unknown quantities. This problem can only be properly undertaken when an approximate knowledge of the orbit has been already attained, which is afterwards to be corrected so as to satisfy all the observations in the most accurate manner possible."

Note that this single paragraph, written over 150 years ago, embodies the concepts that

a) mathematical models may be incomplete,

b) physical measurements are inconsistent,

c) all that can be expected from computations based on inconsistent measurements are <u>estimates</u> of the "truth",

d) redundant measurements will reduce the effect of measurement inconsistencies,

e) an initial approximation to the final estimate should be used, and finally,

f) this initial approximation should be corrected in such a way as to minimize the inconsistencies between measurements (by which Gauss meant his method of least squares).

2. STATISTICAL DEFINITIONS AND CONCEPTS

Statistical terms are in everyday use, and as such are often used imprecisely or erroneously. Most of this Chapter is based on a comparative reading of Kendall [1957], Spiegel [1961], Hamilton [1964], Kendall [1969] and Carnahan et al [1969].

2.1 STATISTICAL TERMS

Statistics is the scientific method of collecting, arranging, summarizing, presenting, analyzing, drawing valid conclusions from, and making reasonable decisions on the basis of data. Statistical data include numerical facts and measurements or observations of natural phenomena or experiments. A statistic is a quantitative item of information deduced from the application of statistical methods.

A variable is a quantity which varies, and may assume any one of the values of a specified set. A continuous variable is a variable which can assume any value within some continuous range. A discrete variable (also called a discontinuous variable) is a variable which can assume only certain discrete values. A constant is a discrete variable which can assume only one value. In general, the result of a

measurement is a continuous variable, while the result of counting is a discrete variable.

A variate (also called a random variable) is a quantity which may assume any one of the values of a specified set, but only with a specified relative frequency or probability. A variate is defined not merely by a set of permissible values (as is an ordinary variable), but also by an associated frequency (probability) function expressing how often those values appear.

A population is the collection of all objects having in common a particular measurable variate. A population can be finite or infinite. For example, the population consisting of all possible outcomes of a single toss of a coin is finite (consisting of two members, heads and tails), while the population consisting of all possible outcomes in successive tosses of a coin is infinite, and the population consisting of all real numbers between 0 and 1 is also infinite. An individual is a single member of a population.

A sample is a group of individuals drawn from a population, and a random sample is a sample which is selected such that each individual in the population is equally likely to be selected. Usually sample implies random sample. Often the terms sample space, sample point, and event respectively, are used instead of population, individual and random sample (for example in Hamilton [1964]).

The individuals in a sample may be grouped according to convenient divisions of the variate-range. A group so determined is called a class. The variate-values determining the upper and lower limits of a class are called the class boundaries. The interval between class

boundaries is called the class interval. The number of individuals falling into a specified class is called the class frequency. The relative frequency (also called the proportional frequency) is the class frequency expressed as a proportion of the total number of individuals in the sample.

No single definition of the concept of probability is accepted by all statisticians. The classical definition is that the probability Pr (A) that an individual selected with equal likelihood from a population will fall into a particular class A is equal to the fraction of all individuals in the population which would, if selected, fall into A. This is a circular definition since the words "equal likelihood" really mean "equal probability", therefore defining probability in terms of itself. This problem can be resolved in two different ways, neither entirely satisfactory. The first is to define the empirical probability Pr (A) that an individual selected from a population will fall into a particular class A as the limit of the relative frequency of A for a series of n selections, as n tends to infinity. The second is to accept "probability" as an undefinable concept, and proceed to state the rules governing probabilities as axioms.

2.2 FREQUENCY FUNCTIONS

This discussion will be restricted to continuous variates only. Most of the results can be applied to discrete variates simply by replacing integrals by summations.

The frequency function ϕ (x) (also called the probability density function or p.d.f.) of the variate x is the relative frequency of x (assuming a class interval dx) as a function of the value of x, that is

$$\phi\ (x_o)\ dx = P_r(x_o \leq x \leq x_o + dx)\ , \qquad 2\text{-}1$$

where the term on the right is read "the probability that the value of the variate x lies between x_o and x_o + dx inclusive".

The cumulative frequency function Φ (x) (also called the distribution function, the cumulative distribution function or c.d.f., and the cumulative probability function), of the variate x is the integral of the frequency function ϕ (x)

$$\Phi\ (x_o) = \int_{-\infty}^{x_o} \phi\ (x)\ dx = P_r\ (x \leq x_o)\ , \qquad 2\text{-}2$$

where the term on the right is read "the probability that the value of the variate x is less than or equal to x_o".

The dependency of the frequency function ϕ (x) on x is called the frequency distribution. A typical frequency distribution is shown in Figure 2-1.

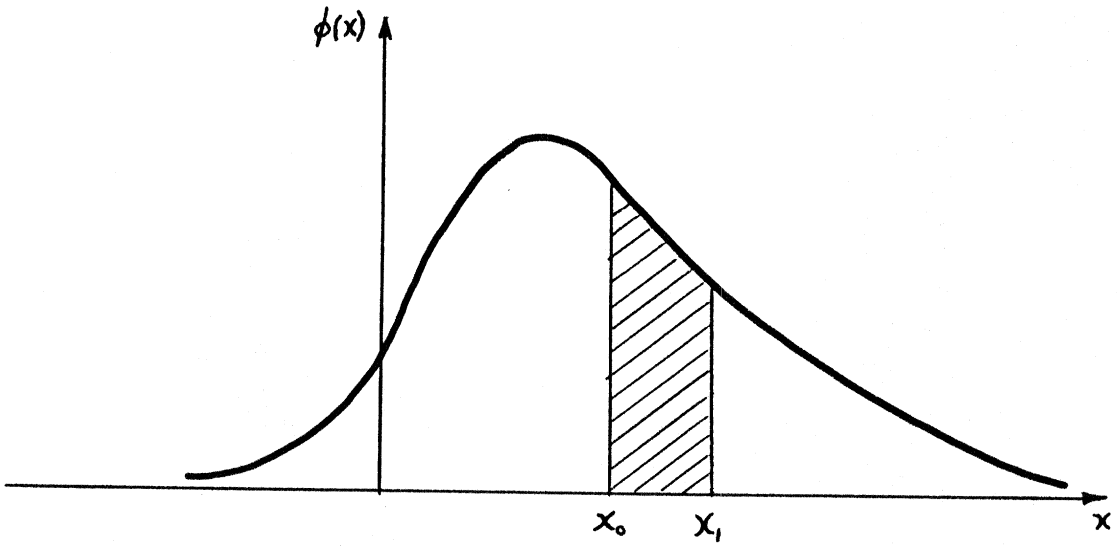
Probability is represented by an area under this curve. For example, the probability that x lies between x_o and x_1 is shown as the shaded area

$$P_r\ (x_o \leq x \leq x_1) = \int_{x_o}^{x_1} \phi\ (x)\ dx\ . \qquad 2\text{-}3$$

Note that the probability that x lies somewhere between the extreme limits is certainty (or probability of unity).

$$P_r\ (-\infty \leq x \leq +\infty) = \int_{-\infty}^{\infty} \phi(x)\ dx = 1\ . \qquad 2\text{-}4$$

Figure 2-1.



TYPICAL FREQUENCY DISTRIBUTION

The probability that x is less than or equal to x_o is the value of $\Phi(x_o)$ and is represented by the total area under the curve from $-\infty$ to x_o shown as the shaded area in Figure 2-2. Note $\Phi(-\infty)=0$; $\Phi(+\infty)=1$.

Frequency distributions have two important characteristics called central tendency and dispersion, and two less important characteristics called skewness (or departure from symmetry) and kurtosis (or peakedness). Measures of central tendency include the arithmetic mean (or simply the mean), the median (the value dividing the distribution in two equal halves), the mode (the most frequently occurring value), the geometric mean and the harmonic mean, of which the mean is most often used. Measures of dispersion include the standard deviation, the mean deviation and the range, of which the standard deviation is most often used.

The expected value of a function f(x) is an arithmetic average of f(x) weighted according to the frequency distribution of the variate x and is defined

$$E\,[f(x)] = \int_{-\infty}^{\infty} f(x)\,\phi(x)\,dx \quad . \qquad 2\text{-}5$$

Expected values have the following properties

$$E\,[k\,f(x)] = k\,E[f(x)] \quad , \qquad 2\text{-}6a$$

$$E\,[f_1(x) + f_2(x)] = E\,[f_1(x)] + E[f_2(x)] \quad , \qquad 2\text{-}6b$$

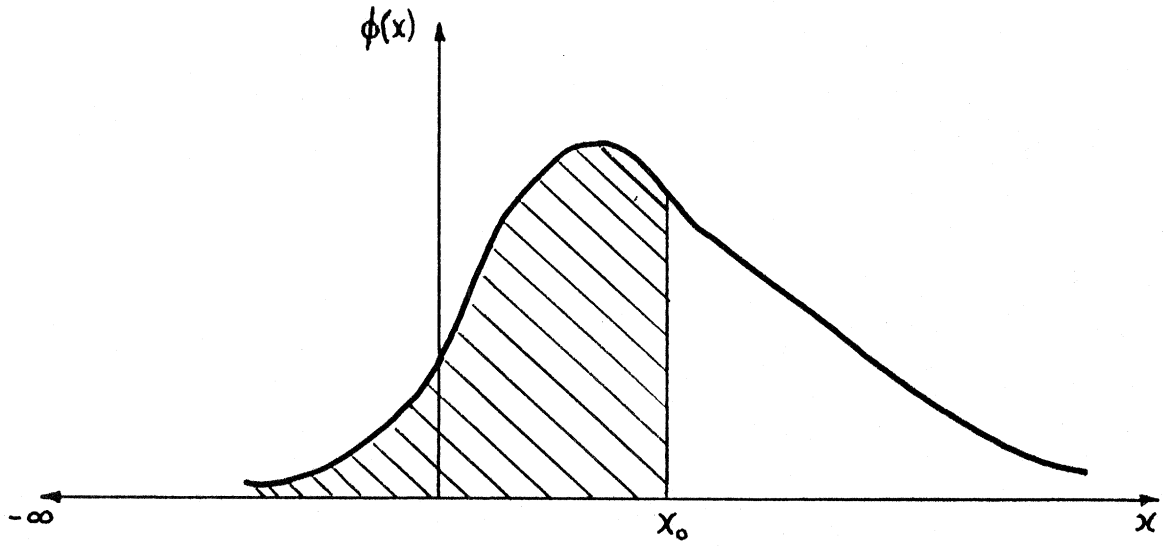
$$E\,[E[f(x)]] = E\,[f(x)] \quad , \qquad 2\text{-}6c$$

$$E\,[\Sigma f(x)] = \Sigma E\,[f(x)] \quad . \qquad 2\text{-}6d$$

The mean μ of a distribution is the expected value of the variate x itself

$$\mu = E\,[x] = \int_{-\infty}^{\infty} x\,\phi\,(x)\,d\,x \quad . \qquad 2\text{-}7$$

Figure 2-2.



CUMULATIVE DISTRIBUTION FUNCTION

The nth moment of a distribution about its origin is defined

$$E\,[x^n] = \int_{-\infty}^{\infty} x^n\, \phi\,(x)\, d\,x \quad . \qquad 2\text{-}8$$

The nth moment of a distribution about its mean is defined

$$E\,[(x-\mu)^n] = \int_{-\infty}^{\infty} (x-\mu)^n\, \phi(x)\, d\,x \quad . \qquad 2\text{-}9$$

The second moment about the mean is called the variance.

$$\sigma^2 = \text{var}\,(x) = E[(x-\mu)^2] = E[x^2] - \mu^2 \quad , \qquad 2\text{-}10$$

and the standard deviation σ is defined as the positive square root of the variance.

The moment generating function or m.g.f. of a variate x is defined as

$$M(t) = E[e^{tx}] = \int_{-\infty}^{\infty} e^{tx}\, \phi(x)\, d\,x \quad , \qquad 2\text{-}10a$$

Moments of a distribution can be deduced directly from the moment generating function. For example, the mean of a distribution is

$$\mu = E[x] = \left.\frac{dM\,(t)}{dt}\right|_{t=0} = M'\,(0) \quad , \qquad 2\text{-}10b$$

and the variance is

$$\sigma^2 = E\,[x^2] - \mu^2 = M''\,(0) - [M'\,(0)]^2 \quad . \qquad 2\text{-}10c$$

A distribution can be completely defined by specifying any one of;the frequency function $\phi(x)$ (or p.d.f.); the cumulative distribution function $\Phi(x)$; or the moment generating function $M(t)$.

2.3 MULTIVARIATE FREQUENCY FUNCTIONS

Thus far we have considered only univariate distributions (dis-

tributions of a single variate x). We will now extend the above results to multivariate distributions (distributions having several variates $x_1, x_2, \ldots x_n$ associated with each individual in the population). Let

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{bmatrix}$$

be the vector of variates. Then the multivariate frequency function (also called the joint density function) is defined

$$\phi\ (X^o)\ dx_1\ dx_2 \ldots dx_n = P_r(X^o \leq X \leq X^o + dX)\ , \qquad 2\text{-}11$$

where

$$X^o = \begin{bmatrix} x_1^o \\ x_2^o \\ \cdot \\ x_n^o \end{bmatrix}\ , \quad dX = \begin{bmatrix} dx_1 \\ dx_2 \\ \cdot \\ dx_n \end{bmatrix}$$

and $P_r(X^o \leq X \leq X^o + dX)$ is the probability that

$$x_1^o \leq x_1 \leq x_1^o + dx_1$$

$$\cdot$$

$$\cdot$$

$$x_n^o \leq x_n \leq x_n^o + dx_n$$

all hold simultaneously. The multivariate cumulative frequency function (also called the joint cumulative distribution function) is defined

$$\Phi(X^o) = \int_{-\infty}^{x_1^o} \cdots \int_{-\infty}^{x_n^o} \phi(X)\ dx_1\ dx_2 \ldots dx_n\ . \qquad 2\text{-}12$$

$$= P_r\,(X \leq X^o)$$

If the variate vector X can be partitioned into two vectors

$$X_1 = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_m \end{bmatrix} \quad \text{and} \quad X_2 = \begin{bmatrix} x_{m+1} \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

such that

$$\phi(X) = \phi_1(X_1)\ \phi_2(X_2) \quad ,$$

then the two sets of variates X_1 and X_2 are called statistically independent.

The expected value of a multivariate function f(X) is defined

$$E[f(X)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(X)\ \phi(X)\ dx_1\ dx_2 \cdots dx_n \quad . \qquad 2\text{-}13$$

The mean vector of a multivariate distribution is defined

$$U_X = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_n] \end{bmatrix} = E \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = E[X] \quad . \qquad 2\text{-}14$$

The second moments of the elements of X about their means forms a symmetric matrix called the covariance matrix (also called the variance-covariance matrix).

$$\Sigma_X = E[(X-U_X)\ (X-U_X)^T] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ & & \vdots & \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} , \qquad 2\text{-}15$$

where σ_i^2 is called the variance of x_i

$$\sigma_i^2 = \text{var}\ (x_i) = E[(x_i-\mu_i)^2] = E[x_i^2]-\mu_i^2 \quad , \qquad 2\text{-}16$$

and σ_{ij} is called the covariance between x_i and x_j

$$\sigma_{ij} = \text{cov}\ (x_i\ x_j) = E[(x_i-\mu_i)\ (x_j-\mu_j)] = E[x_i\ x_j] - E[x_i]\ E[x_j] \quad , \quad 2\text{-}17$$

and $\sigma_{ij} = \sigma_{ji}$ since Σ_X is symmetric.

The correlation coefficient between x_i and x_j is

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i\ \sigma_j} \quad , \qquad 2\text{-}17a$$

and has values

$$-1 \leq \rho_{ij} \leq +1 \quad . \tag{2-17b}$$

If x_i and x_j are statistically independent

$$\phi(x_i, x_j) = \phi_1(x_i)\ \phi_2(x_j) \quad ,$$

and

$$E[x_i\ x_j] = E[x_i]\ E[x_j] \quad ,$$

so that

$$\sigma_{ij} = \rho_{ij} = 0 \quad .$$

In fact the covariance σ_{ij} and the correlation coefficient ρ_{ij} are measures of the statistical dependence or correlation between x_i and x_j.

2.4 THE COVARIANCE LAW

Assume we have a second variate vector Y linearly related to X by

$$Y = C\ X \quad . \tag{2-18}$$

Then

$$U_Y = E[Y] = E[CX] = C\ E[X] = C\ U_X \quad , \tag{2-19}$$

and

$$\Sigma_Y = E[(Y-U_Y)\ (Y-U_Y)^T] = E[(C\ X - C\ U_X)(CX-CU_X)^T]$$

$$= E[C(X-U_X)\ (X-U_X)^T\ C^T] = C\ E[(X-U_X)(X-U_X)^T]C^T \quad ,$$

or

$$\boxed{\Sigma_Y = C\ \Sigma_X\ C^T \quad .} \tag{2-20}$$

This is known as the <u>covariance law</u> (also called the <u>law of covariances</u> and <u>law of propagation of covariances</u>).

If Y is nonlinearly related to X

$$Y = F(X) \quad , \tag{2-21}$$

then we choose some value X° and replace $F(X)$ by its Taylor's series linear approximation about X°, that is

$$Y = F(X) \simeq F(X^\circ) + \frac{\partial F}{\partial X}\Big|_{X^\circ} (X - X^\circ) \quad .$$

Then

$$U_Y = E[Y] = E[F(X^\circ) + \frac{\partial F}{\partial X}\Big|_{X^\circ} (X - X^\circ)]$$

$$= F(X^\circ) + \frac{\partial F}{\partial X}\Big|_{X^\circ} (E[X] - X^\circ) = F(X^\circ) + \frac{\partial F}{\partial X}\Big|_{X^\circ} (U_X - X^\circ), \tag{2-22}$$

and

$$Y - U_Y = C(X - U_X) \quad ,$$

where

$$C = \frac{\partial F}{\partial X}\Big|_{X^\circ} \quad .$$

Thus

$$\Sigma_Y = E[(Y-U_Y)\ (Y-U_Y)^T] = C\Sigma_X C^T \quad , \tag{2-23}$$

which is identical to the covariance law (equation 2-20), with

$$C = \frac{\partial F}{\partial X}\Big|_{X^\circ} \quad .$$

2.5 STATISTICAL POINT ESTIMATION

A characteristic of a given distribution (for example its mean or variance) is a <u>statistic</u> of that distribution. A distinction is drawn between <u>population statistics</u> (also called <u>population parameters</u>, or simply <u>parameters</u>), which are usually denoted by Greek letters,

and sample statistics (also called simply statistics), which are usually denoted by Latin letters. For example, the population standard deviation is denoted by σ, and the sample standard deviation by s.

Statistical estimation is that branch of the statistical method which is concerned with the problem of inferring the nature of a population from a knowledge of samples drawn from the population. A sample statistic e whose value is used to infer the value of a population statistic ε is called an estimator (or point estimator) of ε, and is denoted $\hat{\varepsilon}$. The value of e is called an estimate of the value of ε. Sample statistics which might be used as estimators include sample mean, sample variance, sample standard deviation, sample median, and sample range. The most often used estimators are the sample mean and the sample variance, which for a sample consisting of n observations of a single variate x are usually defined

$$\bar{x} = \frac{1}{n} \sum_i x_i \qquad 2\text{-}24$$

$$s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 \qquad 2\text{-}25$$

If we were to draw another sample from the same population, it would be surprising if the sample mean and variance of this new sample were identical to the mean and variance of the original sample. We see that the value of a sample statistic will in general vary from sample to sample, that is the sample statistic itself is a variate and will have a distribution, called its sampling distribution. We now have three distributions under consideration: the distribution of individuals in the population; the distribution of individuals in a single sample; and the distribution of the value of a sample statistic over all possible samples.

Consider the sampling distribution of the sample mean statistic, called the sampling distribution of means. This distribution itself has a mean and variance. The mean (or expected value) is

$$E[\bar{x}] = E[\frac{1}{n}\sum_i x_i] = \frac{1}{n}\sum_i E[x_i] = \frac{1}{n}\sum_i \mu = \mu \quad , \qquad 2\text{-}26$$

that is the expected value of the sample mean is the population mean. The variance is

$$\begin{aligned} \text{var}(\bar{x}) &= E[(\bar{x}-\mu)^2] = E[(\bar{x}^2 - 2\bar{x}\mu + \mu^2)] = E[\bar{x}^2] - \mu^2 \\ &= E[(\frac{1}{n}\sum_i x_i)^2] - \mu^2 = \frac{1}{n^2} E[(\sum_i x_i)(\sum_j x_j)] - \mu^2 \\ &= \frac{1}{n^2}(\sum_i E[x_i^2] + \sum_{i\neq j} E[x_i\, x_j]) - \mu^2 \quad . \end{aligned}$$

But

$$E[x_i^2] = \sigma^2 + \mu^2 \text{ , and } E[x_i\, x_j] = E[x_i]\, E[x_j] = \mu^2,$$

and

$$\sum_{i\neq j} E[x_i\, x_j] = n(n-1)\mu^2 \quad .$$

Therefore

$$\text{var}(\bar{x}) = \frac{1}{n^2}(n(\sigma^2+\mu^2) + n(n-1)\mu^2) - \mu^2 = \frac{\sigma^2}{n} \quad . \qquad 2\text{-}27$$

Consider now the sampling distribution of the sample variance statistic, called the sampling distribution of variances. The mean (or expected value) of this distribution is

$$\begin{aligned} E[s^2] &= E[\frac{1}{n-1}\sum_i (x_i - \bar{x})^2] \\ &= \frac{1}{n-1} E[\sum_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2)] = \frac{1}{n-1}\left(\sum_i E[x_i^2] - nE[\bar{x}^2]\right) \quad . \end{aligned}$$

But

$$E[x_i^2] = \sigma^2+\mu^2 \text{ , and } E[\bar{x}^2] = \frac{\sigma^2}{n} + \mu^2 \quad .$$

Therefore

$$E[s^2] = \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) = \sigma^2 \qquad 2\text{-}28$$

that is the expected value of the sample variance is the population variance.

An unbiased estimator is an estimator whose expected value (that is the mean of the sampling distribution of the estimator) is equal to the population statistic it is estimating. We have seen that the sample mean as defined above is an unbiased estimator of the population mean and that the sample variance as defined above is an unbiased estimator of the population variance, that is

$$E[\hat{\mu}] = \mu$$

where

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_i x_i \quad , \qquad 2\text{-}29$$

$$E[\hat{\sigma}^2] = \sigma^2$$

where

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_i (x_i-\bar{x})^2 \quad . \qquad 2\text{-}30$$

A particular population statistic may have several possible estimators. There are a number of criteria for deciding which of these is the "best" estimator. The minimum variance estimator is the estimator whose variance (that is the variance of the sampling distribution of the estimator) is less than that of the other possible estimators. Another criterion is the maximum likelihood estimator, the definition of which we will leave until Chapter 6.

2.6 STATISTICAL INTERVAL ESTIMATION AND HYPOTHESIS TESTING

So far we have discussed only point estimation, that is the inference of the value of a population statistic ε from the value of a sample statistic e. Statistical estimation includes two other procedures, called interval estimation and hypothesis testing.

In point estimation, we specify an estimate $\hat{\varepsilon}$ for the population statistic ε. In interval estimation we specify a range of values bounded by an upper and lower limit.

$$e_1 \leq \varepsilon \leq e_2$$

within which the population statistic is estimated to lie. If the probability

$$P_r\ (e_1 \leq \varepsilon \leq e_2) = \alpha \qquad 2\text{-}31$$

then the interval between e_1 and e_2 is called the 100α% confidence interval for ε. For example if $\alpha = 0.95$, the interval is the 95% confidence interval. This means that the statement that ε lies between e_1 and e_2 will be true 95% of the time that such a claim is made.

In hypothesis testing we make an a priori statement (hypothesis) about the population (for example that it is normally distributed with mean μ and variance σ^2), and then based on the value of the sample statistics, test whether to accept or reject the hypothesis. There are four possibilities

a) hypothesis true and accepted,

b) hypthesis true but rejected (called a Type I error),

c) hypothesis false and rejected,

d) hypothesis false but accepted (called a Type II error).

If the probability of a Type I error is

$$P_r \text{ (hypothesis true but rejected)} = \alpha \qquad 2\text{-}32$$

then $100\alpha\%$ is called the <u>significance level</u> of the test. This probability can often be determined from the sampling distribution, and is the probability that a sample from the hypothesized population will have values for the sample statistics which indicate that the sample is from some other population.

If the probability of a Type II error is

$$P_r \text{ (hypothesis false but accepted)} = \beta \qquad 2\text{-}33$$

then $(1-\beta)$ is called the <u>power</u> of the test. This probability can be determined only for a restricted class of hypotheses. Therefore, although a Type II error is more serious than a Type I error, usually less can be said about its probability of occurrence.

To summarize statistical estimation, point estimates can be made without assuming a particular population distribution, however, both interval estimation and hypothesis testing require that a particular population distribution be assumed or specified.

3. STATISTICAL DISTRIBUTION FUNCTIONS

In this Chapter we introduce several distribution functions each of which serves as a mathematical representation of the variation of a given random variable over some domain. When one random variable is involved, the distribution is called univariate, while in the case of several random variables the distribution is called multivariate. We will discuss some special distributions which are derived from basic mathematical functions; they are the normal, chi-square, student's (t), and F distributions.

This Chapter is based on Hogg and Craig [1965] and Hamilton [1964].

3.1 THE NORMAL DISTRIBUTION

3.1.1 The Distribution Function

The basic mathematical function from which the normal distribution function is deduced, is

$$I = \int_{-\infty}^{\infty} \exp\,(-y^2/2)\, dy \quad . \qquad 3\text{-}1$$

The integral is evaluated by first squaring it, that is

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\,\left(-\frac{y^2+z^2}{2}\right) dy\, dz \quad , \qquad 3\text{-}2$$

and then transforming from Cartesian to polar coordinates as follows:

$$\begin{bmatrix} y \\ z \end{bmatrix} = r \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} . \qquad 3\text{-}3$$

Thus 3-2 becomes

$$I^2 = \int_0^{2\pi} \int_0^{\infty} \exp(\frac{-r^2}{2}) \; r \; dr \; d\theta$$

$$= \int_0^{2\pi} d\theta = 2\pi \quad , \qquad 3\text{-}4$$

and

$$I = (2\pi)^{1/2} \quad . \qquad 3\text{-}5$$

Knowing the value of the integral I, 3-1 becomes

$$\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp(\frac{-y^2}{2}) \; dy = 1 \; . \qquad 3\text{-}6$$

By making the following change of variable in the integration,

$$y = \frac{x-a}{b} \quad , \quad b > 0 \qquad 3\text{-}7$$

we see that the integral now becomes

$$\int_{-\infty}^{\infty} \frac{1}{b(2\pi)^{1/2}} \exp\left[-\frac{(x-a)^2}{2b^2}\right] dx = 1 \qquad 3\text{-}8$$

This integral has the properties of a cummulative distribution function (c.d.f.); its corresponding probability density function (p.d.f.) is

$$\boxed{\phi(x) = \frac{1}{b(2\pi)^{1/2}} \exp\left[-\frac{(x-a)^2}{2b^2}\right] \; . \quad \text{where } -\infty < x < \infty} \qquad 3\text{-}9$$

This p.d.f. is said to be that of a continuous normal random variable.

3.1.2 The Moment Generating Function

The moment generating function (m.g.f.) of a normal distribution is

$$M(t) = \int_{-\infty}^{\infty} e^{tx}\, \phi(x)\, dx$$

$$= \int_{-\infty}^{\infty} e^{tx}\, \frac{1}{b(2\pi)^{1/2}}\, \exp\left[-\frac{(x-a)^2}{2b^2}\right] dx \qquad 3\text{-}10$$

and by letting $y = \frac{x-a}{b} - bt$ we have

$$x = by + b^2t + a \ ,$$

$$M(t) = \int_{-\infty}^{\infty} \exp\left[t(by + b^2t + a)\right] \frac{1}{b(2\pi)^{1/2}}\, \exp\left[-\frac{(by+b^2t)^2}{2b^2}\right] b\,dy$$

$$= \exp\left[at + \frac{b^2t^2}{2}\right] \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}}\, \exp\left(\frac{-y^2}{2}\right) dy, \qquad 3\text{-}11$$

and the final result for M(t) is:

$$\boxed{M(t) = \exp\left[at + \frac{b^2t^2}{2}\right] .} \qquad 3\text{-}12$$

From equation 2-10b, the mean μ of distribution is related to its moment generating function by

$$\mu = M'(0).$$

For the normal distribution

$$M'(t) = M(t)\ (a + b^2t) \ .$$

By setting t=0, the result is

$$\boxed{\mu = M'(0) = a} \qquad 3\text{-}13$$

From equation 2-10c the variance σ^2 of a distribution is related to the m.g.f. by

$$\sigma^2 = M''(0) - [M'(0)]^2 \ .$$

For the normal distribution

$$M''(t) = M(t)\ (b^2) + M(t)\ (a + b^2t)^2 \ ,$$

thus

$$\boxed{\sigma^2 = M''(0) - [M'(0)]^2 = (b^2+a^2) - a^2 = b^2} \qquad 3\text{-}14$$

proof for the above is as follows.

Given the cumulative distribution function

$$\Phi(\omega) = P_r\left(\frac{x-\mu}{\sigma} \leq \omega\right) = P_r\left(x \leq \omega\sigma + \mu\right)$$

or in integral form

$$\Phi(\omega) = \int_{-\infty}^{\omega\sigma+\mu} \frac{1}{\sigma(2\pi)^{1/2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \quad .$$

With the change of variable $y = (x-\mu)/\sigma$

$$\Phi(\omega) = \int_{-\infty}^{\omega} \frac{1}{(2\pi)^{1/2}} \exp\left(\frac{-y^2}{2}\right) dy,$$

and the corresponding p.d.f. is

$$\phi(\omega) = \Phi'(\omega) \quad ,$$

$$\boxed{\phi(\omega) = \frac{1}{(2\pi)^{1/2}} \exp\left(\frac{-\omega^2}{2}\right) \quad .} \qquad 3\text{-}17$$

Comparing the above equation to 3-15 it is evident that $\mu = 0$ and $\sigma^2 = 1$, thus the proof is completed.

The graph of $n(0, 1)$ has similar characteristics to $n(\mu, \sigma^2)$; that is substituting $\mu = 0$ and $\sigma^2 = 1$ we get:

1) symmetry about the vertical axis through x=0,

2) maximum value of $1/[(2\pi)^{1/2}]$ at x=0,

3) x axis as horizontal asymtote,

4) points of inflection at $x = \pm\sigma$.

3.1.5 Computations Involving the Normal Distribution

We have seen that the mean μ and variance σ^2 are two parameters of a univariate normal p.d.f. To facilitate computations, precomputed tables have been prepared by statisticians, the arguments of which are in part a function of the parameters of the distribution. The arguments of the normal distribution (Appendix B-1) are the probability Pr and the abscissa value c. The abscissa value is a particular value of the independent variable of the p.d.f. which corresponds to a given probability value.

The direct problem is to enter the table with an abscissa value and exit the table with a probability value; while the inverse problem is to enter with a probability and exit with an abscissa value c.

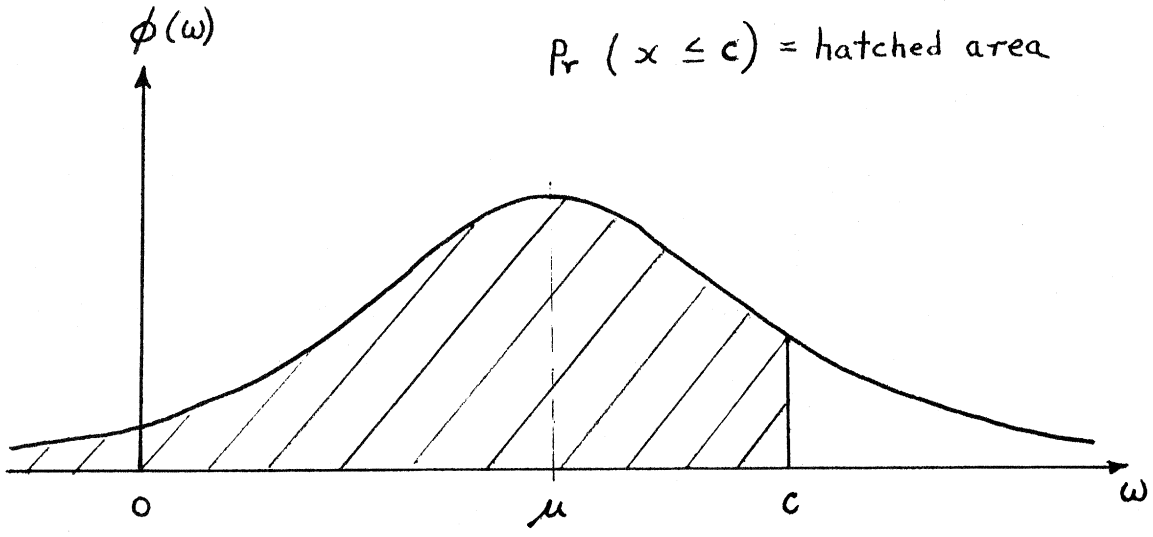
Basic to the solution of problems associated with the normal distribution is the following relationship between the theoretical probability Pr, the abscissa value c, and the tabulated probability N. If a random variable x is $n(\mu, \sigma^2)$, then the probability that x is less than or equal to some value c is computed from (see Figure 3-2):

$$\Pr(x \leq c) = \Pr\left(\frac{x-\mu}{\sigma} \leq \frac{c-\mu}{\sigma}\right)$$

$$= \int_{-\infty}^{(c-\mu)/\sigma} \frac{1}{(2\pi)^{1/2}} \exp\left(\frac{-\omega^2}{2}\right) d\omega$$

$$= N\left(\frac{c-\mu}{\sigma}\right) .$$

Figure 3-2



PROBABILITY - NORMAL DISTRIBUTION

The value N for the above integral is tabulated for a random variable n(0, 1). Normalization, that is $(x-\mu)/\sigma$, allows probabilities associated with $x[n(\mu, \sigma^2)]$ to be expressed and computed in terms of probabilities of $\omega[n(0, 1)]$.

Example 1 - Direct Problem, One Abscissa

Given: x is n(2, 16) that is $\mu=2$, $\sigma^2=16$

Required: Pr $(x \leq 4)$

Solution:

$$\Pr\left(\frac{x-\mu}{\sigma} \leq \frac{c-\mu}{\sigma}\right) = N\left(\frac{c-\mu}{\sigma}\right) = N\left(\frac{4-2}{4}\right) = N(0.5)$$

from Table B-1 $N(0.5) = \underline{0.6915}$.

Example 2 - Direct Problem, Two Abscissa

Given: x is n(2, 16) that is $\mu=2$, $\sigma^2=16$

Required: Pr $(1 \leq x \leq 4)$

Solution:

$$\Pr(c_1 \leq x \leq c_2) = \Pr\left(\frac{c_1-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{c_2-\mu}{\sigma}\right)$$

$$= \Pr\left(\frac{x-\mu}{\sigma} \leq \frac{c_2-\mu}{\sigma}\right) - \Pr\left(\frac{x-\mu}{\sigma} \leq \frac{c_1-\mu}{\sigma}\right)$$

$$= N\left(\frac{c_2-\mu}{\sigma}\right) - N\left(\frac{c_1-\mu}{\sigma}\right) = N\left(\frac{4-2}{4}\right) - N\left(\frac{1-2}{4}\right) = N(0.5) - N(-0.25)$$

$$= N(0.5) - [1-N(0.25)] = N(0.5) + N(0.25) - 1$$

From Table B-1 $N(0.5) + N(0.25) - 1 = 0.6915 + 0.5987 - 1 = \underline{0.2902}$.

Example 3 - Inverse Problem, One Abscissa

Given: x is n(2, 16) that is $\mu=2$, $\sigma^2=16$

Required: Find c such that Pr $(x \leq c) = 0.95$

Solution:

$$\Pr\left(\frac{x-\mu}{\sigma} \leq \frac{c-\mu}{\sigma}\right) = 0.95$$

$$N\left(\frac{c-\mu}{\sigma}\right) = 0.95$$

From Table B-1 $\left(\frac{c-\mu}{\sigma}\right) = 1.645$

$$c = 1.645\sigma + \mu = \underline{8.56}.$$

Example 4 - Inverse Problem Two Abscissa

Given: x is $n(\mu, \sigma^2)$

Required: Find c such that $Pr(|x-\mu| \leq c-\mu) = Pr[-(c-\mu) \leq x-\mu \leq c-\mu] =$ $= 0.95$

Solution:

$$Pr\left(-\frac{c-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \leq \frac{c-\mu}{\sigma}\ ) = 0.95$$

$$Pr\left(\frac{x-\mu}{\sigma} \leq \frac{c-\mu}{\sigma}\right) - Pr\left(\frac{x-\mu}{\sigma} \leq -\frac{c-\mu}{\sigma}\right) = 0.95$$

$$N\left(\frac{c-\mu}{\sigma}\right) - N\left(-\frac{c-\mu}{\sigma}\right) = 0.95$$

$$N\left(\frac{c-\mu}{\sigma}\right) - \left[1 - N\left(\frac{c-\mu}{\sigma}\right)\right] = 0.95$$

$$N\left(\frac{c-\mu}{\sigma}\right) = \frac{1 + 0.95}{2} = 0.975$$

From Table B-1 $\left(\frac{c-\mu}{\sigma}\right) = 1.96$

$$c = 1.96\ \sigma + \mu$$

(Note when $\mu = 0$, $c \simeq 2\sigma$ for $Pr = 0.95$).

3.1.6 Multivariate Normal Distribution

The normal distribution pertaining to a single random variable has been given. When several parameters (random variables) are being estimated simultaneously, the normal distribution characterizing all

these parameters together is called a multivariate normal distribution. An example of this is in the case of a geodetic control network where the coordinates of all the stations are being estimated and are thus considered as random variables and are said to have a multivariate normal distribution.

For m random variables, the m-dimensional multivariate normal p.d.f. is

$$\boxed{\phi\ (X) = C \exp\ [-\frac{(X-U)^T\ \Sigma_x^{-1}\ (X-U)}{2}\ ]\ ,} \qquad 3\text{-}22$$

where the vector of random variables is

$$\underset{mx1}{X} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_m \end{bmatrix}\ , \qquad 3\text{-}23$$

with corresponding means

$$\underset{mx1}{U} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ \mu_m \end{bmatrix}\ , \qquad 3\text{-}24$$

and covariance matrix

$$\underset{mxm}{\Sigma}x = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_m} \\ \sigma_{x_2x_1} & \sigma^2_{x_2} & & \\ & & \ddots & \\ \sigma_{x_mx_1} & & & \sigma^2_{x_m} \end{bmatrix}; \qquad 3\text{-}25$$

the constant

$$C = \frac{[\det\ (\Sigma_x^{-1})]^{1/2}}{(2\pi)^{m/2}}\ . \qquad 3\text{-}26$$

Note the similarities of the univariate normal p.d.f. (3-15) with the multivariate normal p.d.f. (3-22), namely

$$\text{a)}\quad \left(\frac{1}{\sigma^2}\right)^{1/2} \quad \text{vs.} \quad [\det(\Sigma_x^{-1})]^{1/2} \quad ,$$

$$\text{b)}\quad \frac{1}{(2\pi)^{1/2}} \quad \text{vs.} \quad \frac{1}{(2\pi)^{m/2}} \quad ,$$

$$\text{c)}\quad \frac{(x-\mu)^2}{2\sigma^2} \quad \text{vs.} \quad \frac{(X-U)^T \Sigma_x^{-1} (X-U)}{2} \quad ;$$

For zero means the multivariate normal, p.d.f. is

$$\boxed{\phi(X) = C \exp\left(-\frac{X^T \Sigma_x^{-1} X}{2}\right) .} \qquad 3\text{-}27$$

3.2 THE CHI-SQUARE DISTRIBUTION

3.2.1 The Distribution Function

The Chi-Square distribution is a special case of the gamma distribution, with the latter being derived from the following integral called the gamma function of α:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy \quad , \qquad 3\text{-}28$$

where the integral exists for $\alpha > 0$ and has a positive value. When $\alpha = 1$

$$\Gamma(1) = \int_0^\infty e^{-y}\, dy = 1 \quad , \qquad 3\text{-}29$$

and if $\alpha > 1$, then integration by parts shows that

$$\Gamma(\alpha) = (\alpha-1) \int_0^\infty y^{\alpha-2} e^{-y}\, dy = (\alpha-1)\,\Gamma(\alpha-1). \qquad 3\text{-}30$$

Further, if α is a positive integer and greater than one,

$$\Gamma(\alpha) = (\alpha-1)\ (\alpha-2)\ \ldots\ (3)\ (2)\ (1)\Gamma(1) = (\alpha-1)!$$

Making the change of variable $y = x/\beta$ in the integral for $\Gamma(\alpha)$ for $\beta > 0$ yields,

$$\Gamma(\alpha) = \int_0^\infty \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \frac{1}{\beta}\, dx$$

and

$$\boxed{1 = \int_0^\infty \frac{1}{\Gamma(\alpha)\ \beta^\alpha}\ x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) dx} \qquad 3\text{-}31$$

Note the integral equals unity. Since the above integral meets the requirements of a cumulative distribution function, the corresponding p.d.f. is

$$\boxed{\begin{aligned}\phi(x) &= \frac{1}{\Gamma(\alpha)\ \beta^\alpha}\ x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) ;(0 < x < \infty) \\ &= 0 \text{ elsewhere}\end{aligned}} \qquad 3\text{-}32$$

and is said to have a gamma distribution with parameters α and β.

As mentioned earlier, the Chi-square distribution is a special case of a gamma distribution in which

$$\alpha = \frac{\nu}{2} \ ,$$

ν being a positive integer, and $\beta = 2$. Thus from 3-32, a random variable χ of the continuous type is said to have a Chi-square p.d.f. if it has the form

$$\boxed{\begin{aligned}\phi(x) &= \frac{1}{\Gamma(\frac{\nu}{2})2^{\nu/2}}\ x^{(\nu/2-1)}\ e^{-x/2} ;(0 \leq x < \infty) \\ &= 0 \text{ elsewhere .}\end{aligned}} \qquad 3\text{-}33$$

Note the distribution is defined by the parameter ν which is called the number of degrees of freedom. The number of degrees of freedom is a very practical quantity and has a relationship to the least squares estimation problem discussed in Chapter 6. A continuous random variable having the above p.d.f. is written in abbreviated form as $\chi^2(\nu)$.

3.2.2 The Moment Generating Function

The moment generating function for the Chi-square distribution is derived from the basic definition as

$$M(t) = \int_0^\infty e^{tx}\,\phi(x)\,dx$$

$$= \int_0^\infty e^{tx}\,\frac{1}{\Gamma(\frac{\nu}{2})2^{\nu/2}}\,x^{(\frac{\nu}{2}-1)}\exp(-\frac{x}{2})\,dx$$

$$= \int_0^\infty \frac{1}{\Gamma(\frac{\nu}{2})2^{\nu/2}}\,x^{(\frac{\nu}{2}-1)}\exp(\frac{-x(1-2t)}{2})\,dx. \quad 3\text{-}34$$

The change of variable $y = x(1-2t)/2$ or $x=2y/(1-2t)$, yields

$$M(t) = \int_0^\infty \frac{2/(1-2t)^{\nu/2}}{\Gamma(\frac{\nu}{2})2^{\nu/2}}\,(\frac{2y}{1-2t})^{\frac{\nu}{2}-1}\,e^{-y}\,dy$$

$$= \left(\frac{1}{1-2t}\right)^{\nu/2}\int_0^\infty \frac{1}{\Gamma(\frac{\nu}{2})}\,y^{(\frac{\nu}{2}-1)}\,e^{-y}\,dy$$

$$\boxed{M(t) = \frac{1}{(1-2t)^{\frac{\nu}{2}}}} \quad 3\text{-}35$$

Computing

$$M'(t) = (-\frac{\nu}{2})\,(1-2t)^{(\frac{\nu}{2}-1)}(-2)$$

and

$$M''(t) = (-\frac{\nu}{2})\,(-\frac{\nu}{2}-1)\,(1-2t)^{(-\frac{\nu}{2}-2)}(-2)^2$$

the mean and variance of the Chi-squared distribution respectively become

$$\boxed{\mu = M'(0) = \nu} \quad 3\text{-}36$$

$$\sigma^2 = M''(0) - [M'(0)]^2 = (\nu^2+2\nu)-\nu^2 = 2\nu \qquad 3\text{-}37$$

3.2.3 The Graph of the Chi-Square Distribution

The graph of a χ^2 distribution has the following characteristics (see Figure 3-3):

a) a value of zero when x=0,

b) a maximum value in the interval $0 < x < \infty$,

c) the positive x-axis as an asymtote,

d) has one point of inflection on each side of the maximum.

3.2.4 Computations Involving the Chi-Square Distribution

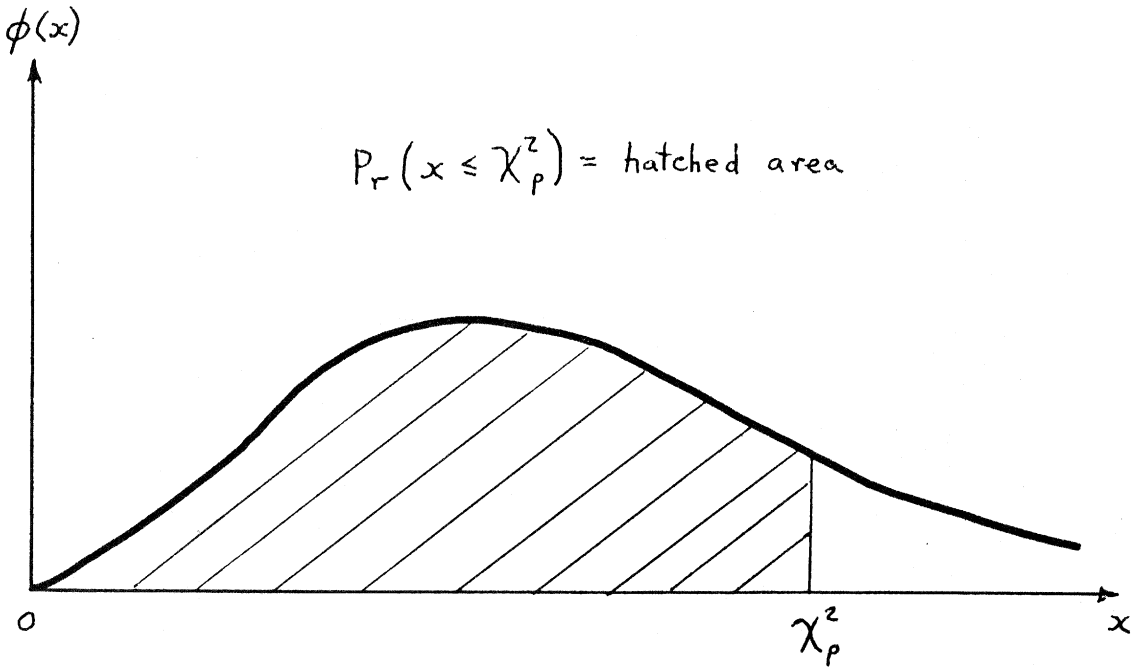
The possible arguments with which to enter the Chi-square table are the Probability Pr, the abscissa value χ_p^2 and chi-square distribution parameter - the degrees of freedom ν. The direct problem is to enter the table with χ_p^2 and ν and exit with Pr, while the inverse problem is to enter with Pr and ν and exit with χ_p^2.

The use of the tables (Appendix B-2) is based on the following relationship between the probability Pr, χ_p^2, and ν. If a random variable x is $\chi^2(\nu)$, then the

$$\text{pr}\ (x \leq \chi_p^2) = \int_0^{\chi_p^2} \frac{1}{\Gamma(\frac{\nu}{2})2^{\nu/2}} x^{\left(\frac{\nu}{2}-1\right)} e^{-\frac{x}{2}}\, dx$$

The above integral has been precomputed and the results tabulated in the body of the table for particular values of χ_p^2 which correspond to different values of ν and Pr; these values χ_p^2 are called <u>percentiles</u> of the <u>chi square distribution</u>, and Pr takes on certain probability

Figure 3-3.



GRAPH OF THE CHI-SQUARE DISTRIBUTION

values between 0 and 1.

Example 1 - Direct Problem, One Abscissa

Given: x is χ^2 (10) that is $\nu = 10$

Required: Pr ($x \leq 18.31$)

Solution: From Table B-2

$$\chi^2_{0.95} (10) = 18.31$$

$$Pr = \underline{0.95}$$

Example 2 - Direct Problem, Two Abscissa

Given: x as χ^2 (20) that is $\nu = 20$

Required: Pr ($34.17 \geq x \geq 9.59$)

Solution: From Table B-2

$$\chi^2_{0.975} = 34.17 \text{ and } \chi^2_{0.025} = 9.59$$

$$Pr (\chi^2_{0.975} \geq x \geq \chi^2_{0.025})$$

$$= Pr (x \leq \chi^2_{0.975}) - Pr (x \leq \chi^2_{0.025})$$

$$= 0.975 - 0.025 = \underline{0.95}$$

Example 3 - Inverse Problem, One Abscissa

Given: x is $\chi^2(10)$ that is $\nu = 10$

Required: χ^2_p such that Pr ($x \leq \chi^2_p$) = 0.90

Solution:

$$Pr (x \leq \chi^2_{0.90}) = 0.90$$

From Table B-2 $\chi^2_{0.90} = \underline{15.99}$

Example 4 - Inverse Problem, Two Abscissa

Given: x is $\chi^2(20)$ that is $\nu = 20$

Required: $\chi^2_{p_1}, \chi^2_{p_2}$ such that Pr $(\chi^2_{p_1} \leq x \leq \chi^2_{p_2}) = 0.99$

Solution: $\chi^2_{p_1}$ and $\chi^2_{p_2}$ will be chosen such that the remaining probability of 0.01 is divided equally, thus

$P_1 = 0.005$ and $P_2 = 0.995$.

$$P_r\ (\chi^2_{0.005} \leq x \leq \chi^2_{0.995}) = 0.99$$

$$P_r\ (x \leq \chi^2_{0.995}) - P_r\ (x \leq \chi^2_{0.005}) =$$

$$= 0.995 - 0.005 = 0.99$$

From Table B-2 $\chi^2_{0.005} = \underline{7.43}$

$\chi^2_{0.995} = \underline{40.00}$

3.3 The Student's t Distribution

3.3.1 The Distribution Function

The (student's) t distribution is derived on the basis of the normal and chi-square distributions, and is useful in the statistical procedures to be described in Chapter 5.

Let us first consider two random variables ω, which is $n(0,1)$ and v, which is $\chi^2(\nu)$; ω and v are stipulated to be statistically independent. The joint p.d.f. of ω and v is the product of the two individual p.d.f.'s, namely

$$\phi(\omega, v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\omega^2}{2}\right) \frac{1}{\Gamma(\frac{\nu}{2}) 2^{\nu/2}}\ v^{(\frac{\nu}{2} - 1)} \exp\ (-\frac{v}{2}) \qquad 3\text{-}38$$

$$\left\{\begin{array}{c} -\infty < \omega < \infty \\ 0 < \nu < \infty \end{array}\right\}$$

$= 0$ elsewhere.

Next consider the definition of third variable t as

$$t = \frac{\omega}{(\nu/\nu)^{1/2}} . \qquad 3\text{-}39$$

The p.d.f. corresponding to the two original variables ω and ν can be transformed into a new p.d.f., e.g. in terms of the new variables t and u through the transformation equations

$$t = \frac{\omega}{(\nu/\nu)^{1/2}} , \quad u = \nu, \qquad 3\text{-}40$$

or

$$\omega = \frac{tu^{1/2}}{\nu^{1/2}} , \quad \nu = u. \qquad 3\text{-}41$$

The Jacobian of the transformation (see Wells [1971]) is

$$|J| = \left|\begin{array}{cc} \frac{\partial\omega}{\partial t} & \frac{\partial\omega}{\partial u} \\ \frac{\partial\nu}{\partial t} & \frac{\partial\nu}{\partial u} \end{array}\right| = \left|\begin{array}{cc} (\frac{u}{\nu})^{1/2} & \frac{t}{2}(u\nu)^{-1/2} \\ 0 & 1 \end{array}\right| = (\frac{u}{\nu})^{1/2}, \qquad 3\text{-}42$$

and the new p.d.f. is

$$\phi(t,u) = \phi(\omega,\nu)|J| = \phi\left(\frac{tu^{1/2}}{\nu^{1/2}} , u\right)|J|$$

$$= \frac{1}{(2\pi)^{1/2}\,\Gamma(\frac{\nu}{2})2^{\nu/2}}\, u^{(\nu/2-1)}$$

$$\exp\left[-\frac{u}{2}\left(1+\frac{t^2}{\nu}\right)\right](\frac{u}{\nu})^{1/2} \qquad 3\text{-}43$$

$$\left\{\begin{array}{c} -\infty < t < \infty \\ 0 < u < \infty \end{array}\right\}$$

$= 0$ elsewhere.

Since we are interested in t only, u is integrated out of the above expression; the following result is then the marginal p.d.f. corresponding to t:

$$\phi(t) = \int_{-\infty}^{\infty} \phi(t,u)du$$

$$= \int_{o}^{\infty} \frac{1}{(2\pi)^{1/2}\,\Gamma(\frac{\nu}{2})\,2^{\nu/2}}\, u^{[(\nu+1)/2-1]}$$

$$\exp\left[-\frac{u}{2}\left(1 + \frac{t^2}{\nu}\right)\right]du \quad . \qquad 3\text{-}45$$

The change of variable in the integration

$$Z = u[1 + (\frac{t^2}{\nu})]/2 \qquad 3\text{-}46$$

yields

$$\phi(t) = \int_{o}^{\infty} \frac{1}{(2\pi\nu)^{1/2}\,\Gamma(\frac{\nu}{2})\,2^{\nu/2}} \left(\frac{2Z}{1 + t^2/\nu}\right)^{(\nu+1)/2-1} \exp(-z)\left(\frac{2}{1 + t^2/\nu}\right)dZ \qquad 3\text{-}47$$

$$\boxed{\phi(t) = \frac{\Gamma[(\nu+1)/2]}{(\pi\nu)^{1/2}\Gamma(\frac{\nu}{2})} \quad \frac{1}{(1+t^2/\nu)^{(\nu+1)/2}}, \quad -\infty < t < \infty} \qquad 3\text{-}48$$

The random variable t is said to have the above t distribution if

$$t = \frac{\omega}{(v/\nu)^{1/2}}$$

where ω is $n(0,1)$ and v is $\chi^2(\nu)$, and is written in the abbreviated form $t(\nu)$. Note that the degrees of freedom ν is the single parameter defining the distribution.

3.3.2 The Graph of the t Distribution

The graph of the t distribution is rather complicated in that it is an intricate combination of a normal curve and a chi-square

curve. It is similar to the normal curve in the following respects (Figure 3-4):

1) $\phi(t)$ has values for $-\infty < t < \infty$,

2) the maximum value of $\phi(t)$ is at $t = 0$,

3) has the t axis as its horizontal asymtote,

4) has two points of inflection one on each side of the maximum.

3.3.3 Computations Involving the t Distribution

As for the chi-square distribution, the arguments for entering the t table (Appendix B-3) are Pr, t_p, and ν. The direct and inverse problems are the same.

The use of the tables is based on the following relationship between Pr, t_p, and ν. If a random variable x is $t(\nu)$, then the

$$\Pr\ (x \leq t_p) = \int_{-\infty}^{t_p} \phi(t)\ dt\ ,$$

where $\phi(t)$ is the t p.d.f. of 3-48. The body of the table contains percentiles t_p of the t distribution corresponding to certain degrees of freedom and probability values between 0 and 1.

Example 1 - Direct Problem, One Abscissa

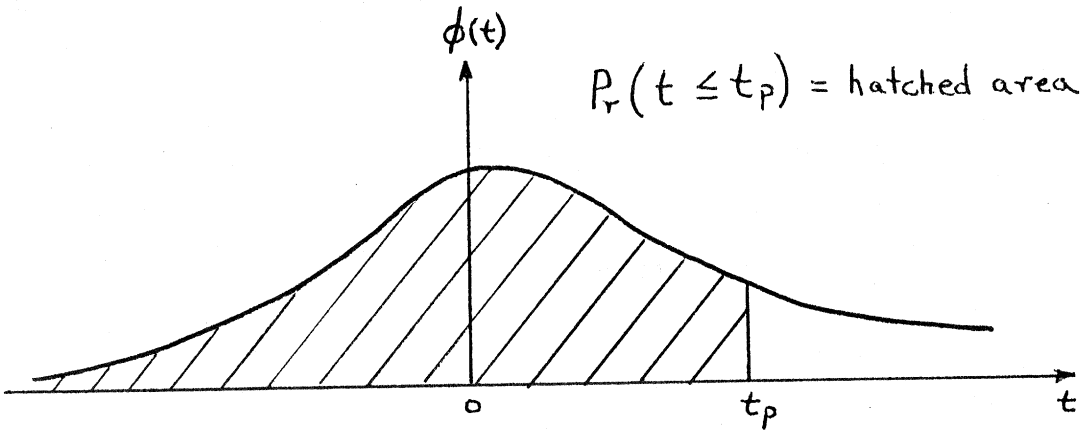
Given: x is t(10) that is $\nu = 10$

Required: $\Pr\ (x \leq 1.372)$

Solution: from Table B-3 $t_{0.90} = 1.372$

$$\therefore \quad P_r = \underline{0.90}$$

Figure 3-4.



GRAPH OF THE t DISTRIBUTION

Example 2 - Direct Problem

Given: x is t(10) that is $\nu = 10$

Required: $\Pr (|x| \geq 2.228)$

Solution: From Table B-3 $t_{0.975} = 2.228$

therefore $\Pr (x \geq t_{0.975})$

$= 1 - \Pr (x \leq t_{0.975})$

$= 1 - 0.975 = 0.025$

since x can also be negative $\Pr = 2(0.025) = \underline{0.05}$

Example 3 - Inverse Problem, Two Abscissa

Given: x as t(14) that is $\nu = 14$

Required: t_p such that $\Pr(-t_p \leq x \leq t_p) = 0.90$

Solution: $\Pr(-t_p \leq x \leq t_p)$

$= \Pr (x \leq t_p) - \Pr (x \leq -t_p)$

$= \Pr (x \leq t_p) - [1 - \Pr (x \leq t_p)]$

$= 2 \Pr (x \leq t_p) - 1 = 0.90$

or $\Pr (x \leq t_p) = 0.95$

From Table B-3 $t_{0.95} = \underline{1.761}$

3.4 THE F DISTRIBUTION

3.4.1 The Distribution Function

The F distribution is derived on the basis of two chi-square distributions and is the last of the basic distributions to be covered in this Chapter.

Let us first consider two independent chi-square random variables u and v having ν_1 and ν_2 degrees of freedom, respectively. The joint p.d.f. of u and v is

$$\phi(u, v) = \frac{1}{\Gamma(\frac{\nu_1}{2})\,\Gamma(\frac{\nu_2}{2})\,2^{(\nu_1+\nu_2)/2}}\; u^{\left(\frac{\nu_1}{2}-1\right)} v^{\left(\frac{\nu_2}{2}-1\right)} e^{-(u+v)/2} \qquad 3\text{-}49$$

$$0 < u < \infty$$

$$0 < v < \infty$$

$$= 0 \text{ elsewhere.}$$

Next consider a new random variable

$$f = \frac{u/\nu_1}{v/\nu_2} \qquad 3\text{-}50$$

whose marginal p.d.f. $\phi(f)$ is to be determined. The transformation equations are

$$f = \frac{u/\nu_1}{v/\nu_2}\,, \quad z = v \qquad 3\text{-}51$$

or

$$u = \frac{f\,\nu_1\,z}{\nu_2}\,, \quad v = z\,, \qquad 3\text{-}52$$

with the Jacobian of the transformation being

$$|J| = \begin{vmatrix} \frac{\partial u}{\partial f} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial f} & \frac{\partial v}{\partial z} \end{vmatrix} = \begin{vmatrix} \frac{\nu_1 z}{\nu_2} & \frac{\nu_1 f}{\nu_2} \\ 0 & 1 \end{vmatrix} = \frac{z\nu_1}{\nu_2} \qquad 3\text{-}53$$

The joint p.d.f. of the random variables f and z is then

$$\phi(f,z) = \phi(u,v)\ \det(J) = \frac{1}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})\ 2^{(\nu_1+\nu_2)/2}} \left(\frac{\nu_1 z f}{\nu_2}\right)^{\left(\frac{\nu_1}{2}-1\right)} z^{\left(\frac{\nu_2}{2}-1\right)}$$

3-54

$$\exp\left[\frac{-z}{2}\left(\frac{\nu_1 f}{\nu_2}+1\right)\right]\frac{\nu_1 z}{\nu_2} \quad ,$$

The marginal p.d.f. $\phi(f)$ is obtained by integrated out z, namely

$$\phi(f) = \int_{-\infty}^{\infty} \phi(f,\ z)\ dz$$

$$= \int_{o}^{\infty} \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} (f)^{\left(\nu_1/2\ -\ 1\right)}}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})\ 2^{(\nu_1+\nu_2)/2}} z^{(\nu_1+\nu_2)/2-1} \exp[-\frac{z}{2}(\frac{\nu_1 f}{\nu_2}+1)]\ dz. \qquad 3\text{-}55$$

By making the following change of variable

$$y = \frac{z}{2}\left(\frac{\nu_1 f}{\nu_2}+1\right) \quad , \qquad 3\text{-}56$$

$\phi(f)$ becomes

$$\phi(f) = \int_{o}^{\infty} \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} (f)^{\left(\nu_1/2\ -\ 1\right)}}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})\ 2^{(\nu_1+\nu_2)/2}} \left(\frac{2y}{\nu_1 f/\nu_2+1}\right)^{(\nu_1+\nu_2)/2-1} e^{-y} \left(\frac{2}{\nu_1 f/\nu_2+1}\right) dy \ , \qquad 3\text{-}57$$

and after integrating,

$$\phi(f) = \frac{\Gamma[(\nu_1+\nu_2)/2]\ (\nu_1/\nu_2)^{\nu_1/2}}{\Gamma(\frac{\nu_1}{2})\ \Gamma(\frac{\nu_2}{2})} \frac{(f)^{\nu_1/2-1}}{(1+\nu_1 f/\nu_2)^{(\nu_1+\nu_2)/2}}$$

$$(0 < f < \infty) \qquad 3\text{-}58$$

$$= 0 \text{ elsewhere.}$$

The random variable

$$f = \frac{u/\nu_1}{v/\nu_2}$$

where u is $\chi^2(\nu_1)$ and v is $\chi^2(\nu_2)$, is said to have the above F distribution, and is written in the abbreviated form $F(\nu_1,\nu_2)$. Note that two degrees of freedom ν_1 and ν_2 are the sole defining parameters of this distribution.

A very useful fact is that 1/f has an F distribution with parameters ν_2 and ν_1. This result can be proved by a procedure similar to the one used above. That is, $\frac{1}{F_p(\nu_1, \nu_2)} = F_{1-p}(\nu_2, \nu_1)$. Note also $\nu_1 F(\nu_1, \infty) = \chi^2(\nu_1)$.

3.4.2 The Graph of the F Distribution

The graph of the F distribution is rather complicated as it is an intricate combination of two chi-square distributions. It has the characteristics (Figure 3-5) similar to the chi-square distribution.

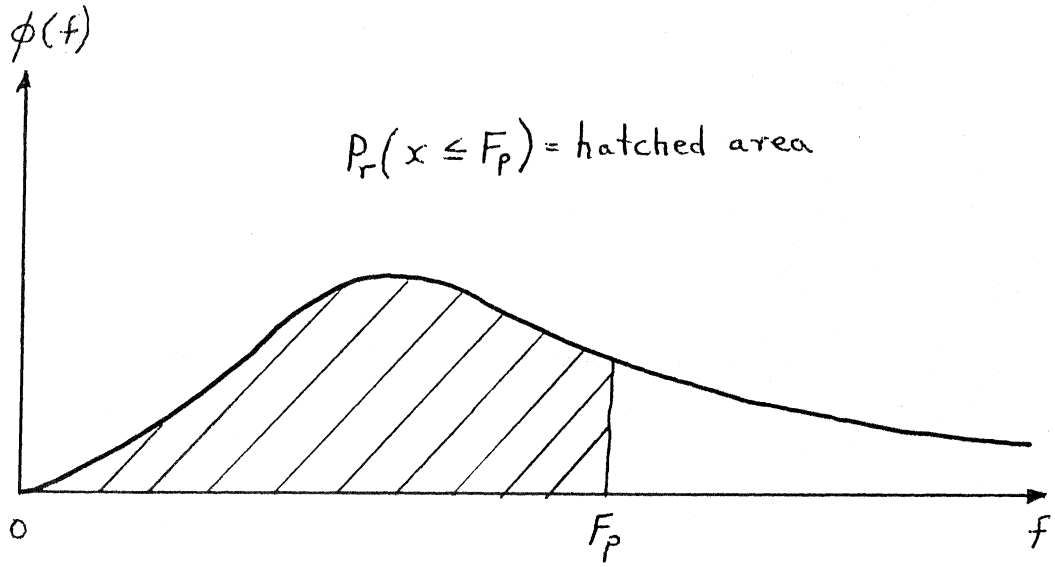
3.4.3 Computations Involving the F Distribution

The possible arguments with which to enter the F-tables (Appendix B-4) are the probability Pr, the abscissa value F_p, and the two degrees of freedom ν_1 and ν_2. The direct problem is to enter the table with F_p, ν_1 and ν_2, and exit with Pr. The inverse problem is to enter with Pr, ν_1 and ν_2, and exit with F_p.

The use of the tables is based on the relationship between Pr, F_p, ν_1, and ν_2, that is, if x is $F(\nu_1, \nu_2)$, then

$$Pr(x \leq F_p) = \int_0^{F_p} \phi(f)\, df \ ,$$

Figure 3-5.



GRAPH OF THE F DISTRIBUTION

where $\phi(f)$ is the F p.d.f. given by 3-58. The above integral has been precomputed and the results tabulated for particular values of F_p which correspond to different values of ν_1 and ν_2 and Pr; these values are called the percentiles of the F distribution, where Pr takes on certain probability values between 0 and 1.

Example 1 - Direct Problem, One Abscissa

Given: x is F(5, 10) that is $\nu_1 = 5$ and $\nu_2 = 10$

Required: Pr $(x \leq 2.52)$

Solution: From the first of Tables B-4

$$F_{0.90}\,(5, 10) = 2.52$$

$$\text{Pr} = \underline{0.90}$$

Example 2 - Inverse Problem, One Abscissa

Given: x is F(4, 8) that is $\nu_1 = 4$ and $\nu_2 = 8$

Required: F_p such that Pr $(x \leq F_p) = 0.95$

Solution: From the second set of Tables B-4

$$F_{0.95}\,(4, 8) = \underline{3.84}$$

Example 3 - Inverse Problem, One Abscissa

Given: x is F(4, 8)

Required: F_p such that Pr $(x \leq F_p) = 0.05$

Solution:
$$\text{Pr}\,(x \leq F_p) = \text{Pr}\left(\frac{1}{x} \geq \frac{1}{F_p}\right)$$

$$= 1 - \text{Pr}\left(\frac{1}{x} \leq \frac{1}{F_p}\right) = 0.05$$

or

$$\text{Pr}\left(\frac{1}{x} \leq \frac{1}{F_p}\right) = 0.95$$

where $F_p = F_{0.05}(4,8)$

Recall that if x is F(4,8) then $\frac{1}{x}$ is F(8,4), which gives a second probability statement for $\frac{1}{x}$, that is

$$\Pr\left(\frac{1}{x} \leq F_{0.95}(8,4)\right) = 0.95 \ .$$

From the second of Tables B-4

$$F_{0.95}(8,4) = 6.04$$

therefore, equating the two probability statements for $\frac{1}{x}$ we can solve for F_p, that is

$$F_p = \frac{1}{F_{0.95}(8,4)} = \frac{1}{6.04} = \underline{0.166} \ .$$

Example 4 - Inverse Problem, Two Abscissa

Given: x is F(5,10)

Required: F_{p_1} and F_{p_2} such that $\Pr\left(F_{p_1} \leq x \leq F_{p_2}\right) = 0.90$

Solution: $\Pr\left(F_{p_1} \leq x \leq F_{p_2}\right) = \Pr\left(x \leq F_{p_2}\right) - \Pr\left(x \leq F_{p_1}\right)$

F_{p_1} and F_{p_2} will be chosen such that the remaining probability of 0.10 is divided equally, thus

a) $\Pr\left(x \leq F_{p_2}\right) = 0.95$ where $F_{p_2} = F_{0.95}(5,10)$

b) $\Pr\left(x \leq F_{p_1}\right) = 0.05$ where $F_{p_1} = F_{0.05}(5,10)$

Taking a) from the second of Tables B-4

$$F_{0.95}(5,10) = \underline{3.33}$$

Taking b) $\Pr\left(x \leq F_{p_1}\right) = \Pr\left(\frac{1}{x} \geq \frac{1}{F_{p_1}}\right)$

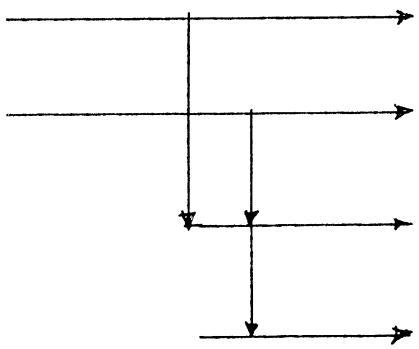
$$= 1 - \Pr\left(\frac{1}{x} \leq \frac{1}{F_{p_1}}\right) = 0.05$$

$$\Pr\left(\frac{1}{x} \leq \frac{1}{F_{p_1}}\right) = 0.95$$

and $\Pr\left(\frac{1}{x} \leq F_{0.95}(10,5)\right) = 0.95$ as in example 3.

From the second of Tables B-4 $F_{0.95}(10,5) = 4.74$ and from $F_p(n_1, n_2) = \frac{1}{F_{1-p}(n_2,n_1)}$ we have $F_{0.05}(5, 10) = \frac{1}{4.74} = 0.211$.

3.5 SUMMARY OF THE BASIC DISTRIBUTIONS



Normal $\quad n\ (0,\ 1)$

Chi-square $\quad \chi^2\ (\nu)$

Student's $\quad t(\nu) = \dfrac{n\ (0,\ 1)}{(\chi^2(\nu)/\nu)^{1/2}}$

F $\quad F = \dfrac{\chi^2(\nu_1)/\nu_1}{\chi^2(\nu_2)/\nu_2}$

4. DISTRIBUTIONS OF FUNCTIONS OF RANDOM VARIABLES

We have introduced the normal, chi-square, student's (t), and F distributions in Chapter 3. Any random variable having a p.d.f. corresponding to any of these distributions was said to have that particular distribution. We now introduce several very useful random variables which are functions of these random variables. A function containing one or more random variables that does not depend upon any unknown parameter is called a <u>statistic</u>. Two examples of statistics are

$$y = \sum_{i=1}^{n} x_i$$

where the x_i are $n(\mu, \sigma^2)$, and

$$y = \left(\frac{x_i - \mu}{\sigma}\right)^2 ,$$

where μ and σ are known.

Two other statistics are the <u>mean</u> of the sample

$$\bar{x} = \frac{x_1 + x_2 + \cdots x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

and the <u>variance</u> of the sample

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

In this Chapter, we will derive the distributions of these and other statistics which serve two purposes:

1) used in the derivations of the distributions of other functions of random variables,

2) used as "test statistics" in Chapter 5 on hypothesis testing.

4.1 DISTRIBUTION OF A NORMALIZED NORMAL RANDOM VARIABLE

Given: A random sample $x_1, x_2 \ldots x_n$, where the x_i are all independent and $x_i \overset{d}{\rightarrow} n(\mu, \sigma^2)$

Required to prove:

$$\boxed{\frac{x-\mu}{\sigma} \overset{d}{\rightarrow} n\ (0, 1)}$$

Proof: The proof was given in section 3.1 where the main idea was to take the p.d.f. of x and make the change of variable $y=(x-\mu)/\sigma$ in the integration. The resultant p.d.f. had $\mu = 0$ and $\sigma^2 = 1$ (3-17).

Comment: This result is used for further derivations in Sections 4-3, 4-4, and 4-10, and for hypothesis testing in Chapter 5.

4.2 DISTRIBUTION OF THE SAMPLE MEAN

Given: A random sample $x_1, x_2 \ldots x_n$, where x_i are independent and $x_i \overset{d}{\rightarrow} n\ (\mu, \sigma^2)$

Required to prove:

$$\boxed{\bar{x} \overset{d}{\rightarrow} n\ (\mu, \frac{\sigma^2}{n})}$$

Proof: The moment generating function of $\bar{x}$ is

$M(t) = E\left[\exp\left[\frac{t}{n}(x_1+x_2+\ldots+x_n)\right]\right]$ and for x_i statistically independent $M(t) = E\left[e^{\frac{t}{n}x_1}\right] E\left[e^{\frac{t}{n}x_2}\right] \ldots E\left[e^{\frac{t}{n}x_n}\right]$.

For x_i distributed as $n(\mu, \sigma^2)$, the m.g.f. is (3-16)

$$M(t) = E\left[e^{tx}\right] = \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right].$$

thus in the present case for $\bar{x}$

$$M(t) = \prod_{i=1}^{n} \exp\left[\mu_i \frac{t}{n} + \frac{\sigma_i^2 \left(\frac{t}{n}\right)^2}{2}\right]$$

$$= \exp\left[\left(\frac{1}{n}\sum^{n} \mu_i\right)t + \frac{\frac{1}{n^2}\sum_{1}^{n} \sigma^2 t^2}{2}\right].$$

$$\boxed{M(t) = \exp\left[\mu t + \frac{\frac{\sigma^2}{n} t^2}{2}\right],}$$

since $\mu_1 = \mu_2 = \ldots = \mu_n = \mu$ and $\sigma_1^2 = \sigma_2^2 = \ldots \sigma_n^2 = \sigma^2$. We recognize that the form of the m.g.f. is still normal and has parameters μ and σ^2/n, thus it is proved that $\bar{x}$ is $n(\mu, \sigma^2/n)$.

Comment: We use the above result in a subsequent derivation in Section 4.3.

4.3 DISTRIBUTION OF A NORMALIZED SAMPLE MEAN

Given: A sample mean $\bar{x} \xrightarrow{d} n(\mu, \sigma^2/n)$

Required to Prove:

$$\boxed{\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} n(0, 1)}$$

Proof: From section 4.1, if $x_i \xrightarrow{d} n(\mu, \sigma^2)$ then

$$(x_i - \mu)/\sigma \xrightarrow{d} n(0, 1).$$

The only difference in this case is that the variances are scaled by $1/n$.

Comment: We use this result in further derivations in Section 4-8 and in Chapter 8.

4.4 DISTRIBUTION OF THE SQUARE OF A NORMALIZED NORMAL RANDOM VARIABLE

Given: $x \overset{d}{\rightarrow} n(\mu, \sigma^2)$

Required to Prove: $\boxed{\left(\frac{x-\mu}{\sigma}\right)^2 \overset{d}{\rightarrow} \chi^2(1)}$

Proof: From section 4.1 $w = (x-\mu)/\sigma \overset{d}{\rightarrow} n(0, 1)$, so the c.d.f. of $v \quad w^2$ is

$$\Phi(v) = \Pr(w^2 \leq v) = \Pr(-\sqrt{v} \leq w \leq \sqrt{v})$$

$$= 2 \int_0^{\sqrt{v}} \frac{1}{(2\pi)^{1/2}} e^{-w^2/2} \, dw \quad , \quad 0 \leq v$$

$$= 0 \quad , \; v < 0 \; .$$

Next we perform the change of variable $w = y^{1/2}$; the result is

$$\Phi(v) = \int_0^v \frac{1}{(2\pi)^{1/2}} y^{-1/2} e^{-y/2} \, dy \quad 0 \leq v \; .$$

The associated p.d.f. is $\phi(v) = \Phi'(v)$, namely

$$\boxed{\phi(v) = \frac{1}{(\pi)^{1/2}(2)^{1/2}} v^{(1/2-1)} e^{-v/2}, \quad 0 < v < \infty \; . \\ = 0 \text{ elsewhere}}$$

Comparing the last expression with that of the basic form of χ^2 p.d.f. (3-33), we see that the degrees of freedom $\nu = 1$, and knowing that the

gamma function $\Gamma(\frac{1}{2}) = \pi^{1/2}$, the χ^2 p.d.f. form of $[(x-\mu)/\sigma]^2$ is verified.

Comment: We will use this result for further derivations in Sections 4-6 and 4-7.

4.5 DISTRIBUTION OF THE SUM OF SEVERAL CHI-SQUARE RANDOM VARIABLES

Given: A random sample $y_1, y_2 \ldots y_n$ where y_i are independent and $y_i \overset{d}{\rightarrow} \chi^2(\nu_i)$.

Required to prove:

$$\boxed{\sum_1^n y_i \overset{d}{\rightarrow} \chi^2(\nu_1 + \nu_2 + \ldots \nu_n)}$$

Proof: The moment generating function of $\sum_1^n y_i$ is

$$M(t) = E\left[\exp\left[t(y_1 + y_2 + \ldots y_n)\right]\right]$$

$$= E\left[e^{ty_1}\right] E\left[e^{ty_2}\right] \ldots E\left[e^{ty_n}\right].$$

Since the m.g.f. of a χ^2 variable is (3-35)

$$M(t) = (1-2t)^{-\nu/2} ,$$

the m.g.f. for this case is

$$M(t) = (1 - 2t)^{-(\nu_1 + \nu_2 + \ldots \nu_n)/2} .$$

which corresponds to a chi-square random variable with $\nu_1 + \nu_2 + \ldots \nu_n$ degrees of freedom.

Comment: We use this result for a further derivation in Section 4-6.

4.6 DISTRIBUTION OF THE SUM OF SQUARES OF SEVERAL NORMALIZED NORMAL RANDOM VARIABLES

Given: A random sample $x_1, x_2 \ldots x_n$, where x_i are independent.

and $x_i \xrightarrow{d} n(\mu, \sigma^2)$

Required to Prove:

$$\boxed{\sum_{1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2 \xrightarrow{d} \chi^2(n)}$$

Proof: From section 4.4 if $y_i = [(x_i - \mu)/\sigma]^2$ then $y_i \xrightarrow{d} \chi^2(1)$.

From Section 4.5

$$\sum_{1}^{n} y_i \xrightarrow{d} \chi^2(\nu_1 + \nu_2 + \ldots \nu_n)$$

In our case $\nu_1 = \nu_2 = \ldots \nu_n = 1$, thus

$$\sum_{1}^{n} y_i = \sum_{1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2 \xrightarrow{d} \chi^2(n) \quad .$$

Comment: We use this result for a further derivation in Sections 4-7 and 4-10, as well as for hypothesis testing in Chapter 8.

4.7 DISTRIBUTION OF A FUNCTION OF THE SAMPLE VARIANCE

Given: The sample variance $s^2 = \sum_{1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$ where the

$$x_i \xrightarrow{d} n(\mu, \sigma^2)$$

Required to Prove:

$$\boxed{\frac{(n-1)\, s^2}{\sigma^2} = \sum_{1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} \xrightarrow{d} \chi^2(n-1)}$$

Proof: We begin by writing

$$\sum_{1}^{n} (x_i - \mu)^2 = \sum_{1}^{n} (x_i - \bar{x} + \bar{x} - \mu)^2$$

$$= \sum_{1}^{n} [(x_i - \bar{x})^2 + (\bar{x} - \mu)^2 + 2(x_i - \bar{x})(\bar{x} - \mu)]$$

$$= \sum_{1}^{n} (x_i - \bar{x})^2 + \sum_{1}^{n} (\bar{x} - \mu)^2 + \sum_{1}^{n} 2(x_i - \bar{x})(\bar{x} - \mu)$$

$$= \sum_{1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 + 2(\bar{x} - \mu) \sum_{1}^{n} (x_i - \bar{x}) \quad .$$

But

$$\sum_{1}^{n}(x_i-\bar{x}) = \sum_{1}^{n}(x_i-\sum_{1}^{n}\frac{x_i}{n}) = \sum_{1}^{n} x_i - \sum_{1}^{n}\sum_{1}^{n}\frac{x_i}{n} = \sum_{1}^{n}x_i - \sum_{1}^{n} x_i = 0 \quad .$$

Therefore

$$\sum_{1}^{n}(x_i-\mu)^2 = \sum_{1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2 \quad .$$

Dividing by σ^2 yields

$$\frac{\sum_{1}^{n}(x_i-\mu)^2}{\sigma^2} = \frac{\sum_{1}^{n}(x_i-\bar{x})^2}{\sigma^2} + \frac{n(\bar{x}-\mu)^2}{\sigma^2}$$

$$= \frac{(n-1)\,s^2}{\sigma^2} + \frac{n(\bar{x}-\mu)^2}{\sigma^2} \quad .$$

Writing the m.g.f. of this equation

$$M(t) = E\left\{\exp\left[t\,\frac{\sum_{1}^{n}(x_i-\mu)^2}{\sigma^2}\right]\right\} = E\left\{\exp\left[t\left(\frac{(n-1)s^2}{\sigma^2} + \frac{n(\bar{x}-\mu)^2}{\sigma^2}\right)\right]\right\}$$

and we can write*

$$M(t) = E\left\{\exp\left[t\,\frac{(n-1)s^2}{\sigma^2}\right]\right\} E\left\{\exp\left[t\,\frac{n(\bar{x}-\mu)^2}{\sigma^2}\right]\right\} .$$

From section 4.6

$$\sum_{1}^{n}\frac{(x_i-\mu)^2}{\sigma^2} \xrightarrow{d} \chi^2\,(n)$$

and section 4.4

$$\frac{n(\bar{x}-\mu)^2}{\sigma^2} \xrightarrow{d} \chi^2\,(1)$$

so that these have the corresponding m.g.f.'s

$$(1-2t)^{-n/2} \quad \text{and} \quad (1-2t)^{-1/2} \quad .$$

---

* Since $\frac{n-1}{\sigma^2}\,s^2$ and $n\left(\frac{\bar{x}-\mu}{\sigma}\right)^2$ are statistically independent [see Hogg and Craig, 1965, p. 233].

Therefore

$$(1-2t)^{-n/2} = (1-2t)^{-1/2} \; E\left\{\exp\left[t\,\frac{(n-1)s^2}{\sigma^2}\right]\right\} .$$

Thus the m.g.f. of $(n-1)s^2/\sigma^2$ is

$$\boxed{E\left[e^{t(n-1)s^2/\sigma^2}\right] = (1-2t)^{-(n-1)/2}}$$

This m.g.f. corresponds to a chi-square random variable with n-1 degrees of freedom. Therefore

$$\frac{(n-1)s^2}{\sigma^2} \xrightarrow{d} \chi^2\,(n-1) .$$

Comment: This result is used for a subsequent derivation in Section 4-9 and for hypothesis testing in Chapter 8.

4.8 DISTRIBUTION OF THE RATIO OF THE NORMALIZED SAMPLE MEAN TO (s/σ)

Given: a) $\bar{x} \xrightarrow{d} n(\mu, \frac{\sigma^2}{n})$,

b) $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} n\,(0, 1)$,

c) $\dfrac{(n-1)\,s^2}{\sigma^2} \xrightarrow{d} \chi^2\,(n-1)$

Required to Prove:

$$\boxed{\frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\left[\dfrac{(n-1)\,s^2}{\sigma^2}\Big/(n-1)\right]^{1/2}} = \frac{(\bar{X}-\mu)n^{1/2}}{S} \xrightarrow{d} t(n-1)}$$

Proof: The result for the above follows immediately from the definition of a t random variable. Recall from Section 3.3 that

$$\frac{n(0,\,1)}{[\chi^2\,(\nu\,)/\nu]^{1/2}} \xrightarrow{d} t\,(\nu) \quad ,$$

and in the present case we have

$$\frac{n\,(0,\,1)}{[\chi^2\,(n-1)/(n-1)]^{1/2}} \xrightarrow{d} t\,(n-1) \; .$$

Comment: This result is used in hypothesis testing in Chapter 5.

4.9 DISTRIBUTION OF THE RATIO OF TWO SAMPLE VARIANCES FROM THE SAME POPULATION

Given: a) $\dfrac{(n_1-1)\,s_1^2}{\sigma^2} \xrightarrow{d} \chi^2 \quad (n_1-1)$

b) $\dfrac{(n_2-1)s_2^2}{\sigma^2} \xrightarrow{d} \chi^2 \quad (n_2-1)$

Required to Prove:

$$\frac{\dfrac{(n_1-1)\,s_1^2/\sigma^2}{(n_1-1)}}{\dfrac{(n_2-1)\,s_2^2/\sigma^2}{(n_2-1)}} = \frac{S_1^2}{S_2^2} = \frac{\sum_1^{n_1}(X_i-\bar{X}_1)^2}{\sum_1^{n_2}(X_i-\bar{X}_2)^2}\,\frac{(n_2-1)}{(n_1-1)} \xrightarrow{d} F(n_1-1,\,n_2-1)$$

Proof: The above result follows immediately from the definition of an F random variable. Recall from Section 3.4 that

$$\frac{\chi^2(\nu_1)/\nu_1}{\chi^2(\nu_2)/\nu_2} \xrightarrow{d} F(\nu_1,\,\nu_2)$$

and in the present case we have

$$\frac{\chi^2(n_1-1)/(n_1-1)}{\chi^2(n_2-1)/(n_2-1)} \xrightarrow{d} F(n_1-1,\,n_2-1) \quad .$$

Comment: This result is used in Chapter 5.

4.10 DISTRIBUTION OF A MULTIVARIATE NORMAL QUADRATIC FORM

Given: The quadratic form $\underset{1xm}{X^T}\ \underset{mxm}{\Sigma_X^{-1}}\ \underset{mx1}{X}$ (equation 3-27), where X is a vector of m normally distributed random variables with zero means and variance - covariance matrix $\underset{mxm}{\Sigma_X}$.

Required to Prove:

$$\underset{1xm}{X^T}\ \underset{mxm}{\Sigma_X^{-1}}\ \underset{mx1}{X} \xrightarrow{d} \chi^2(m)$$

Proof: First make the orthogonal transformation of X to Y by

$$\underset{mx1}{Y} = \underset{mxm}{T^{-1}}\ \underset{mx1}{X} = T^T X; \quad X = TY$$

such that in the process Σ_X^{-1} is diagonalized and the variables $\underset{mx1}{Y}$ are made independent. The quadratic form becomes

$$\underset{1xm}{X^T}\ \underset{mxm}{\Sigma_X^{-1}}\ \underset{mx1}{X} = \underset{1xm}{Y^T}\ (\ \underset{mxm}{T^T}\ \underset{mxm}{\Sigma_X^{-1}}\ \underset{mxm}{T}\ )\ \underset{mx1}{Y}$$

$$= \frac{y_1^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2} + \cdots + \frac{y_m^2}{\sigma_m^2}.$$

where each $\frac{y_i}{\sigma_i} \xrightarrow{d} n(0, 1)$ according to Section 4.1, and that $(\frac{y_i}{\sigma_i})^2 \xrightarrow{d} \chi^2(1)$ according to Section 4.4, and finally according to Section 4.6, it follows that the sum of m random variables (each distributed $\chi^2(1)$) is distributed as $\chi^2(m)$.

Comments: This result for quadratic forms is the basis for statistical testing in multivariate least squares estimation problems discussed in Chapter 8.

4.11 SUMMARY OF DISTRIBUTIONS OF FUNCTIONS OF RANDOM VARIABLES

This is a summary of all distributions introduced thus far. In Chapter 3 we derived the normal, chi square, Student's (t), and F

distributions; these are special distributions and are the basis for the distributions of functions of random variables given in this Chapter. The latter are summarized in Table 4-1 in such a way as to show their uses in:

1) subsequent derivations,

2) hypothesis testing.

Table 4-1. Interplay of Random Variables - Their Uses

Section Derived	Random Variable		Used in Derivation	Used For Statistical Test(Chap.5)
4.1	$x_i \xrightarrow{d} n(\mu, \sigma^2)$	$\frac{x_i-\mu}{\sigma} \xrightarrow{d} n(0, 1)$		5.2 5.3
4.2	$x_i \xrightarrow{d} n(\mu, \sigma^2)$	$\bar{x} = \frac{\sum_1^n x_i}{n} \xrightarrow{d} n(\mu, \sigma^2/n)$		no
4.3	$\bar{x} \xrightarrow{d} n(\mu, \sigma^2/n)$	$\frac{\bar{x}-\mu}{\sigma/(n)^{1/2}} \xrightarrow{d} n(0, 1)$		5.4 5.5
4.4	$x_i \xrightarrow{d} n(\mu, \sigma^2)$	$(\frac{x_i-\mu}{\sigma})^2 \xrightarrow{d} \chi^2(1)$		no
4.5	$y_i \xrightarrow{d} \chi^2(\nu_i)$	$\sum_1^n y_i \xrightarrow{d} \chi^2(\nu_1+\nu_2+..\nu_n)$		no
4.6	$x_i \xrightarrow{d} n(\mu, \sigma^2)$	$\sum_1^n (\frac{x_i-\mu}{\sigma})^2 \xrightarrow{d} \chi^2(n)$		5.8
4.7	$x_i \xrightarrow{d} n(\mu, \sigma^2)$	$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_1^n (x_i-\bar{x})^2}{\sigma^2} \xrightarrow{d} \chi^2(n-1)$		5.9 5.10
4.8	$\bar{x} \xrightarrow{d} n(\mu, \sigma^2/n)$	$\frac{\bar{x}-\mu}{s/\sqrt{n}} \xrightarrow{d} t(n-1)$		5.6 5.7
4.9	$x_i \xrightarrow{d} n(\mu, \sigma^2)$	$\frac{s_1^2}{s_2^2} \xrightarrow{d} F(n_1-1, n_2-1)$		yes
4.10	multivariate normal	$X^T \Sigma^{-1} X \xrightarrow{d} \chi^2(m)$ lxm mxm mxl		Chap. 8

Basis for Multivariate Testing

5. UNIVARIATE INTERVAL ESTIMATION AND HYPOTHESIS TESTING

5.1 INTRODUCTION

Recall from Section 2.5 that point estimation deals with the estimation of population parameters from a knowledge of sample statistics. For example the sample mean $\bar{x}$ and sample variance s^2 are unbiased point estimators for the population mean μ and population variance σ^2, that is,

$$\bar{x} = \hat{\mu} = \frac{\sum_{1}^{n} x_i}{n} ,$$

$$s^2 = \hat{\sigma}^2 = \frac{\sum_{1}^{n} (x_i - \bar{x})^2}{n - 1} .$$

In this chapter we treat interval estimation which is the determination of the region or limits associated with point estimates.

Recall from section 2.6 that interval estimation involves a probability statement

$$P_r (e_1 \leq \varepsilon \leq e_2) = \alpha$$

where ε is a statistic of a known p.d.f., α is a probability value between 0 and 1 which must be specified (often $\alpha = 0.95$), and e_1 and e_2 are abscissa values of the known p.d.f. which are determined by the specified α. Note, finding e_1 and e_2 given the p.d.f. and α is the inverse problem described in sections 3.1.5, 3.2.4, 3.3.3 and 3.4.3 for the normal, chi-square, t and F distributions respectively. The interval

$$[e_1 \leq \varepsilon \leq e_2]$$

is called the confidence interval. If $\alpha = 0.95$ for example, it is called the 95% confidence interval. In general ε is not a single statistic, but is a function of several statistics, the values of all but one of which are computed or specified. Therefore, the confidence interval for the unknown statistic (say μ) in ε is found by operations on the inequalities in the above confidence interval to yield

$$[f(e_1,\varepsilon) \leq \mu \leq f(e_2,\varepsilon)] .$$

It often happens that an a priori hypothesis about the value of the unknown statistic μ can be made. The hypothesis

$$H_o : \mu = \mu_H$$

is called the null hypothesis H_o and is read "the mean μ is hypothesized to have the particular value μ_H." The alternative hypothesis is

$$H_1 : \mu \neq \mu_H .$$

The confidence interval can be used to determine whether the null hypothesis should be rejected. If the hypothesis is rejected, then the probability value α in the probability statement is called the significance level of the test. Failure to reject the hypothesis does not mean that the hypothesis is true. No statement about the hypothesis, the hypothesis

test, or the significance level can be made if the hypothesis is not rejected.

In this chapter, we will restrict our discussions to population distribution functions which involve only one random variable-univariate case (for example the set of measurements of the distance between two points). This concept is extended in Chapter 8 to include population distribution functions involving several random variables, and is called multivariate interval estimation.

Interval estimates are the basis for hypothesis testing and are developed in this chapter for the following quantities:

1. a single observation, x_i
2. the population mean, μ
3. the sample mean, $\bar{x}$
4. the population variance, σ^2
5. the sample variance, s^2
6. the ratio of two population variances, (σ_2^2/σ_1^2)
7. the ratio of two sample variances, (s_2^2/s_1^2).

We make extensive use of the distributions of random variables given in Chapter 4 in the remaining portion of this chapter.

5.2 EXAMINATION OF A SINGLE MEASUREMENT X_i IN TERMS OF THE MEAN μ AND VARIANCE σ^2

Consider a single measurement x_i as being a sample drawn from a normal population with known mean μ and known variance σ^2, that is

$$x_i \overset{d}{\rightarrow} n\ (\mu,\sigma^2)\ .$$

From section 4.1, we know that

$$\frac{x_i - \mu}{\sigma} \xrightarrow{d} n(0,1) .$$

The associated probability statement is

$$\Pr\left(-c \leq \frac{x_i - \mu}{\sigma} \leq c\right) = \alpha$$

while the confidence interval for x_i is

$$\boxed{[\mu - c\,\sigma \leq x_i \leq \mu + c\,\sigma]}$$

The bounds of this confidence interval are evaluated from

(1) the known value for the mean μ

(2) the known value for the variance σ^2

(3) the tabulated value of c (Appendix B-1) corresponding to $n(0,1)$ and α.

This confidence interval is used to test the hypothesis

$$H_o: x_i = x_H .$$

5.3 EXAMINATION OF THE MEAN μ IN TERMS OF AN OBSERVATION x_i AND VARIANCE σ^2

The mean μ can be examined in terms of the variance σ^2 and an observation x_i as follows. Consider a random sample where

$$x_i \xrightarrow{d} n(\mu,\sigma^2) .$$

From section 4.1, we know that

$$\frac{x_i - \mu}{\sigma} \xrightarrow{d} n(0,1) .$$

The associated probability statement is

$$\Pr\left(-c \leq \frac{x_i - \mu}{\sigma} \leq c\right) = \alpha ,$$

while the confidence interval for μ is

$$\boxed{[x_i - c\,\sigma \leq \mu \leq x_i + c\,\sigma] \ .}$$

The bounds of the interval are evaluated from:

(1) the measurement value x_i

(2) the known value for the variance σ^2

(3) the tabulated (Table B-1 Appendix) value of c corresponding to $n(0,1)$ and α.

The above is used to test the hypothesis

$$H_o : \mu = \mu_H \ .$$

5.4 EXAMINATION OF THE MEAN μ IN TERMS OF THE SAMPLE MEAN $\bar{X}$ AND VARIANCE σ^2/n

The mean μ can be examined in terms of the known variance σ^2/n and sample mean $\bar{x}$. From section 4.3, we know that

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} n(0,1) \ .$$

The associated probability statement is

$$\Pr\left(-c \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq c\right) = \alpha \ ,$$

while the confidence interval for μ is

$$\left[\bar{x} - c\,\frac{\sigma}{(n)^{1/2}} \leq \mu \leq \bar{x} + c\,\frac{\sigma}{(n)^{1/2}}\right]$$

The bounds of the interval are evaluated from:

(1) the computed value for the sample mean $\bar{x}$,

(2) the known value for the variance σ^2/n ,

(3) the tabulated value for c (Table B-1 Appendix corresponding to n(0,1) and α.

This confidence interval is used to test the hypothesis

$$H_o : \mu = \mu_H .$$

5.5 EXAMINATION OF THE SAMPLE MEAN $\bar{X}$ IN TERMS OF THE MEAN μ AND VARIANCE σ^2/n

The sample mean $\bar{x}$ can be examined in terms of mean μ and variance σ^2/n as follows. Again we begin by considering that from section 4.3

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} n\,(0,1) .$$

The associated probability statement is

$$\Pr\left(-c \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq c\right) = \alpha ,$$

while the confidence interval for $\bar{x}$ is

$$\left[\mu - c\,\frac{\sigma}{(n)^{1/2}} \leq \bar{x} \leq \mu + c\,\frac{\sigma}{(n)^{1/2}}\right] .$$

The bounds of the interval are evaluated from:

(1) the known mean μ,

(2) the known variance σ^2/n,

(3) the tabulated value for c (Appendix B-1) corresponding to n(0,1) and α.

The above is used to test the hypothesis

$$H_o: \bar{x} = \bar{x}_H \ .$$

5.6 EXAMINATION OF THE MEAN μ IN TERMS OF THE SAMPLE MEAN $\bar{X}$ AND SAMPLE VARIANCE s^2

The mean μ can be examined in terms of the sample mean $\bar{x}$ and sample variance s^2 as follows. From section 4.8

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \xrightarrow{d} t(n-1) \ .$$

The associate probability statement is

$$\Pr\left(-t_p \leq \frac{\bar{x} - \mu}{s/\sqrt{n}} \leq t_p\right) = \alpha \ ,$$

while the confidence interval for μ is

$$\boxed{\left[\bar{x} - t_p \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t_p \frac{S}{\sqrt{n}}\right] .}$$

The bounds of the interval are evaluated from:

(1) the computed value for the sample mean $\bar{x}$,

(2) the computed value for the sample variance s^2,

(3) the sample size n,

(4) the tabulated value of t_p (Appendix B-3) corresponding to t(n-1) and α.

The above is used to test the hypothesis

$$H_o: \mu = \mu_H .$$

5.7 EXAMINATION OF THE SAMPLE MEAN $\bar{X}$ IN TERMS OF THE MEAN μ AND SAMPLE VARIANCE s^2

The sample mean $\bar{x}$ can be examined in terms of the mean μ and sample variance s^2 as follows. Again . . . from section 4.8

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \xrightarrow{d} t(n-1) .$$

The associated probability statement is

$$\Pr\left(-t_p \leq \frac{\bar{x} - \mu}{s/\sqrt{n}} \leq t_p\right) = \alpha ,$$

while the confidence interval for $\bar{x}$ is

$$\boxed{\left[\mu - t_p \frac{s}{(n)^{1/2}} \leq \bar{x} \leq \mu + t_p \frac{s}{(n)^{1/2}}\right] .}$$

The bounds of the interval are evaluated from:

(1) the known mean μ,

(2) the computed value for the sample variance s^2, and sample size n,

(3) the tabulated value t_p (Appendix B-3) corresponding to t(n-1) and α.

The above is used to test the hypothesis

$$H_o: \bar{x} = \bar{x}_H .$$

5.8 EXAMINATION OF THE VARIANCE σ^2 IN TERMS OF THE MEAN μ AND SEVERAL MEASUREMENTS $x_1, x_2, \ldots, x_n$

The variance σ^2 can be examined in terms of the mean μ and several measurements $x_1, x_2, \ldots, x_n$ as follows. From section 4.6, we know that

$$\sum_{1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 \xrightarrow{d} \chi^2(n) .$$

The associated probability statement is

$$\Pr\left(\chi^2_{p_1} \leq \sum_{1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 \leq \chi^2_{p_2} \right) = \alpha ,$$

while the confidence interval for σ^2 is

$$\left[\frac{\sum_{1}^{n} (x_i - \mu)^2}{\chi^2_{p_2}} \leq \sigma^2 \leq \frac{\sum_{1}^{n} (x_i - \mu)^2}{\chi^2_{p_1}} \right] .$$

The bounds of the interval are evaluated from:

(1) the known mean μ,

(2) the several measurements $x_1, x_2, \ldots, x_n$,

(3) the tabulated values $\chi^2_{p_1}$ and $\chi^2_{p_2}$ (Appendix B-2) corresponding to $\chi^2(n)$, $p_1 = \frac{1 - \alpha}{2}$, and $p_2 = \frac{1+\alpha}{2}$

The above is used to test the hypothesis

$$H_o: \ \sigma^2 = \sigma^2_H .$$

5.9 EXAMINATION OF THE VARIANCE σ^2 IN TERMS OF THE SAMPLE VARIANCE s^2

The variance σ^2 can be examined in terms of the sample variance s^2 as follows. From section 4.7, we know that

$$\frac{(n-1)\,s^2}{\sigma^2} \xrightarrow{d} \chi^2\,(n-1) \ .$$

The associated probability statement is

$$\Pr\left(\chi^2_{p_1} \leq \frac{(n-1)\,s^2}{\sigma^2} \leq \chi^2_{p_2}\right) = \alpha \ ,$$

while the confidence interval for σ^2 is

$$\left[\frac{(n-1)\,s^2}{\chi^2_{p_2}} \leq \sigma^2 \leq \frac{(n-1)\,s^2}{\chi^2_{p_1}}\right] .$$

The bounds of the interval are evaluated from:

(1) the computed value of the sample variance s^2,

(2) the sample size n,

(3) the tabulated values $\chi^2_{p_1}$ and $\chi^2_{p_2}$ (Appendix B-2) corresponding to $\chi^2(n-1)$, $p_1 = \frac{1-\alpha}{2}$ and $p_2 = \frac{1+\alpha}{2}$.

The above is used to test the hypothesis

$$H_o: \ \sigma^2 = \sigma^2_H \ .$$

5.10 EXAMINATION OF THE SAMPLE VARIANCE s^2 IN TERMS OF THE VARIANCE σ^2

The sample variance s^2 can be examined in terms of the variance σ^2 as follows. Again from section 4.7, we know that

$$\frac{(n-1)\,s^2}{\sigma^2} \xrightarrow{d} \chi^2(n-1) .$$

The associated probability statement is

$$\Pr\left(\chi^2_{p_1} \leq \frac{(n-1)\,s^2}{\sigma^2} \leq \chi^2_{p_2}\right) = \alpha ,$$

while the confidence interval for s^2 is

$$\left[\chi^2_{p_1}\frac{\sigma^2}{(n-1)} \leq s^2 \leq \chi^2_{p_2}\frac{\sigma^2}{(n-1)}\right] .$$

The bounds of the interval are evaluated from:

(1) the known value of the variance σ^2,

(2) the sample size n,

(3) the tabulated values $\chi^2_{p_1}$ and $\chi^2_{p_2}$ (Appendix B-2) corresponding to $\chi^2(n-1)$, $p_1 = \frac{1-\alpha}{2}$ and $p_2 = \frac{1+\alpha}{2}$.

The above is used to test the hypothesis

$$H_o: \; s^2 = s^2_H .$$

5.11 EXAMINATION OF THE RATIO OF TWO VARIANCES (σ_2^2/σ_1^2) IN TERMS OF THE SAMPLE VARIANCES s_1^2 AND s_2^2

The ratio of two variances σ_2^2/σ_1^2 can be examined in terms of the sample variances s_1^2 and s_2^2 as follows. Using section 4.9, we can write that

$$\frac{\left(\frac{(n_1 - 1)\, s_1^2}{\sigma_1^2}\right) / (n_1 - 1)}{\left(\frac{(n_2 - 1)\, s_2^2}{\sigma_2^2}\right) / (n_2 - 1)} \xrightarrow{d} F\,(n_1 - 1,\ n_2 - 1)\ ,$$

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \xrightarrow{d} F\,(n_1 - 1,\ n_2 - 1)\ .$$

The associated probability statement is

$$\Pr\left(F_{p_1} \leq \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \leq F_{p_2}\right) = \alpha\ ,$$

while the confidence interval for σ_2^2 / σ_1^2 is

$$\left[\ F_{p_1} \frac{s_2^2}{s_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq F_{p_2} \frac{s_2^2}{s_1^2}\ \right]\ .$$

The bounds of the interval are evaluated from:

(1) the computed values of the sample variances s_1^2 and s_2^2 ,

(2) the tabulated values of F_{p_1} and F_{p_2} (Appendix B-4) corresponding to $F(n_1 - 1,\ n_2 - 1)$, $p_1 = \frac{1 - \alpha}{2}$, and $p_2 = \frac{1 + \alpha}{2}$.

The above is used to test the hypothesis

$$H_o: (\sigma_2^2/\sigma_1^2) = (\sigma_2^2/\sigma_1^2)_H \ .$$

Note that if in the above $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then $\mathbf{s}_1^2$ and $\mathbf{s}_2^2$ are simply sample variances of the same population ($n\ (\mu,\sigma^2)$), and the ratio $\sigma_2^2 \ / \ \sigma_1^2 = 1$.

5.12 EXAMINATION OF THE RATIO OF TWO SAMPLE VARIANCES $(\mathbf{s}_1^2/\mathbf{s}_2^2)$ IN TERMS OF THE VARIANCES σ_1^2 AND σ_2^2 .

The ratio of two sample variances $\mathbf{s}_1^2/\mathbf{s}_2^2$ can be examined in terms of the variances σ_1^2 and σ_2^2 as follows. Again using section 4.9, we can write that

$$\frac{\mathbf{s}_1^2/\sigma_1^2}{\mathbf{s}_2^2/\sigma_2^2} \ \xrightarrow{d} \ F\ (n_1 - 1,\ n_2 - 1) \ .$$

The associated probability statement is

$$\Pr\ (F_{p_1} \leq \frac{\mathbf{s}_1^2/\sigma_1^2}{\mathbf{s}_2^2/\sigma_2^2} \leq F_{p_2}) = \alpha \ ,$$

while the confidence interval for $\mathbf{s}_1^2/\mathbf{s}_2^2$ is

$$\boxed{[\ F_{p_1}\ \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{\mathbf{s}_1^2}{\mathbf{s}_2^2} \leq F_{p_2}\ \frac{\sigma_1^2}{\sigma_2^2}\ ] \ .}$$

The bounds of the interval are evaluated from:

(1) the values of the variances σ_1^2 and σ_2^2 ,

(2) the tabulated values of F_{p_1} and F_{p_2} (Appendix B-4) corresponding to $F(n_1 - 1, n_2 - 1)$, $p_1 = \frac{1-\alpha}{2}$ and $p_2 = \frac{1+\alpha}{2}$.

The above is used to test the hypothesis

$$H_o: \ s_1^2/s_2^2 = (s_1^2/s_2^2)_H \ .$$

Note that when $\sigma_1^2 = \sigma_2^2$, the bounds of the confidence interval become described by the F percentilies.

5.13 EXAMINATION OF THE RATIO OF TWO VARIANCES (σ_2^2/σ_1^2) IN TERMS OF SEVERAL MEASUREMENTS FROM TWO SAMPLES

The ratio of two variances σ_2^2/σ_1^2 can be examined in terms of measurements $x_1, x_2, \ldots, x_{n_1}$ sampled from a population which is $n(\mu_1, \sigma_1^2)$, and measurements $x_1, x_2, \ldots, x_{n_2}$ sampled from a population which is $n(\mu_2, \sigma_2^2)$. Using section 4.6, we know that

$$\frac{\sum_1^{n_1} (x_i - \mu_1)^2}{\sigma_1^2} \xrightarrow{d} \chi^2(n_1) \ ,$$

and

$$\frac{\sum_1^{n_2} (x_i - \mu_2)^2}{\sigma_2^2} \xrightarrow{d} \chi^2(n_2) \ .$$

Further, using section 3.4, we know that

$$\frac{\dfrac{\sum_{1}^{n_1}(x_i - \mu_1)^2}{\sigma_1^2 n_1}}{\dfrac{\sum_{1}^{n_2}(x_i - \mu_2)^2}{\sigma_2^2 n_2}} = \frac{(1)}{(2)} \xrightarrow{d} F(n_1, n_2) \quad .$$

The associated probability statement is

$$\Pr\left(F_{p_1} \leq \frac{(1)}{(2)} \leq F_{p_2}\right) = \alpha$$

while the confidence interval for σ_2^2/σ_1^2 is

$$\left[F_{p_1} \frac{n_1}{n_2} \frac{\sum_{1}^{n_2}(x_i - \mu_2)^2}{\sum_{1}^{n_1}(x_i - \mu_1)^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq F_{p_2} \frac{n_1}{n_2} \frac{\sum_{1}^{n_2}(x_i - \mu_2)^2}{\sum_{1}^{n_1}(x_i - \mu_1)^2} \right] .$$

The bounds of the confidence interval are evaluated from:

(1) the sample sizes n_1 and n_2,

(2) the means μ_1 and μ_2 of the two populations,

(3) the sum of squares of the differences of the measurements from each of the respective means μ_1 and μ_2,

(4) the tabulated values of F_{p_1} and F_{p_2} (Appendix B-4) corresponding to $F(n_1, n_2)$, $p_1 = \frac{1 - \alpha}{2}$ and $p_2 = \frac{1+\alpha}{2}$.

The above is used to test the hypothesis

$$H_o: \quad (\sigma_2^2/\sigma_1^2) = (\sigma_2^2/\sigma_1^2)_H \quad .$$

5:14 SUMMARY OF UNIVARIATE CONFIDENCE INTERVALS

In table 5-1 we summarize univariate confidence intervals and hypothesis testing according to

a) column one - the chapter section discussing the particular confidence interval,

b) column two - the quantity being examined by the confidence interval,

c) column three - the quantities which must be known for the confidence interval,

d) column four - the statistic,

e) column five - the confidence interval.

TABLE 5-1. UNIVARIATE CONFIDENCE INTERVALS

Section	Quantity Examined	Known Quantities	Statistic	Confidence Interval
5.2	single observation x_i	μ, σ	$\frac{x_i-\mu}{\sigma} \xrightarrow{d} n(0,1)$	$[\mu-c\sigma \leq x_i \leq \mu+c\sigma]$
5.3	mean μ	x_i, σ	as above	$[x_i - c\sigma \leq \mu \leq x_i + c\sigma]$
5.4	mean μ	n, $\bar{x}$, σ	$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} n(0,1)$	$[\bar{x} - c\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + c\frac{\sigma}{\sqrt{n}}]$
5.5	sample mean $\bar{x}$	n, μ, σ	as above	$[\mu - c\frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + c\frac{\sigma}{\sqrt{n}}]$
5.6	mean μ	n, $\bar{x}$, s	$\frac{\bar{x} - \mu}{s/\sqrt{n}} \xrightarrow{d} t(n-1)$	$[\bar{x} - t_p\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_p\frac{s}{\sqrt{n}}]$
5.7	sample mean $\bar{x}$	n, μ, s	as above	$[\mu - t_p\frac{s}{\sqrt{n}} \leq \bar{x} \leq \mu + t_p\frac{s}{\sqrt{n}}]$
5.8	variance σ^2	n, x_i, μ	$\sum_1^n(\frac{x_i - \mu}{\sigma})^2 \xrightarrow{d} \chi^2(n)$	$[\frac{\sum_1^n(x_i - \mu)^2}{\chi^2_{p_2}} \leq \sigma^2 \leq \frac{\sum_1^n(x_i - \mu)^2}{\chi^2_{p_1}}]$

TABLE 5-1 (continued)

Section	Quantity Examined	Known Quantities	Statistic	Confidence Interval
5.9	variance σ^2	n, s	$\frac{(n-1)s^2}{\sigma^2} \xrightarrow{d} \chi^2(n-1)$	$[\frac{(n-1)s^2}{\chi^2_{p_2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{p_1}}]$
5.10	sample variance s^2	n, σ	as above	$[\chi^2_{p_1} \frac{\sigma^2}{(n-1)} \leq s^2 \leq \chi^2_{p_2} \frac{\sigma^2}{(n-1)}]$
5.11	ratio of two variances σ_2^2/σ_1^2	n_1, n_2, s_2/s_1	$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \xrightarrow{d} F(n_1-1, n_2-1)$	$[F_{p_1} (s_2/s_1)^2 \leq (\sigma_2/\sigma_1)^2 \leq F_{p_2} (s_2/s_1)^2]$
5.12	ratio of two sample variances s_1^2/s_2^2	n_1, n_2, σ_1/σ_2	as above	$[F_{p_1} (\sigma_1/\sigma_2)^2 \leq (s_1/s_2)^2 \leq F_{p_2} (\sigma_1/\sigma_2)^2]$
5.13	ratio of two variances σ_2^2/σ_1^2	n_1, n_2, x_i, μ_1, μ_2	$\frac{\frac{\sum_1^{n_1} (x_i-\mu_1)^2}{\sigma_1^2 n_1}}{\frac{\sum_1^{n_2} (x_i-\mu_2)^2}{\sigma_2^2 n_2}} \xrightarrow{d} F(n_1, n_2)$	$[F_{p_1} \frac{\frac{\sum_1^{n_2} (x_i-\mu_2)^2}{n_1}}{\frac{\sum_1^{n_1} (x_i-\mu_1)^2}{n_2}} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq F_{p_2} \frac{\frac{\sum_1^{n_2} (x_i-\mu_2)^2}{n_1}}{\frac{\sum_1^{n_1} (x_i-\mu_1)^2}{n_2}}]$

6. LEAST SQUARES POINT ESTIMATORS: LINEAR MATHEMATICAL MODELS

In this Chapter we follow the approaches of Schwarz [1967] and Hamilton [1964] in discussing the linear mathematical model

$$A X - L = V \quad , \tag{6.1}$$

where ${}_{n}L_{1}$ is called the observation vector and is the column vector whose elements are the observed values, ${}_{n}V_{1}$ is called the residual vector and is the column vector whose elements are the unknown measurement errors (inconsistencies), ${}_{u}X_{1}$ is called the solution vector for which we want a point estimate and whose elements are the unknown parameters, and ${}_{n}A_{u}$ is known and is called the design matrix. Note that there are n observations and u unknowns. The least squares estimation process is applied only when there are redundant measurements, that is $n > u$. The number $(n - u)$ is called the redundancy or number of degrees of freedom.

In addition to these quantities we also have weights associated with each of the observations L. The weight matrix P is the matrix whose elements are these weights.

6.1 THE LEAST SQUARES UNBIASED ESTIMATOR FOR X

The least squares criterion states that the "best" estimator $\hat{X}$ of X is the estimator which will minimize the sum of the squares of the weighted residuals, that is

$$\hat{V}^T P \hat{V} = \text{minimum} \quad , \tag{6.2}$$

where from Equation 6.1 we have

$$\hat{V} = A \hat{X} - L \quad . \tag{6.3}$$

Combining Equations 6.2 and 6.3 we have the criterion

$$\phi = (A \hat{X} - L)^T P(A \hat{X} - L) = \text{minimum} \quad . \tag{6.4}$$

To minimize this function we set

$$\frac{\partial \phi}{\partial \hat{X}} = \frac{\partial \phi}{\partial (A\hat{X}-L)} \frac{\partial (A\hat{X}-L)}{\partial \hat{X}}$$

$$= 2(A \hat{X} - L)^T P A = 0 \ , \ ,$$

or after transposing and separating

$$A^T P A \hat{X} - A^T P L = 0 \quad . \tag{6.5}$$

Equations 6.5 are called the <u>normal equations</u>. If $(A^T P A)$, called the matrix of the normal equations, is nonsingular then there is a unique least squares estimator for X which is

$$\hat{X} = (A^T P A)^{-1} A^T P L \quad . \tag{6.6}$$

An estimator $\hat{X}$ of X is called unbiased if

$$E[\hat{X}] = X \quad . \tag{6.7}$$

The estimator $\hat{X}$ of Equation 6.6 will be unbiased if the expected value of the residuals is zero, that is

$$E[V] = 0 \quad , \tag{6.8}$$

in which case, from Equation 6.1

$$E[V] = E[AX-L] = E[AX] - E[L] = AX - E[L] = 0$$

or

$$E[L] = AX \ , \tag{6.9}$$

where we have used the relation $E[X] = X$, which is trivial since X is the "true" value.

From Equation 6.6

$$E[\hat{X}] = E[(A^TPA)^{-1}A^TPL] = (A^TPA)^{-1}A^TP\ E[L]$$

$$= (A^TPA)^{-1}A^TP\ AX = X$$

thus by definition $\hat{X}$ is an unbiased estimator of X.*

6.2 CHOICE OF THE WEIGHT MATRIX P

So far we have not specified how the weight matrix P should be chosen. If $E[V] = 0$ then the covariance matrix of the "true" values of the residuals is

$$\Sigma_V = E[(V-E[V])(V-E[V])^T] = E[VV^T] \quad . \tag{6.10}$$

Also if $E[V] = 0$, from Equation 6.9

$$E[L] = A\ X$$

and from Equation 6.1

$$L - E[L] = -V$$

and the covariance matrix of the observations

---

* Note that we have made no statements about the weight matrix P as yet. Hence $\hat{X}$ is unbiased, independent of the choice of P.

$$\Sigma_L = E\,[(L-E[L])(L-E\,[L])^T] = E\,[VV^T] = \Sigma_V \quad , \tag{6.11}$$

that is the observations and the "true" values of the observation errors have the same covariance matrix, as should be expected. This does not mean that $\Sigma_L = \Sigma_{\hat{V}}$ where $\hat{V}$ is the estimator of V which results from the least squares process.

We set

$$\Sigma_L = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots \\ \sigma_{21} & \sigma_2^2 & \\ \vdots & & \ddots \\ & & \sigma_n^2 \end{bmatrix},$$

where σ_i^2 is <u>variance</u>,

and σ_i is the <u>standard deviation</u> associated with the observation ℓ_i,

and σ_{ij} is the <u>covariance</u> between observations ℓ_i and ℓ_j.

The variance of an observation is larger when it is less accurately determined. In combining many observations more importance should be attached to those having smaller variances. One reasonable choice for the weight matrix P therefore would be to set

$$P = \Sigma_L^{-1} \tag{6.12}$$

In this case values must be assigned to the variances and covariances in Σ_L before a least squares estimation can be made. The values arise from a knowledge of the measuring instruments and procedures being used. It is often possible only to assign <u>relative</u> values among the variances and covariances, so that we know Σ_L only to within a scale factor, that

is if

$$\Sigma_L = \sigma_o^2 \, Q \tag{6.13}$$

we know the relative covariance matrix Q but not the variance factor σ_o^2. In such a case if we use

$$P = Q^{-1} = \sigma_o^2 \, \Sigma_L^{-1} \tag{6.14}$$

in Equation 6.6, then we have

$$\hat{X} = (A^T \sigma_o^2 \Sigma_L^{-1} A)^{-1} A^T \sigma_o^2 \Sigma_L^{-1} L = (A^T \Sigma_L^{-1} A)^{-1} A^T \Sigma_L^{-1} L \tag{6.15}$$

that is, the variance factor drops out and either weights 6.12 or 6.14 result in the same estimator $\hat{X}$.

6.3 THE MINIMUM VARIANCE POINT ESTIMATOR FOR X

If for an estimator $\hat{X}$ of X, a matrix B exists such that

$$\hat{X} = B \, L \tag{6.16}$$

that is, the elements of $\hat{X}$ are linear functions of the observations, then $\hat{X}$ is called a linear estimator of X.

The minimum variance estimator $\hat{X}$ of X is the linear unbiased estimator whose covariance matrix

$$\Sigma_{\hat{X}} = E\,[(\hat{X} - E\,[\hat{X}])(\hat{X} - E\,[\hat{X}])^T] \tag{6.17}$$

is "less than" that of any other linear unbiased estimator of X. We need some criterion by which we can decide when one matrix is "less than" another matrix. One such criterion useful for square matrices, which is

conveniently a scalar quantity, is the sum of the diagonal elements, called the trace of the matrix. Therefore we will define the minimum variance condition to be

$$\text{Trace } (\Sigma_{\hat{X}}) = \text{minimum} \quad , \tag{6.18}$$

and we will now proceed to find the matrix B in Equation 6.16 which will satisfy this condition.

We have already seen that if E [V] = 0 then $\hat{X}$ is unbiased, that is

$$E [\hat{X}] = X \quad .$$

From the linear condition of Equation 6.16, and from the assumption E[V] = 0

$$E [\hat{X}] = E [BL] = B\ E [L] = B\ A\ X \quad .$$

Therefore

$$B\ A = I \tag{6.19}$$

From the covariance law, since

$$\hat{X} = BL$$

then

$$\Sigma_{\hat{X}} = B\Sigma_L B^T \quad . \tag{6.20}$$

The problem may now be stated that we want to find that value of B such that

$$\text{Trace } (B\Sigma_L B^T) = \text{minimum} \quad ,$$

under the constraint

$$B\,A - I = 0 \quad .$$

We will use the standard method for solving such minimization problems, called the method of Lagrange, which will be fully explained in Chapter 7. The procedure is that we form the variation function

$$\phi = B\Sigma_L B^T + 2(B\,A - I)\,K$$

where K is a matrix of undetermined constants, called Lagrange multipliers, and we then set

$$\frac{\partial\, Tr(\phi)}{\partial B} = 0$$

and from the peroperties of traces (Appendix E) we have

$$Tr\,(\phi) = Tr\,(B\Sigma_L B^T) + 2\,Tr\,(BAK) - 2Tr\,(k)$$

$$\frac{\partial Tr\,(B\Sigma_L B^T)}{\partial B} = B\,(\Sigma_L + \Sigma_L^T) = 2B\Sigma_L$$

$$\frac{\partial Tr\,(BAK)}{\partial B} = K^T A^T$$

$$\frac{\partial Tr\,(K)}{\partial B} = 0$$

so that

$$\frac{\partial Tr\,(\phi)}{\partial B} = 2B\Sigma_L + 2K^T A^T = 0$$

or

$$B = -\,K^T\,A^T\Sigma_L^{-1} \quad .$$

But from Equation 6.19

$$B\,A = I = -\,K^T\,A^T\,\Sigma_L^{-1}\,A$$

or

$$K^T = -\,(A^T\,\Sigma_L^{-1}\,A)^{-1}$$

and

$$B = (A^T\,\Sigma_L^{-1}\,A)^{-1}\,A^T\,\Sigma_L^{-1} \tag{6.21}$$

finally giving us

$$\hat{X} = B\,L = (A^T\,\Sigma_L^{-1}\,A)^{-1}\,A^T\,\Sigma_L^{-1}\,L \tag{6.22}$$

as the minimum variance estimator of X. By comparing Equations 6.6 and 6.22 we see that the least squares estimator is the minimum variance estimator when $P=\Sigma_L^{-1}$. Note that when $P=\sigma_o^2\Sigma_L^{-1}$ the equivalence between equations 6.22 and 6.6 is still valid, since the variance factor drops out of equation 6.6.

6.4 THE MAXIMUM LIKELIHOOD POINT ESTIMATOR FOR X

If the observation errors V have a normal (Gaussian) distribution, then their probability density function can be written

$$\phi(V) = C\,\exp\,[-\tfrac{1}{2}\,(V - E[V])^T\,\Sigma_L^{-1}\,(V - E[V])] \tag{6.23}$$

where the constant C has been defined in equation 3.26. If

$$E\,[V] = 0$$

$$\phi(V) = C\,\exp\,[-\tfrac{1}{2}\,V^T\,\Sigma_L^{-1}\,V] \quad . \tag{6.24}$$

It can be seen that the least squares criterion

$$\hat{V}^T\,\Sigma_L^{-1}\,\hat{V} = \text{minimum} \tag{6.25}$$

leads to a maximum probability density $\phi(\hat{V})$, also called a maximum likelihood. Therefore the estimator $\hat{X}$ of X which satisfies 6.25 is called the maximum likelihood estimator of X.

6.5 UNBIASED POINT ESTIMATORS FOR THE VARIANCE FACTOR AND THE COVARIANCE MATRIX OF X

In this section we will show that unbiased estimators exist for the variance factor σ_o^2 and the covariance matrix of the unknowns $\Sigma_{\hat{X}}$ given by

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{n-u} \tag{6.26}$$

$$\hat{\Sigma}_{\hat{X}} = \hat{\sigma}_o^2 (A^T PA)^{-1} \tag{6.27}$$

where

$$P = \sigma_o^2 \Sigma_L^{-1} \quad . \tag{6.28}$$

The covariance matrix for $\hat{X}$ is given by

$$\Sigma_{\hat{X}} = E [(\hat{X} - E [\hat{X}])(\hat{X} - E [\hat{X}])^T] \quad . \tag{6.29}$$

If E [V] = 0 we have seen that $\hat{X}$ is an unbiased estimator, that is

$$E [\hat{X}] = X \tag{6.30}$$

therefore

$$\Sigma_{\hat{X}} = E [(\hat{X} - X)(\hat{X} - X)^T] \quad . \tag{6.31}$$

Now from Equation 6.6 the expression for $\hat{X}$ is

$$\hat{X} = (A^T P A)^{-1} A^T P L \tag{6.32}$$

which can be written

$$\hat{X} = B L \tag{6.33}$$

where

$$B = (A^T P A)^{-1} A^T P \tag{6.34}$$

From the covariance law we have

$$\Sigma_{\hat{X}} = B\Sigma_L B^T$$

and thus

$$\Sigma_{\hat{X}} = (A^T PA)^{-1} A^T P\Sigma_L PA (A^T PA)^{-1} \tag{6.40}$$

and if

$$P = \sigma_o^2 \Sigma_L^{-1} \tag{6.41}$$

then

$$\Sigma_L = \sigma_o^2 P^{-1} \tag{6.42}$$

and

$$\Sigma_{\hat{X}} = (A^T PA)^{-1} A^T P \sigma_o^2 P^{-1} PA (A^T PA)^{-1} \tag{6.43}$$

$$= \sigma_o^2 (A^T PA)^{-1} \quad .$$

Therefore

$$\hat{\Sigma}_{\hat{X}} = \hat{\sigma}_o^2 \, (A^T PA)^{-1} \tag{6.44}$$

is an unbiased estimator of $\Sigma_{\hat{X}}$ if and only if an unbiased estimator $\hat{\sigma}_o^2$ of σ_o^2 can be found. We will spend the remainder of this section showing that such an unbiased estimator is given by Equation 6.26, that is we will show that

$$E\,[\hat{\sigma}_o^2] = E\left[\frac{\hat{V}^T P\, \hat{V}}{n-u}\right] = \frac{1}{n-u}\; E\,[\hat{V}^T P\, \hat{V}] = \sigma_o^2 \tag{6.45}$$

or

$$E\,[\hat{V}^T P\, \hat{V}] = \sigma_o^2\,(n-u) \quad . \tag{6.46}$$

First we recall the normal equations (Equation 6.5)

$$A^T PA\, \hat{X} = A^T PL \tag{6.47}$$

which can be written

$$A^T P\,(A\hat{X} - L) = 0 \tag{6.48}$$

or

$$(A\hat{X} - L)^T PA = 0 \tag{6.49}$$

and from which

$$A^T PL = A^T PA\hat{X} \tag{6.50}$$

and

$$L^T PA = \hat{X}^T A^T PA \quad . \tag{6.51}$$

Based on these relations we now show that

$$V^T PV - \hat{V}^T P\, \hat{V} = (\hat{X} - X)^T A^T PA(\hat{X} - X) \tag{6.52}$$

where

$$V = AX - L \tag{6.53}$$

$$\hat{V} = A\hat{X} - L \quad . \tag{6.54}$$

We begin by considering

$$\hat{V}^T P \hat{V} = (A\hat{X} - L)^T P(A\hat{X} - L)$$

$$= (A\hat{X} - L)^T PA\hat{X} - \hat{X}^T A^T PL + L^T PL \quad .$$

From Equation 6.49 the first term is zero. From Equation 6.50 the second term is $\hat{X}^T A^T PA\hat{X}$. Therefore

$$\hat{V}^T P \hat{V} = L^T PL - \hat{X}^T A^T PA\hat{X} \quad . \tag{6.55}$$

Next we consider

$$V^T P V = (AX - L)^T P(AX - L)$$

$$= X^T A^T PAX - L^T PAX - X^T A^T PL + L^T PL.$$

From Equation 6.51 the second term is $\hat{X}^T A^T PAX$. From Equation 6.50 the third term is $X^T A^T PA\hat{X}$. Therefore

$$V^T P V - \hat{V}^T P \hat{V} = \hat{X}^T A^T PA\hat{X} + X^T A^T PAX - \hat{X}^T A^T PAX - X^T A^T PA\hat{X}$$

$$= (\hat{X} - X)^T A^T PA(\hat{X} - X) \quad , \tag{6.56}$$

which can be written

$$\hat{V}^T P \hat{V} = V^T P V - (\hat{X} - X)^T A^T PA (\hat{X} - X) \quad . \tag{6.57}$$

The next step is to prove that the value of the quadratic form $Y^T A Y$ is equal to the trace of the product $Y Y^T A$, that is

$$Y^T A Y = \text{Trace } (Y Y^T A) \quad . \tag{6.58}$$

The quadratic form

$$Y^T A Y$$

is a scalar, therefore it is equal to its trace

$$\text{Tr } (Y^T A Y) = Y^T A Y.$$

From the properties of traces (Appendix E) we see

$$\text{Tr } ((Y^T A)Y) = \text{Tr } (Y(Y^T A))$$

[ ]

so that

$$\text{Trace } (YY^T A) = Y^T AY \quad .$$

Applying this to Equation 6.57 we have

$$\hat{V}^T P \hat{V} = \text{Trace } (VV^T P) - \text{Trace } ((\hat{X} - X)(\hat{X} - X)^T A^T PA) . \tag{6.59}$$

This is true for any weight P. If we define P as in Equation 6.28 then from Equations 6.38 and 6.43

$$A^T PA = \sigma_o^2 \Sigma_{\hat{X}}^{-1} \tag{6.60}$$

and using Equation 6.11, Equation 6.59 becomes

$$\hat{V}^T P \hat{V} = \text{Trace } (VV^T \sigma_o^2 \Sigma_V^{-1}) - \text{Trace } ((\hat{X}-X)(\hat{X}-X)^T \sigma_o^2 \Sigma_{\hat{X}}^{-1}) \tag{6.61}$$

and the expected value of this equation is

$$E[\hat{V}^T P \hat{V}] = \sigma_o^2 \, E[\text{Trace } (VV^T \Sigma_V^{-1})] - \sigma_o^2 \, E[\text{Trace } ((\hat{X}-X)(\hat{X}-X)^T \Sigma_{\hat{X}}^{-1})]$$

$$= \sigma_o^2 \, \text{Trace } (E[VV^T \Sigma_V^{-1}]) - \sigma_o^2 \, \text{Trace } (E[(\hat{X}-X)(\hat{X}-X)^T \Sigma_{\hat{X}}^{-1}])$$

$$= \sigma_o^2 \, \text{Trace } (E[VV^T] \Sigma_V^{-1}) - \sigma_o^2 \, \text{Trace } (E[(\hat{X}-X)(\hat{X}-X)^T] \Sigma_{\hat{X}}^{-1}) \quad . \quad (6.62)$$

Therefore if there are n observations and u unknowns (taking into consideration Equations 6.11 and 6.31)

$$E[\hat{V}^T P \hat{V}] = \sigma_o^2 \, (\text{Trace } (\Sigma_V \Sigma_V^{-1}) - \text{Trace } (\Sigma_{\hat{X}} \Sigma_{\hat{X}}^{-1}))$$

$$= \sigma_o^2 \, (\text{Trace } I_n - \text{Trace } I_u)$$

$$= \sigma_o^2 \, (n - u) \quad . \quad (6.63)$$

Therefore the correct P must be used to obtain an unbiased $\hat{\sigma}_o^2$, and we have shown Equation 6.46 is true, which means that Equation 6.26 defines an unbiased estimator of σ_o^2, and from Equation 6.44 we see that Equation 6.27 defines an unbiased estimator of $\Sigma_{\hat{X}}$.

6.6 SUMMARY

In this Chapter we have shown that for the linear mathematical model

$$A X - L = V$$

i) the least squares estimator $\hat{X}$ of X is

$$\hat{X} = (A^T P A)^{-1} A^T P L$$

ii) this estimator is <u>unique</u> if $(A^T P A)$ is nonsingular

iii) this estimator is <u>unbiased</u> if E [V] = 0

iv) this estimator is the <u>minimum variance</u> estimator of X if

$$P = \sigma_o^2 \, \Sigma_L^{-1}$$

v) this estimator is the <u>maximum likelihood</u> estimator of X if V has a normal distribution

vi) the least squares unbiased estimator $\hat{V}$ of V is

$$\hat{V} = A \hat{X} - L$$

vii) the least squares unbiased estimator $\hat{\sigma}_o^2$ of the variance factor σ_o^2 is

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{n - u}$$

where

$$P = \sigma_o^2 \, \Sigma_L^{-1}$$

viii) the least squares unbiased estimator $\hat{\Sigma}_{\hat{X}}$ of $\Sigma_{\hat{X}}$ is

$$\hat{\Sigma}_{\hat{X}} = \hat{\sigma}_o^2 \, (A^T P A)^{-1}$$

where

$$P = \sigma_o^2 \, \Sigma_L^{-1} \, .$$

7. LEAST SQUARES POINT ESTIMATORS: NONLINEAR MATHEMATICAL MODELS

In Chapter 6 we concentrated on demonstrating the statistical significance of the least squares point estimators under a variety of assumptions, and assuming a linear mathematical model.

Nonlinear mathematical models occur far more frequently than do linear models. Therefore in this Chapter the emphasis is shifted to considering the steps required to obtain expressions which can be used for numerical calculations from nonlinear mathematical models. There are three steps: Linearization of the mathematical model; derivation of the normal equations from the least squares criterion; and derivation of expressions for the least squares point estimators from the normal equations. It is only these last expressions which are coded into a computer program , or otherwise set up for numerical computation.

In dealing with nonlinear mathematical models it is usual to specify an initial approximation to the solution vector, which we will denote by X°, and then to determine a correction to this value, which we will denote by X. We will call the sum of the two the total solution vector

$$\bar{X} = X^{\circ} + X \quad . \tag{7.1}$$

Similarly we will often want to refer to the sum of the observation

vector L and the residual vector V as the total observation vector

$$\bar{L} = L + V \ . \tag{7.2}$$

7.1 LINEARIZING THE MATHEMATICAL MODEL

Mathematical models expressing the relationship between the total observed and solution vectors $\bar{X}$ and $\bar{L}$ have the general form

$$F(\bar{X}, \bar{L}) = 0 \ , \tag{7.3}$$

which is the mathematical model of what is known as the combined method for reasons which will soon be clear. The vector function F represents r equations relating n observations and u unknowns. The method of least squares can be applied only when $n + u > r > u$. The quantity $(r - u)$ is called the redundancy or number of degrees of freedom.

If the observed quantities can be explicitly expressed as functions of the parameters $\bar{X}$ the mathematical model becomes

$$F(\bar{X}) = \bar{L} \ , \tag{7.4}$$

and the method is called the parametric method (also called the method of observation equations, the method of parametric equations, and the method of indirect observations). In this case the vector function F represents n equations (one equation per observation), and the redundancy is given by $(n - u)$.

If the mathematical model consists of conditions between the observed quantities, that is

$$F(\bar{L}) = 0 \ , \tag{7.5}$$

then the method is called the condition method (also called the method of condition equations, the method of conditional observations, and the method of correlates). In this case the vector function F represents

(n - u) equations (one equation per degree of freedom).

Linearization of these models is accomplished by replacing the nonlinear functions F by their Taylor's series linear approximation, expanded about the point defined by the initial approximation to the solution vector (X°), and the measured values of the observation vector (L).

For the combined method linearization gives

$$F(\bar{X}, \bar{L}) = F(X^\circ, L) + \left.\frac{\partial F}{\partial \bar{X}}\right|_{X^\circ,L} X + \left.\frac{\partial F}{\partial \bar{L}}\right|_{X^\circ,L} V = 0$$

or

$$W + AX + BV = 0 \ , \tag{7.6}$$

where ${}_{r}W_{1} = F(X^\circ, L)$ is called the misclosure vector and

${}_{r}A_{u} = \left.\frac{\partial F}{\partial \bar{X}}\right|_{X^\circ,L}$ and ${}_{r}B_{n} = \left.\frac{\partial F}{\partial \bar{L}}\right|_{X^\circ,L}$ are called the design matrices.

For the parametric method linearization gives

$$F(\bar{X}) - \bar{L} = F(X^\circ) + \left.\frac{\partial F}{\partial \bar{X}}\right|_{X^\circ} X - (L + V) = 0$$

or

$$W + AX - V = 0 \ , \tag{7.7}$$

where ${}_{n}W_{1} = F(X^\circ) - L$ and ${}_{n}A_{u} = \left.\frac{\partial F}{\partial \bar{X}}\right|_{X^\circ}$.

For the condition method linearization gives

$$F(\bar{L}) = F(L) + \left.\frac{\partial F}{\partial \bar{L}}\right|_{L} V = 0$$

or

$$W + BV = 0 \ , \tag{7.8}$$

where ${}_{n-u}W_1 = F(L)$ and ${}_{n-u}B_n = \left.\frac{\partial F}{\partial \bar{L}}\right|_{L}$.

The parametric and condition methods are the classical methods of least squares estimation and are special cases of the combined method. Many problems can be solved by either method. The condition method has two drawbacks. Specifying the (n - u) conditions to be used is usually difficult compared to writing the parametric equations involved. If a solution for the unknown parameters are desired, as is usually the case, then after the condition method solution is complete, further work must be done to obtain this solution.

On the other hand the condition method requires the solution of only (n - u) equations rather than n equations for the parametric method. This consideration overrode the drawbacks in the days of hand computations, and most least squares work was done by the condition method. However with the use of the digital computer the advantage of fewer equations has been erased, so now the parametric and more recently the combined methods are usually used. The three methods are summarized in the following table.

	Combined	Parametric	Condition
Mathematical Model	$F(\bar{X}, \bar{L}) = 0$	$F(\bar{X}) - \bar{L} = 0$	$F(\bar{L}) = 0$
number of equations	r	n	n - u
number of observations	n	n	n
number of unknowns	u	u	-
linearized math model	$W + AX + BV = 0$	$W + AX - V = 0$ special case of combined with $\begin{cases} B = -I \\ r = n \end{cases}$	$W + BV = 0$ special case of combined with $\begin{cases} A = 0 \\ r = n - u \end{cases}$

7.2 LINEARIZATION EXAMPLES

In this section we give two examples of mathematical models and their linearization: One using the combined method only; one using both parametric and condition methods. A third numerical example, using both combined and parametric methods is given in section A.2 of the Appendix.

Example One. This example is the fitting of the "best" straight line to a set of data points, using the least squares criterion. The combined method is the only one which can be used. The mathematical model

$$y_i = mx_i + b$$

relates r sets of observed coordinates (x_i, y_i) to a straight line with slope m and intercept b. There are r equations, n = 2r observations, and 2 unknowns. The total solution vector is

$$_2\bar{X}_1 = X^\circ + X = \begin{bmatrix} \bar{m} \\ \\ \bar{b} \end{bmatrix} = \begin{bmatrix} m^\circ \\ \\ b^\circ \end{bmatrix} + \begin{bmatrix} m \\ \\ b \end{bmatrix}$$

the total observation vector is

$$_{2r}\bar{L}_1 = L + V = \begin{bmatrix} \bar{x}_1 \\ \bar{y}_1 \\ \cdot \\ \cdot \\ \bar{x}_r \\ \bar{y}_r \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ \cdot \\ \cdot \\ x_r \\ y_r \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ \cdot \\ v_{2r} \end{bmatrix}$$

and the combined method mathematical model consists of r equations like

$$f_i(\bar{X}, \bar{L}) = \bar{m}\,\bar{x}_i + \bar{b} - \bar{y}_i = 0$$

which, after linearizing about (X°, L) becomes

$$AX + BV + W = 0$$

where

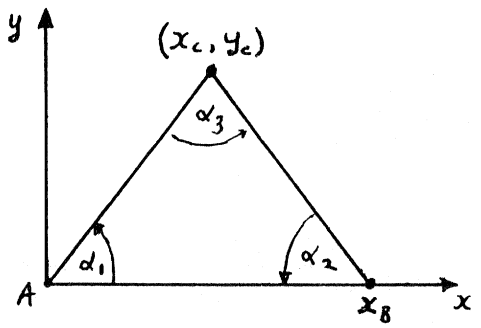
$$r^{A}_{2} = \begin{bmatrix} x_1 & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ x_r & 1 \end{bmatrix},$$

$$r^{B}_{2r} = \begin{bmatrix} m^\circ & -1 & \vdots & 0 & 0 \\ \cdot & & \vdots & & \cdot \\ \cdot & & \vdots & & \cdot \\ 0 & 0 & \vdots & m^\circ & -1 \end{bmatrix}$$

and

$$r^{W}_{1} = \begin{bmatrix} m^\circ x_1 + b^\circ - y \\ \cdot \\ \cdot \\ \cdot \\ m^\circ x_r + b^\circ - y_r \end{bmatrix}.$$

Example Two. This example illustrates the use of both parametric and condition methods. The triangle ABC has known points A and B at (0, 0) and $(x_B, 0)$ and unknown point C at (x_c, y_c). The three interior angles of the triangle are measured.



The total solution vector is

$$2\bar{X}_1 = X^\circ + X = \begin{bmatrix} \bar{x}_c \\ \bar{y}_c \end{bmatrix} = \begin{bmatrix} x_o \\ y_o \end{bmatrix} + \begin{bmatrix} x \\ y \end{bmatrix}$$

the total observation vector is

$$3\bar{L}_1 = L + V = \begin{bmatrix} \bar{\alpha}_1 \\ \bar{\alpha}_2 \\ \bar{\alpha}_3 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

and the parametric equations are

$$\bar{\alpha}_1 = \arctan\,(\bar{y}_c/\bar{x}_c)$$

$$\bar{\alpha}_2 = \arctan\,(\bar{y}_c/(x_B - \bar{x}_c))$$

$$\bar{\alpha}_3 = \arctan\,(\bar{x}_c/\bar{y}_c) + \arctan((x_B - \bar{x}_c)/\bar{y}_c)$$

which, when linearized about (X°, L) becomes

$$V = AX + W$$

where

$$3A_2 = \begin{bmatrix} \dfrac{-y_o}{x_o^2+y_o^2} & \dfrac{x_o}{x_o^2+y_o^2} \\ \dfrac{y_o}{(x_B-x_o)^2 + y_o^2} & \dfrac{x_B-x_o}{(x_B-x_o)^2 + y_o^2} \\ \dfrac{y_o}{x_o^2+y_o^2} - \dfrac{y_o}{(x_B-x_o)^2 + y_o^2} & \dfrac{-x_o}{x_o^2+y_o^2} - \dfrac{(x_B-x_o)}{(x_B-x_o)^2 + y_o^2} \end{bmatrix}$$

and

$$ {}_3W_1 = \begin{bmatrix} \arctan\left(y_o/x_o\right) - \alpha_1 \\ \arctan\left(y_o/(x_B - x_o)\right) - \alpha_2 \\ \arctan\left(x_o/y_o\right) + \arctan\left((x_B - x_o)/y_o\right) - \alpha_3 \end{bmatrix} $$

Applying the condition method, there is only $(n - u) = (3 - 2) = 1$ condition which is

$$ F(\bar{L}) = \bar{\alpha}_1 + \bar{\alpha}_2 + \bar{\alpha}_3 - 180^\circ = 0 $$

or

$$ BV + W = 0 \ , $$

where

$$ {}_1B_3 = [1 \quad 1 \quad 1] $$

and

$$ {}_1W_1 = \alpha_1 + \alpha_2 + \alpha_3 - 180^\circ \ . $$

Note however that once this has been solved for V we will still have no knowledge of $\bar{X}$.

7.3 DERIVATION OF THE NORMAL EQUATIONS

The normal equations express the relationship between the least squares estimators $\hat{X}$ of the solution vector X and $\hat{V}$ of the residual vector V (and as we shall soon see the estimator $\hat{K}$ of the vector of Lagrange multipliers K) and the known quantities P, A, B and W. The normal equations result from applying the least squares criterion

$$ \hat{V}^T P \hat{V} = \text{minimum} \qquad (7.9) $$

to the linearized mathematical model, which for the general case in this

Chapter is the combined method

$$A\hat{X} + B\hat{V} + W = 0 \ . \tag{7.10}$$

Mathematically this is called the extremal problem with constraints; that is we wish to find an extremum (a maximum or minimum value) for one function when the variables are related to each other by other functions. The standard method for handling such problems is called Lagrange's method. We will describe this method using a simple two dimensional example.

Suppose we want to minimize the function

$$f_1(x, y) = ax^2 + by^2 \tag{7.11}$$

subject to the constraint

$$f_2(x, y) = ax + by + c = 0 \ . \tag{7.12}$$

Applying Lagrange's method we perform three steps:

i) form the variation function

$$\phi = f_1(x, y) + kf_2(x, y) = ax^2 + by^2 + k(ax + by + c)$$

where k is an undetermined constant called the Lagrange multiplier.

ii) set the derivatives of the variation function to zero

$$\frac{\partial \phi}{\partial x} = 2ax + ka = 0 \quad \text{and} \quad \frac{\partial \phi}{\partial y} = 2by + kb = 0$$

iii) solve the three equations

$$f_2(x, y) = 0 \ , \ \frac{\partial \phi}{\partial x} = 0 \ , \ \frac{\partial \phi}{\partial y} = 0$$

for the three unknowns x, y, k.

The point (x, y) so determined is an extreme point of $f_1(x, y)$. The

value of k so determined is closely related to the value of $f_1(x, y)$ at the extreme point. For our example the solution is

$$x = \frac{-c}{a+b} \quad , \quad y = \frac{-c}{a+b} \quad , \quad k = \frac{2c}{a+b} \qquad \text{and} \qquad f_1(x, y) = \frac{a+b}{4} k^2 .$$

Let us interpret this process geometrically. The first function (Equation 7-11) is a family of ellipses, and the second function (Equation 7.12) is a single straight line. What we have done is to find the particular ellipse which just touches the straight line, and the solution (x, y) is the point of tangency between the ellipse and the straight line, as shown in Figure 7.1.

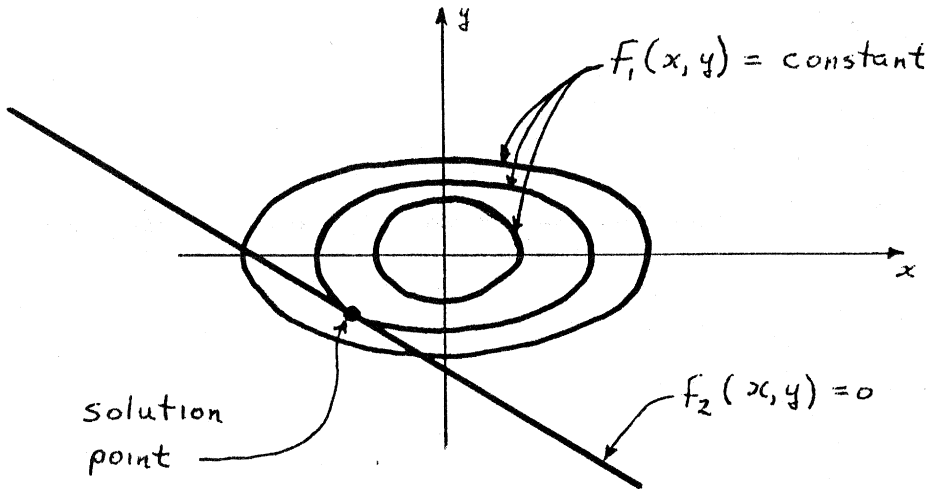


Figure 7.1

Extending this method to the combined case we want to minimize the function

$$\hat{V}^T P \hat{V}$$

subject to the r constraints (because we have r equations)

$$A\hat{X} + B\hat{V} + W = 0 \ . \tag{7.13}$$

Geometrically we want to find the solution point $(\hat{X}, \hat{V})$ in hyperspace which is the point of tangency between one of the family of hyperellipsoids

defined by

$$\hat{V}^T P \hat{V} = \text{constant}$$

and the hyperplane defined by Equation 7.13.

Following the same procedure as in the simple example above, we form the variation function

$$\phi = \hat{V}^T P \hat{V} + 2\hat{K}^T (A\hat{X} + B\hat{V} + W) , \qquad (7.14)$$

where $\hat{K}$ is the least squares estimator of K, a column vector of r Lagrange multipliers, and has been multiplied by a factor 2 for convenience. Setting the derivatives of ϕ with respect to $\hat{X}$ and $\hat{V}$ to zero we have

$$\frac{\partial \phi}{\partial \hat{V}} = 2\hat{V}^T P + 2\hat{K}^T B = 0$$

which, when transposed and divided by 2 becomes

$$P \hat{V} + B^T \hat{K} = 0 \quad , \qquad (7.15)$$

and

$$\frac{\partial \phi}{\partial \hat{X}} = 2\hat{K}^T A = 0 ,$$

which when transposed and divided by 2 becomes

$$A^T \hat{K} = 0 . \qquad (7.16)$$

We now want to find a simultaneous solution of the three equations

$$A\hat{X} + B\hat{V} + W = 0$$

$$P\hat{V} + B^T \hat{K} = 0$$

$$A^T \hat{K} = 0$$

for $\hat{X}$, $\hat{V}$ and $\hat{K}$. These three equations can be combined into the single hypermatrix equation (a hypermatrix being a matrix whose elements are themselves matrices)

$$\begin{bmatrix} P & B^T & 0 \\ B & 0 & A \\ 0 & A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{K} \\ \hat{X} \end{bmatrix} + \begin{bmatrix} 0 \\ W \\ 0 \end{bmatrix} = 0 \ . \tag{7.17}$$

This hypermatrix equation represents the normal equations for the combined method in their most expanded form. For the parametric method B and B^T are replaced by -I to get the normal equations in hypermatrix form

$$\begin{bmatrix} P & -I & 0 \\ -I & 0 & A \\ 0 & A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{K} \\ \hat{X} \end{bmatrix} + \begin{bmatrix} 0 \\ W \\ 0 \end{bmatrix} = 0 \ . \tag{7.18}$$

For the condition method A = 0, which reduces the normal equations to

$$\begin{bmatrix} P & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{K} \end{bmatrix} + \begin{bmatrix} 0 \\ W \end{bmatrix} = 0 \ . \tag{7.19}$$

Note that in all three cases the hypercoefficient matrix has been constructed to be symmetric, with a nonsingular upper left element. This is a necessary condition on the structure of the hypercoefficient matrix.

It would be possible (for small problems at least) merely to program this one equation directly to solve for $\hat{X}$, $\hat{K}$, and $\hat{V}$ as partitions of the hypersolution vector (for the combined case)

$$\begin{bmatrix} \hat{V} \\ \hat{K} \\ \hat{X} \end{bmatrix} = - \begin{bmatrix} P & B^T & 0 \\ B & 0 & A \\ 0 & A^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ W \\ 0 \end{bmatrix} \ . \tag{7.20}$$

However, the time required to invert matrices goes up as the cube of their size, so it always is more efficient to invert the smallest possible matrix, not the largest as we have here. Also this hyper-

coefficient matrix consists in large part of zero elements, the storage of which in a computer is unnecessarily wasteful. Therefore in the next section we will derive more explicit expressions for the solutions for $\hat{X}$, $\hat{K}$ and $\hat{V}$.

7.4 DERIVATION OF EXPLICIT EXPRESSIONS FOR THE SOLUTION TO THE NORMAL EQUATIONS

In this section we will derive expressions for the estimator $\hat{X}$ of the solution vector X, the estimator $\hat{K}$ of the vector of Lagrange multipliers K, and the estimator $\hat{V}$ of the residual vector V. We start with the hypermatrix equation 7.17, treat the hypercoefficient matrix as a partitioned matrix, and use the rules for inverting partitioned matrices to eliminate $\hat{V}$ and $\hat{K}$ from the solution. Once an expression for $\hat{X}$ is obtained it is back-substituted to obtain expressions for $\hat{K}$ and $\hat{V}$. This procedure is equivalent to the more familiar elimination and back-substitution technique which does not use hypermatrices. However the approach used here is a standard step by step procedure which will be of great advantage when we consider more complicated mathematical models in Chapter 9.

The hypermatrix Equation 7.17

$$\begin{bmatrix} P & B^T & 0 \\ B & 0 & A \\ 0 & A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{K} \\ \hat{X} \end{bmatrix} + \begin{bmatrix} 0 \\ W \\ 0 \end{bmatrix} = 0 \qquad (7.21)$$

is of the form

$$NY + U = 0 \ ,$$

which can be partitioned

$$\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \quad \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \quad + \quad \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \quad = 0. \qquad (7.22)$$

The two matrix equations can be stated separately

$$N_{11}\, Y_1 + N_{12}\, Y_2 + U_1 = 0 \qquad (7.23)$$

$$N_{21}\, Y_1 + N_{22}\, Y_2 + U_2 = 0. \qquad (7.24)$$

If N_{11} is nonsingular, then from Equation 7.23

$$Y_1 = - N_{11}^{-1} (U_1 + N_{12}\, Y_2)$$

and eliminating Y_1 from Equation 7.24

$$-N_{21}\, N_{11}^{-1} (U_1 + N_{12}\, Y_2) + N_{22}\, Y_2 + U_2 = 0$$

or

$$(N_{22}-N_{21}N_{11}^{-1}N_{12})\, Y_2 + (U_2-N_{21}N_{11}^{-1}U_1) = 0 \qquad (7.25)$$

To eliminate $\hat{V}$ we partition hypermatrix Equation 7.21 so that $Y_1 = \hat{V}$

$$\left[\begin{array}{c|c|c} P & B^T & 0 \\ \hline B & 0 & A \\ 0 & A^T & 0 \end{array}\right] \quad \left[\begin{array}{c} \hat{V} \\ \hline \hat{K} \\ \hat{X} \end{array}\right] \quad + \quad \left[\begin{array}{c} 0 \\ \hline W \\ 0 \end{array}\right] \quad = 0 .$$

Then from Equation 7.25

$$\left[\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} - \begin{bmatrix} B \\ 0 \end{bmatrix} P^{-1} \begin{bmatrix} B^T & 0 \end{bmatrix}\right] \begin{bmatrix} \hat{K} \\ \hat{X} \end{bmatrix} + \left[\begin{bmatrix} W \\ 0 \end{bmatrix} - \begin{bmatrix} B \\ 0 \end{bmatrix} P^{-1} \begin{bmatrix} 0 \end{bmatrix}\right] = 0 ,$$

or

$$\begin{bmatrix} -BP^{-1}B^T & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{K} \\ \hat{X} \end{bmatrix} + \begin{bmatrix} W \\ 0 \end{bmatrix} = 0 . \qquad (7.26)$$

To eliminate $\hat{K}$ we partition hypermatrix Equation 7.26 so that $Y_1 = \hat{K}$

$$\left[\begin{array}{c|c} -BP^{-1}B^T & A \\ \hline A^T & 0 \end{array}\right] \left[\begin{array}{c} \hat{K} \\ \hline \hat{X} \end{array}\right] + \left[\begin{array}{c} W \\ \hline 0 \end{array}\right] = 0 .$$

Then from Equation 7.25

$$(A^T (BP^{-1}B^T)^{-1} A)\hat{X} + A^T (BP^{-1}B^T)^{-1} W = 0 ,$$

or

$$\boxed{\hat{X} = - \left(A^T (BP^{-1}B^T)^{-1} A\right)^{-1} A^T (BP^{-1}B^T)^{-1} W .} \qquad (7.27a)$$

The first equation from hypermatrix Equation 7.26 is

$$- (BP^{-1}B^T)\hat{K} + A\hat{X} + W = 0 ,$$

from which

$$\boxed{\hat{K} = (BP^{-1}B^T)^{-1} (A\hat{X} + W) .} \qquad (7.27b)$$

The first equation from hypermatrix Equation 7.21 is

$$P\hat{V} + B^T\hat{K} = 0 ,$$

from which

$$\boxed{\hat{V} = - P^{-1}B^T\hat{K} .} \qquad (7.27c)$$

Equations 7.27 are the expressions for the least squares estimators using the combined method. For the parametric method B = - I, therefore equations 7.27 become

$$\boxed{\begin{aligned} \hat{X} &= - (A^TPA)^{-1} A^TPW \\ \hat{K} &= P (A\hat{X} + W) \\ \hat{V} &= P^{-1}\hat{K} = A\hat{X} + W . \end{aligned}} \qquad (7.28)$$

For the condition method A = 0, therefore Equations 7.27b and 7.27c become

$$\boxed{\begin{aligned} \hat{K} &= (BP^{-1}B^T)^{-1} W \\ \hat{V} &= - P^{-1}B^T\hat{K} = - P^{-1}B^T(BP^{-1}B^T)^{-1} W . \end{aligned}} \qquad (7.29)$$

These solutions for $\hat{X}$ and $\hat{V}$ must be added to the initial approximation X° and measured values L to obtain estimates of the total solution vector

$$\hat{\bar{X}} = X^\circ + \hat{X} \qquad (7.30)$$

and the total observation vector

$$\hat{\bar{L}} = L + \hat{V} . \qquad (7.31)$$

7.5 DERIVATION OF EXPRESSIONS FOR COVARIANCE MATRICES

In this section we will apply the covariance law (Equation 2.23) to derive expressions for the covariance matrices of the random variables W, $\hat{X}$, $\hat{K}$, $\hat{V}$, given the covariance matrix Σ_L for the observations L.

From the definition of the misclosure vector W, the design matrix B, and the total observation vector $\bar{L}$

$$W = F(X^\circ, L)$$

$$B = \left.\frac{\partial F}{\partial \bar{L}}\right|_{X^\circ, L}$$

$$\bar{L} = L + V \ ,$$

it is obvious that from the covariance law

$$\Sigma_W = C_W \Sigma_L \, C_W^T \ ,$$

where

$$C_W = \frac{\partial}{\partial L}\left(F(X^\circ, L)\right) = \frac{\partial}{\partial \bar{L}}\left(F(\bar{X}, \bar{L})\right)\Big|_{X^\circ, L} = B \ .$$

Therefore

$$\boxed{\Sigma_W = B \, \Sigma_L \, B^T \ .} \qquad (7.32)$$

From Equation 7-27a

$$\hat{X} = C_X W \ ,$$

and therefore from the covariance law

$$\Sigma_{\hat{X}} = C_X \, \Sigma_W \, C_X^T$$

$$= C_X \, B\Sigma_L \, B^T C_X^T \ ,$$

where

$$C_X = - \left(A^T \, (B \, \Sigma_L \, B^T)^{-1} \, A\right)^{-1} A^T (B \, \Sigma_L \, B^T)^{-1} \ ,$$

and

$$\Sigma_L = \sigma_o^2 \, P^{-1} \ .$$

Therefore

$$\Sigma_{\hat{X}} = (A^T(B\Sigma_L B^T)^{-1}A)^{-1}A^T(B\Sigma_L B^T)^{-1}[B\Sigma_L B^T](B\Sigma_L B^T)^{-1}A(A^T(B\Sigma_L B^T)^{-1}A)^{-1} ,$$

which cancels to become

$$\boxed{\Sigma_{\hat{X}} = (A^T(B\Sigma_L B^T)^{-1}A)^{-1} .} \tag{7.33}$$

From Equations 7-27b

$$\hat{K} = C^*(A\hat{X} + W)$$

$$= C^*(AC_X + I)W = C_K W ,$$

and therefore from the covariance law

$$\Sigma_{\hat{K}} = C_K \Sigma_W C_K^T$$

$$= C_K B\Sigma_L B^T C_K^T ,$$

where

$$C_K = C^*(AC_X + I)$$

$$= \sigma_o^2(B\Sigma_L B^T)^{-1}\left[I - A(A^T(B\Sigma_L B^T)^{-1}A)^{-1}A^T(B\Sigma_L B^T)^{-1}\right] .$$

Therefore, after simplifying

$$\boxed{\Sigma_{\hat{K}} = (\sigma_o^2)^2(B\Sigma_L B^T)^{-1}\left[I - A(A^T(B\Sigma_L B^T)^{-1}A)^{-1}A^T(B\Sigma_L B^T)^{-1}\right] .} \tag{7.34}$$

From Equation 7-27c

$$\hat{V} = - P^{-1} B^T \hat{K}$$

therefore from the covariance law

$$\Sigma_{\hat{V}} = \left(\frac{1}{\sigma_o^2}\right)^2 \Sigma_L B^T \Sigma_{\hat{K}} B \Sigma_L$$

or

$$\Sigma_{\hat{V}} = \Sigma_L B^T (B\Sigma_L B^T)^{-1} \left[B\Sigma_L - A(A^T (B\Sigma_L B^T)^{-1} A)^{-1} A^T (B\Sigma_L B^T)^{-1} B\Sigma_L \right] \quad (7.35)$$

Setting

$$\Sigma_L = \sigma_o^2 P^{-1}$$

in Equations 7.32, 7.33 and 7.35

$$\Sigma_W = \sigma_o^2 (BP^{-1} B^T) \quad (7.36a)$$

$$\Sigma_{\hat{X}} = \sigma_o^2 (A^T (BP^{-1} B^T)^{-1} A)^{-1} \quad (7.36b)$$

$$\Sigma_{\hat{V}} = \sigma_o^2 P^{-1} B^T (BP^{-1} B^T)^{-1} \left[BP^{-1} - A(A^T (BP^{-1} B^T)^{-1} A)^{-1} A^T (BP^{-1} B^T)^{-1} BP^{-1} \right] \quad (7.36c)$$

Recalling the discussion of section 6.5, equations 7.36 become unbiased estimators only when the a priori variance factor σ_o^2 is replaced by the unbiased estimator

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{\nu}$$

where ν is the number of degrees of freedom.

For the parametric method B = - I and Equations 7.36 become

$$\Sigma_W = \sigma_o^2 P^{-1} \quad (7.37a)$$

$$\Sigma_{\hat{X}} = \sigma_o^2 (A^T P A)^{-1} \quad (7.37b)$$

$$\Sigma_{\hat{V}} = \sigma_o^2 (P^{-1} - A(A^T P A)^{-1} A^T) . \quad (7.37c)$$

For the condition method A = 0 and Equations 7.36 become

$$\Sigma_W = \sigma_o^2 \, (BP^{-1} B^T) \tag{7.38a}$$

$$\Sigma_{\hat{V}} = \sigma_o^2 \, (P^{-1}B^T(BP^{-1}B^T)^{-1} BP^{-1}) \tag{7.38b}$$

8. MULTIVARIATE INTERVAL ESTIMATION AND HYPOTHESIS TESTING

8.1 INTRODUCTION

In Chapter 5 we treated the problem of univariate interval estimation which involved population distribution functions containing only one random variable. In this Chapter we treat multivariate interval estimation, which is an extension of the univariate concept to include population distribution functions containing several random variables.

Multivariate point estimation was discussed in detail in Chapters 6 and 7. An example of this type of problem is the solution for the coordinates of points in a network containing a redundant set of observations among the points. On the other hand, multivariate interval estimation involves these many points or the determination of a confidence region for these points taken all together, taken in groups, or considered one at a time (for example, an error ellipse about a point in two dimensions, or an error ellipsoid for a point in three dimensions). We will develop confidence regions for the following quantities (assuming the observations to be normally distributed):

1) the variance factor σ_o^2,
2) ratio of two variance factors - $(\sigma_o^2)_2/(\sigma_o^2)_1$,
3) the quadratic form for deviations from the estimated solution vector $\hat{X}$ when the variance factor is known:

$$(\hat{X} - X)^T \Sigma_{\hat{X}}^{-1} (\hat{X} - X) ,$$

4) the quadratic form for the deviations from the estimated solution vector $\hat{X}$ when the variance factor is not known:

$$[(\hat{X} - X)^T \hat{\Sigma}_{\hat{X}}^{-1} (\hat{X} - X)]/u .$$

Before we begin the Chapter proper, a brief statement on quadratic forms is in order [Wells 1971, pp. 36-40]. A quadratic form is presented by

$$\underset{1xu}{X^T} \quad \underset{uxu}{A} \quad \underset{uxl}{X} = k ,$$

where

$$\underset{uxl}{X} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_u \end{bmatrix}$$

is a uxl vector of random variables, A is a uxu symmetric matrix, and k is the value of the quadratic form. The quadratic form can be taken as the equation of an ellipsoid in u dimensions; this aspect will receive special attention in Section 8.5.

8.2 EXAMINATION OF THE VARIANCE FACTOR

The variance factor, σ_o^2 can be examined in terms of the estimated residual vector $\hat{V}$ and weights P. Recall the relationship of the covariance (Σ_L) and weight (P) matrices, namely

$$\Sigma_L = \sigma_o^2 P^{-1}$$

and

$$P = \sigma_o^2 \Sigma_L^{-1} .$$

The estimated variance factor $\hat{\sigma}_o^2$ is computed from

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{\nu} \quad ,$$

where $\hat{V}$ is the estimated residual vector and ν is the degrees of freedom.

Recall for the univariate case with known mean μ (Section 4.6) that

$$\sum_1^n \left(\frac{x_i - \mu}{\sigma}\right)^2 \xrightarrow{d} \chi^2 (n) \; ;$$

Note that the number of observations n equals the degrees of freedom. With unknown mean, that is $\bar{x}$ taken as an estimator of μ, the random variable

$$(n-1)\,\frac{s^2}{\sigma^2} = \sum_1^n \left(\frac{x_i - \bar{x}}{\sigma}\right)^2 \xrightarrow{d} \chi^2(n-1) \; ; \qquad 8\text{-}1$$

Note in this case that the degrees of freedom is n-1 instead of n as the degrees of freedom have been reduced by one due to the fact that μ is unknown and is estimated by $\bar{x}$.

For the multivariate case, u unknown parameters are to be estimated, that is

$$\underset{1 \times u}{\hat{X}^T} = [\hat{x}_1 \quad \hat{x}_2 \; . \; . \; . \; . \; \hat{x}_u] \; .$$

The combined and parametric cases are used as the estimation technique; in the combined case of r equations, the degrees of freedom $\nu = r-u$, while in the parametric case $\nu = n-u$.

$$\boxed{\nu \frac{\hat{\sigma}_o^2}{\sigma_o^2} = \frac{\hat{V}^T P \hat{V}}{\sigma_o^2} = \hat{V}^T \Sigma_L^{-1} \hat{V} \xrightarrow{d} \chi^2(\nu) \; ,} \qquad 8\text{-}2$$

The probability statement for the above chi-square random variable is

$$\Pr\left(\chi^2_{p_1} \leq \frac{\nu\hat{\sigma}_o^2}{\sigma_o^2} \leq \chi^2_{p_2}\right) = \alpha \quad ,$$

while the associated confidence interval for σ_o^2 is

$$\left[\frac{\nu\hat{\sigma}_o^2}{\chi^2_{p_2}} \leq \sigma_o^2 \leq \frac{\nu\hat{\sigma}_o^2}{\chi^2_{p_1}}\right]$$

$$\left[\frac{\hat{V}^T P \hat{V}}{\chi^2_{p_2}} \leq \sigma_o^2 \leq \frac{\hat{V}^T P \hat{V}}{\chi^2_{p_1}}\right] \quad . \qquad 8\text{-}3$$

The bounds of the interval are evaluated from:

1) the known degrees of freedom ν,

2) the computed value of the estimated variance factor $\hat{\sigma}_o^2$,

3) the tabulated values of $\chi^2_{p_1}$ and $\chi^2_{p_2}$ (Appendix B-2)

corresponding to $\chi^2(\nu), P_1 = \frac{1-\alpha}{2}$ and $P_2 = \frac{1+\alpha}{2}$.

The above confidence interval is used to test the null hypothesis

$$H_o: \quad \sigma_o^2 = (\sigma_o^2)_H \quad .$$

The relevance of the above relates to the choice of the weight matrix, that is $P = \sigma_o^2 \Sigma_L^{-1}$. If σ_o^2 is taken as unity then $P = \Sigma_L^{-1}$. Hypothesized values of σ_o^2 may be made which implies a change of scale of the covariance matrix Σ_L, since $\Sigma_L = \sigma_o^2 P^{-1}$. It should be noted that rejection of H_o could be due to phenomena other than the incorrect scale of the covariance matrix, that is

1) shortcomings in the mathematical model,

2) non-normal distribution of the random variables in the residual vector.

The above two items can also be treated as null hypotheses-of course

keeping in mind that it is only possible to test one at a time.

The confidence interval can as well be written for $\hat{\sigma}_o^2$ or $\hat{V}^T P \hat{V}$ as follows:

$$\left[\chi^2_{p_1} \frac{\sigma_o^2}{\nu} \leq \hat{\sigma}_o^2 \leq \chi^2_{p_2} \frac{\sigma_o^2}{\nu} \right] \qquad 8\text{-}4$$

$$\left[\chi^2_{p_1} \sigma_o^2 \leq \hat{V}^T P \hat{V} \leq \chi^2_{p_2} \sigma_o^2 \right] .$$

8.3 EXAMINATION OF THE RATIO OF TWO VARIANCE FACTORS

The ratio of two variance factors $(\sigma_o^2)_2/(\sigma_o^2)_1$ can be examined in terms of two sets of estimated residuals: (a) $\hat{V}_1$ with a $n_1 x n_1$ weight matrix

$$P_1 = (\sigma_o^2)_1 \Sigma_{L_1}^{-1} ,$$

where Σ_{L_1} and $(\sigma_o^2)_1$ are the covariance matrix and variance factor respectively for measurement set one, (b) $\hat{V}_2$ with a $n_2 x n_2$ weight matrix

$$P_2 = (\sigma_o^2)_2 \Sigma_{L_2}^{-1} ,$$

where Σ_{L_2} and $(\sigma_o^2)_2$ are the covariance matrix and variance factor, respectively for measurement set two.

Recall from section 5.11 for the univariate case that

$$\frac{(\frac{(n_1-1)\, s_1^2}{\sigma_1^2})/(n_1-1)}{(\frac{(n_2-1)\, s_2^2}{\sigma_2^2})/(n_2-1)} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \xrightarrow{d} F(n_1 - 1, n_2 - 1) .$$

For the multivariate case in which u unknown parameters are to be estimated, the statistic of interest is

$$\frac{\frac{\nu_1 \, (\hat{\sigma}_o^2)_1}{(\sigma_o^2)_1}/\nu_1}{\frac{\nu_2 \, (\hat{\sigma}_o^2)_2}{(\sigma_o^2)_2}/\nu_2} = \frac{(\hat{\sigma}_o^2)_1 \; / \; (\sigma_o^2)_1}{(\hat{\sigma}_o^2)_2 \; / \; (\sigma_o^2)_2} \overset{d}{\rightarrow} F(\nu_1, \nu_2) \; ,$$

that is $\nu_1 = n_1 - u$ or $\nu_1 = r_1 - u$, and $\nu_2 = n_2 - u$ or $\nu_2 = r_2 - u$ for the parametric and combined cases, respectively.

The probability statement involving the above random variable is

$$\Pr \left(F_{p_1} \leq \frac{(\hat{\sigma}_o^2)_1 \; / \; (\sigma_o^2)_1}{(\hat{\sigma}_o^2)_2 \; / \; (\sigma_o^2)_2} \leq F_{p_2} \right) = \alpha \qquad 8\text{-}5$$

with associated confidence interval for $(\sigma_o^2)_2/(\sigma_o^2)_1$ being

$$\boxed{\left[F_{p_1} \frac{(\hat{\sigma}_o^2)_2}{(\hat{\sigma}_o^2)_1} \leq \frac{(\sigma_o^2)_2}{(\sigma_o^2)_1} \leq F_{p_2} \frac{(\hat{\sigma}_o^2)_2}{(\hat{\sigma}_o^2)_1} \right] .} \qquad 8\text{-}6$$

The bounds of the interval are evaluated from:

1) the computed values of the two estimated variance factors $(\hat{\sigma}_o^2)_1$ and $(\hat{\sigma}_o^2)_2$,

2) the tabulated value of F_{p_1} and F_{p_2} (Appendix B-4) corresponding to $F(\nu_1, \nu_2)$, $P_1 = \frac{1 - \alpha}{2}$ and $P_2 = \frac{1 + \alpha}{2}$.

The null hypothesis to be tested is

$$H_o: \quad \frac{(\sigma_o^2)_2}{(\sigma_o^2)_1} = \left[\frac{(\sigma_o^2)_2}{(\sigma_o^2)_1} \right]_H \; .$$

8.4 EXAMINATION OF DEVIATIONS FROM THE ESTIMATED SOLUTION VECTOR $\hat{X}$ WHEN THE VARIANCE FACTOR IS KNOWN

We can test deviations from the least squares estimate $\hat{X}$ of the

parameters $\underset{\sim}{X}$. These deviations are represented as the difference between the two uxl vectors, $\hat{X}$-X. The quadratic form (see Sections 3.1.6 and 4.10).

$$(\hat{X}-X)^T \Sigma_{\hat{X}}^{-1} (\hat{X}-X) \overset{d}{\rightarrow} \chi^2 (u) .$$

Note that the variance factor σ_o^2 is assumed known since

$$\Sigma_{\hat{X}} = \sigma_o^2 Q_{\hat{X}} ,$$

where $Q_{\hat{X}}$ is the weight coefficient matrix computed from the design and weight matrices. In the next section, we will introduce a test where σ_o^2 need not be known, that is its estimate $\hat{\sigma}_o^2$ is used.

The probability statement is

$$\Pr (0 \leq (\hat{X}-X)^T \Sigma_{\hat{X}}^{-1} (\hat{X}-X) \leq \chi^2_p) = \alpha$$

The associated confidence region is simply the argument of the probability statement whose bound is the tabulated value χ^2_p (Appendix B-2) corresponding to $\chi^2(u)$ and $p = \alpha$.

The null hypothesis is

$$H_o: \quad X = X_H ,$$

that is if the computed value

$$(\hat{X}-X_H)^T \Sigma_{\hat{X}}^{-1} (\hat{X}-X_H) > \chi^2_p , \qquad 8\text{-}7$$

then the null hypothesis is rejected.

8.5 EXAMINATION OF DEVIATIONS FROM THE ESTIMATED SOLUTION VECTOR $\hat{X}$ WHEN THE VARIANCE FACTOR IS NOT KNOWN

The estimated solution vector $\hat{X}$ can be examined even when the variance factor σ_o^2 is not known, that is only the relative variances

of the measurements need be known and not the absolute variances. We will use the relations:

$$\Sigma_{\hat{X}} = \sigma_o^2 \, Q_{\hat{X}} \text{ and } \hat{\Sigma}_{\hat{X}} = \hat{\sigma}_o^2 \, Q_{\hat{X}}$$

below.

The ratio of two chi-squared random variables divided by their respective degrees of freedom defines the F-statistic we need, namely

$$\frac{(\hat{X}-X)^T \Sigma_{\hat{X}}^{-1} (\hat{X}-X)/u}{\nu\left(\frac{\hat{\sigma}_o^2}{\sigma_o^2}\right)/\nu} \xrightarrow{d} \frac{\chi^2(u)/u}{\chi^2(\nu)/\nu} \xrightarrow{d} F(u, \nu)$$

and after simplifying the first ratio we get

$$\frac{(\hat{X}-X)^T\left(1/\sigma_o^2\right) Q_{\hat{X}}^{-1} (\hat{X}-X)/u}{\left(\frac{\hat{\sigma}_o^2}{\sigma_o^2}\right)} = \frac{(\hat{X}-X)^T \quad Q_{\hat{X}}^{-1} (\hat{X}-X)}{u} \frac{\sigma_o^2}{\hat{\sigma}_o^2} \frac{1}{\sigma_o^2}$$

$$= \frac{(\hat{X}-X)^T \hat{\Sigma}_{\hat{X}}^{-1} (\hat{X}-X)}{u} \xrightarrow{d} F(u, \nu) \quad .$$

The probability statement involving the above random variable is

$$\Pr\left(0 \leq \frac{(\hat{X}-X)^T \hat{\Sigma}_{\hat{X}}^{-1} (\hat{X}-X)}{u} \leq F_p\right) = \alpha \qquad 8\text{-}8$$

while the associated confidence region is given by

$$\boxed{\left[0 \leq \frac{(\hat{X}-X)^T \hat{\Sigma}_{\hat{X}}^{-1} (\hat{X}-X)}{u} \leq F_p\right] .} \qquad 8\text{-}9$$

The bound of the region is computed by table look-up of F_p (Appendix B-4) corresponding to $F(u,\nu)$ and $p = \alpha$. The limits of this confidence region are given by the equation of the hyper-ellipsoid,

$$\boxed{(\hat{X}-X)^T \hat{\Sigma}_{\hat{X}}^{-1} (\hat{X}-X) = u \, F_p \, ,} \qquad 8\text{-}10$$

where $u F_p$ is the constant of the hyper-ellipsoid. Note $1-\alpha$ is the probability that this region does not include the true value X.

If the origin of coordinate system is translated to correspond to the position described by the vector $\hat{X}$, then the above equation becomes

$$X^T \hat{\Sigma}_{\hat{X}}^{-1} X = u F_p . \qquad 8\text{-}11$$

In two dimensions u = 2 and

$$\underset{1x2}{X^T} \quad \underset{2x2}{\hat{\Sigma}_{\hat{X}}^{-1}} \quad \underset{2x1}{X} = 2 F_p$$

$$[x_1 \; x_2] \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{21} & \hat{\sigma}_2^2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2 F_p \qquad 8\text{-}12$$

is the equation of an ellipse.

Similarly in three dimensions

$$\underset{1x3}{X^T} \quad \underset{3x3}{\hat{\Sigma}_{\hat{X}}^{-1}} \quad \underset{3x1}{X} = 3 F_p$$

$$[x_1 \; x_2 \; x_3] \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \hat{\sigma}_{13} \\ \hat{\sigma}_{21} & \hat{\sigma}_2^2 & \hat{\sigma}_{23} \\ \hat{\sigma}_{31} & \hat{\sigma}_{32} & \hat{\sigma}_3^2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 3 F_p \qquad 8\text{-}13$$

is the equation of a tri-axial ellipsoid. Note that in the above two examples, the equations will contain cross product terms since the off diagonal elements are non zero. An equation without cross product terms can be found by rotating the coordinate system through a special angle θ, where θ is computed from the components of one of the eigen vectors [Wells 1971]. The eigen vectors give the direction of maximum and minimum variances, the latter are the eigen values. For example after performing the eigen value problem on the two dimensional case

above, the equation for the limits of the confidence region is given by the equation

$$[y_1\ y_2] \begin{bmatrix} \hat{\sigma}^2_{max} & 0 \\ 0 & \hat{\sigma}^2_{min} \end{bmatrix}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 2F_p \ . \qquad 8\text{-}14$$

where (y_1, y_2) lie along the axes of the rotated coordinate system defined by the eigenvectors of $\hat{\Sigma}_{\hat{X}}$.

To summarize the results for this confidence region, first the origin of the coordinate system was translated to the position given by the vector $\hat{X}$, and secondly this coordinate system was rotated through an angle θ, given by the eigen vectors. Tests of hypothesis are made by considering the null hypothesis

$$H_o: \quad X = X_H \ ,$$

that is, if the computed value

$$\frac{(\hat{X}-X_H)^T \ \hat{\Sigma}^{-1}_{\hat{X}} \ (\hat{X}-X_H)}{u} > F_p \ ,$$

then the null hypothesis is rejected.

8.6 SUMMARY OF MULTIVARIATE CONFIDENCE REGIONS

Table 8-1 summarizes the multivariate confidence regions discussed in this Chapter. The first column of the table gives the quantity examined, the second indicates whether the variance factor is assumed known, the third gives the statistic or random variable upon which the confidence region is based, and finally the fourth column gives the confidence region itself.

Table 8-1. Summary of Multivariate Confidence Regions

Quantity Examined	σ_o^2 Known	Statistic	Confidence Region
variance factor σ_o^2 (Section 8.2)	under hypothesis	$\hat{V}^T \Sigma_{LL}^{-1} \hat{V} \xrightarrow{d} \chi^2(\nu)$	$\left[\frac{\nu\hat{\sigma}_o^2}{\chi^2_{p_2}} \leq \sigma_o^2 \leq \frac{\nu\hat{\sigma}_o^2}{\chi^2_{p_1}} \right]$
ratio of two variances $(\sigma_o^2)_2/(\sigma_o^2)_1$, (Section 8.3)	ratio under hypothesis	$\frac{(\hat{\sigma}_o^2)_1/(\sigma_o^2)_1}{(\hat{\sigma}_o^2)_2/(\sigma_o^2)_2} \xrightarrow{d} F(\nu_1, \nu_2)$	$\left[F_{p_1} \frac{(\hat{\sigma}_o^2)_2}{(\hat{\sigma}_o^2)_1} \leq \frac{(\sigma_o^2)_2}{(\sigma_o^2)_1} \leq F_{p_2} \frac{(\hat{\sigma}_o^2)_2}{(\hat{\sigma}_o^2)_1}\right]$
deviations from estimated solution vector $\hat{X}$ (Section 8.4)	yes	$(\hat{X}-X)^T \Sigma_{\hat{X}\hat{X}}^{-1} (\hat{X}-X) \xrightarrow{d} \chi^2(u)$	$[0 \leq (\hat{X}-X_H)^T \Sigma_{\hat{X}\hat{X}}^{-1} (\hat{X}-X_H) \leq \chi^2_p]$
deviations from estimated solution vector $\hat{X}$ (Section 8.5)	no	$\frac{(\hat{X}-X)^T \hat{\Sigma}_{\hat{X}\hat{X}}^{-1}(\hat{X}-X)}{u} \xrightarrow{d} F(u,\nu)$	$\left[0 \leq \frac{(\hat{X}-X_H)^T \hat{\Sigma}_{\hat{X}\hat{X}}^{-1} (\hat{X}-X_H)}{u} \leq F_p\right]$

9. PARTITIONING THE MATHEMATICAL MODEL

Not all problems in least squares estimation can be conveniently represented by the combined method mathematical model of Chapter 7

$$F(\bar{X}, \bar{L}) = 0 \ .$$

The additions to this model which can and have been made are innumerable. These "additions" are in fact really different schemes for partitioning the above mathematical model. In this Chapter we will consider four of these partitioning schemes, and subject each to the three steps of Chapter 7: Linearization, formation of the normal equations, and derivations of expressions for estimators.

We will illustrate the use of each of these four additions by considering their application in positioning by satellite. We will assume that observations L have been made on satellites from one or more ground stations by some means which we need not specify here. These observations are related both to the ground station coordinates and to the satellite coordinates, which together make up the elements of X.

9.1 ELIMINATION OF "NUISANCE" PARAMETERS

If we are not particularly interested in the satellite coordinates,

except for their role in determining the ground station coordinates, we refer to them as "nuisance" parameters. We want to eliminate them from the solution, so we partition X into the ground station coordinates, denoted by X_1, and the satellite coordinates, denoted by X_2. Then our mathematical model is

$$F(\bar{X}_1, \bar{X}_2, \bar{L}) = 0 \quad , \tag{9.1}$$

where

$$\bar{X}_1 = X_1^\circ + X_1$$

$$\bar{X}_2 = X_2^\circ + X_2$$

$$\bar{L} = L + V \quad ,$$

and the observations L have weight matrix

$$P = \sigma_o^2 \, \Sigma_L^{-1} \quad .$$

We linearize by replacing $F(\bar{X}_1, \bar{X}_2, \bar{L})$ by its Taylor's series linear approximation, expanded about the initial approximations X_1° and X_2° and the observed values L.

$$F(\bar{X}_1,\bar{X}_2,\bar{L}) = F(X_1^\circ,X_2^\circ,L) + \left.\frac{\partial F}{\partial \bar{X}_1}\right|_{X_1^\circ,X_2^\circ,L} X_1 + \left.\frac{\partial F}{\partial \bar{X}_2}\right|_{X_1^\circ,X_2^\circ,L} X_2 + \left.\frac{\partial F}{\partial \bar{L}}\right|_{X_1^\circ,X_2^\circ,L} V = 0$$

or

$$W + A_1 X_1 + A_2 X_2 + B\, V = 0 \quad , \tag{9.2}$$

where the misclosure vector $W = F(X_1^\circ, X_2^\circ, L)$ and the design matrices

$$A_1 = \left.\frac{\partial F}{\partial \bar{X}_1}\right|_{X_1^\circ,X_2^\circ,L} , \; A_2 = \left.\frac{\partial F}{\partial \bar{X}_2}\right|_{X_1^\circ,X_2^\circ,L} , \text{ and } B = \left.\frac{\partial F}{\partial \bar{L}}\right|_{X_1^\circ,X_2^\circ,L} .$$

This mathematical model is equivalent to partitioning the A matrix

and X vector in the linearized combined method mathematical model

$$W + A\,X + B\,V = 0$$

such that $A = [A_1 \vdots A_2]$, $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and $A_1 X_1$ and $A_2 X_2$ are conformable for multiplication.

The normal equations are derived by setting

$$\hat{V}^T P \hat{V} = \text{minimum} \tag{9.3}$$

under the constraint

$$W + A_1\hat{X}_1 + A_2\hat{X}_2 + B\hat{V} = 0 \ . \tag{9.4}$$

The variation function is

$$\phi = \hat{V}^T P \hat{V} + 2\hat{K}^T (W + A_1\hat{X}_1 + A_2\hat{X}_2 + B\hat{V}) \ ,$$

where $\hat{K}$ is the estimator for the vector of Lagrange multipliers. Setting the derivatives of ϕ to zero we have:

$$\frac{\partial\phi}{\partial\hat{V}} = 2\hat{V}^T P + 2\hat{K}^T B = 0$$

or

$$P\hat{V} + B^T\hat{K} = 0 \ , \tag{9.5}$$

$$\frac{\partial\phi}{\partial\hat{X}_1} = 2\hat{K}^T A_1 = 0$$

or

$$A_1^T\hat{K} = 0 \ , \tag{9.6}$$

$$\frac{\partial\phi}{\partial\hat{X}_2} = 2\hat{K}^T A_2 = 0$$

or

$$A_2^T\hat{K} = 0 \ . \tag{9.7}$$

The normal equations are

$$\begin{bmatrix} P & B^T & 0 & 0 \\ B & 0 & A_2 & A_1 \\ 0 & A_2^T & 0 & 0 \\ 0 & A_1^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{K} \\ \hat{X}_2 \\ \hat{X}_1 \end{bmatrix} + \begin{bmatrix} 0 \\ W \\ 0 \\ 0 \end{bmatrix} = 0 , \qquad (9.8)$$

which can again be seen to be a partitioned version of the normal equations for the combined method

$$\begin{bmatrix} P & B^T & 0 \\ B & 0 & A \\ 0 & A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{K} \\ \hat{X} \end{bmatrix} + \begin{bmatrix} 0 \\ W \\ 0 \end{bmatrix} = 0 .$$

Eliminating $\hat{V}$ from Equation 9.8 by the methods of Chapter 7

$$\begin{bmatrix} -BP^{-1}B^T & A_2 & A_1 \\ A_2^T & 0 & 0 \\ A_1^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{K} \\ \hat{X}_2 \\ \hat{X}_1 \end{bmatrix} + \begin{bmatrix} W \\ 0 \\ 0 \end{bmatrix} = 0 . \qquad (9.9)$$

Eliminating $\hat{K}$ from Equation 9.9

$$\begin{bmatrix} A_2^T(BP^{-1}B^T)^{-1}A_2 & A_2^T(BP^{-1}B^T)^{-1}A_1 \\ A_1^T(BP^{-1}B^T)^{-1}A_2 & A_1^T(BP^{-1}B^T)^{-1}A_1 \end{bmatrix} \begin{bmatrix} \hat{X}_2 \\ \hat{X}_1 \end{bmatrix} + \begin{bmatrix} A_2^T(BP^{-1}B^T)^{-1}W \\ A_1^T(BP^{-1}B^T)^{-1}W \end{bmatrix} = 0. \qquad (9.10)$$

Rewriting Equation 9.10 in simpler notation

$$\begin{bmatrix} N_{22} & N_{21} \\ N_{12} & N_{11} \end{bmatrix} \begin{bmatrix} \hat{X}_2 \\ \hat{X}_1 \end{bmatrix} + \begin{bmatrix} U_2 \\ U_1 \end{bmatrix} = 0 , \qquad (9.11)$$

where

$$N_{ij} = A_i^T (BP^{-1}B^T)^{-1} A_j$$

$$U_i = A_i^T (BP^{-1}B^T)^{-1} W \quad .$$

Eliminating $\hat{X}_2$ from Equation 9.11

$$\hat{X}_1 = - (N_{11} - N_{12}N_{22}^{-1}N_{21})^{-1} (U_1 - N_{12}N_{22}^{-1}U_2) \quad . \tag{9.12}$$

From the first of Equations 9.11

$$N_{22} \hat{X}_2 + N_{21} \hat{X}_1 + U_2 = 0$$

or

$$\hat{X}_2 = - N_{22}^{-1} (N_{21} \hat{X}_1 + U_2) \quad . \tag{9.13}$$

From the first of equations 9.9

$$- B P^{-1} B^T \hat{K} + A_2 \hat{X}_2 + A_1 \hat{X}_1 + W = 0$$

or

$$\hat{K} = (BP^{-1} B^T)^{-1} (A_2\hat{X}_2 + A_1\hat{X}_1 + W) \tag{9.14}$$

From the first of equations 9.8

$$P \hat{V} + B^T \hat{K} = 0$$

or

$$\hat{V} = - P^{-1} B^T \hat{K} \quad . \tag{9.15}$$

The estimators for the total solution vectors are given by

$$\hat{\bar{X}}_1 = X_1^o + \hat{X}_1$$

$$\hat{\bar{X}}_2 = X_2^o + \hat{X}_2 \quad .$$

9.2 ADDITIONAL OBSERVATIONS

If both ground and satellite coordinates are of interest then the mathematical model is again

$$F(\bar{X}, \bar{L}) = 0 \quad .$$

However let us assume that we have two sets of observations from the same ground stations. These may be separated by time ("old" and "new" observations) or by technique (two different kinds of observations). The mathematical model now is

$$\begin{aligned} F_1(\bar{X}, \bar{L}_1) &= 0 \\ F_2(\bar{X}, \bar{L}_2) &= 0 \quad , \end{aligned} \tag{9.16}$$

where

$$\begin{aligned} \bar{X} &= X^\circ + X \\ \bar{L}_1 &= L_1 + V_1 \\ \bar{L}_2 &= L_2 + V_2 \quad , \end{aligned}$$

and the observations L_1 have weight matrix $P_1 = \sigma_o^2 \, \Sigma_{L_1}^{-1}$ and L_2 have $P_2 = \sigma_o^2 \, \Sigma_{L_2}^{-1}$.

We linearize by replacing the functions $F_1(\bar{X}, \bar{L}_1)$ and $F_2(\bar{X}, \bar{L}_2)$ by their Taylor's series linear approximations, expanded about the initial approximation X° and the observed values L_1 and L_2

$$F_1(\bar{X}, \bar{L}_1) = F_1(X^\circ, L_1) + \left.\frac{\partial F_1}{\partial \bar{X}}\right|_{X^\circ, L_1} X + \left.\frac{\partial F_1}{\partial \bar{L}_1}\right|_{X^\circ, L_1} V_1 = 0$$

$$F_2(\bar{X}, \bar{L}_2) = F_2(X^\circ, L_2) + \left.\frac{\partial F_2}{\partial \bar{X}}\right|_{X^\circ, L_2} X + \left.\frac{\partial F_2}{\partial \bar{L}_2}\right|_{X^\circ, L_2} V_2 = 0 \ ,$$

or

$$\begin{aligned} W_1 + A_1 X + B_1 V_1 &= 0 \\ W_2 + A_2 X + B_2 V_2 &= 0 \ , \end{aligned} \tag{9.17}$$

where the misclosure vectors $W_1 = F_1(X^\circ, L_1)$ and $W_2 = F_2(X^\circ, L_2)$ and the design matrices

$$A_1 = \left.\frac{\partial F_1}{\partial \bar{X}}\right|_{X^\circ, L_1} , \ A_2 = \left.\frac{\partial F_2}{\partial \bar{X}}\right|_{X^\circ, L_2} , \ B_1 = \left.\frac{\partial F_1}{\partial \bar{L}_1}\right|_{X^\circ, L_1} , \text{ and } B_2 = \left.\frac{\partial F_2}{\partial \bar{L}_2}\right|_{X^\circ, L_2} .$$

This mathematical model is equivalent to partitioning the A and B matrices in the mathematical model

$$W + AX + BV = 0 \ ,$$

so that

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \quad \text{and}$$

$$B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix}$$

and V, W, and P must also be partitioned, giving

$$\begin{bmatrix} W_1 \\ W_2 \end{bmatrix} + \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} X + \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0 \ , \ P = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} ,$$

which is the same as Equations 9.17.

The normal equations are derived by setting

$$\hat{V}^T P \hat{V} = [\hat{V}_1^T \; \hat{V}_2^T] \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} \begin{bmatrix} \hat{V}_1 \\ \hat{V}_2 \end{bmatrix} = \hat{V}_1^T P_1 \hat{V}_1 + \hat{V}_2^T P_2 \hat{V}_2 = \text{minimum} \qquad (9.18)$$

under the constraints

$$\begin{aligned} W_1 + A_1 \hat{X} + B_1 \hat{V}_1 &= 0 \\ W_2 + A_2 \hat{X} + B_2 \hat{V}_2 &= 0 \quad . \end{aligned} \qquad (9.19)$$

The variation function is

$$\phi = \hat{V}_1^T P_1 \hat{V}_1 + \hat{V}_2^T P_2 \hat{V}_2 + 2\hat{K}_1^T (W_1 + A_1 \hat{X} + B_1 \hat{V}_1) + 2\hat{K}_2^T (W_2 + A_2 \hat{X} + B_2 \hat{V}_2) \; ,$$

where $\hat{K} = \begin{bmatrix} \hat{K}_1 \\ \hat{K}_2 \end{bmatrix}$ is the estimator for the Lagrange multipliers.

Then

$$\frac{\partial \phi}{\partial \hat{V}_1} = 2\hat{V}_1^T P_1 + 2\hat{K}_1^T B_1 = 0$$

or

$$P_1 \hat{V}_1 + B_1^T \hat{K}_1 = 0 \; , \qquad (9.20)$$

$$\frac{\partial \phi}{\partial \hat{V}_2} = 2\hat{V}_2^T P_2 + 2\hat{K}_2^T B_2 = 0$$

or

$$P_2 \hat{V}_2 + B_2^T \hat{K}_2 = 0 \; , \qquad (9.21)$$

$$\frac{\partial \phi}{\partial \hat{X}} = 2\hat{K}_1^T A_1 + 2\hat{K}_2^T A_2 = 0$$

or

$$A_1^T\hat{K}_1 + A_2^T\hat{K}_2 = 0 \ . \tag{9.22}$$

The normal equations are

$$\begin{bmatrix} P_1 & 0 & B_1^T & 0 & 0 \\ 0 & P_2 & 0 & B_2^T & 0 \\ B_1 & 0 & 0 & 0 & A_1 \\ 0 & B_2 & 0 & 0 & A_2 \\ 0 & 0 & A_1^T & A_2^T & 0 \end{bmatrix} \begin{bmatrix} \hat{V}_1 \\ \hat{V}_2 \\ \hat{K}_1 \\ \hat{K}_2 \\ \hat{X} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ W_1 \\ W_2 \\ 0 \end{bmatrix} = 0 \ , \tag{9.23}$$

which are a partitioned version of the normal equations for the combined method of Chapter 7.

Eliminating $\hat{V}_1$ from Equation 9.23

$$\begin{bmatrix} P_2 & 0 & B_2^T & 0 \\ 0 & -B_1P_1^{-1}B_1^T & 0 & A_1 \\ B_2 & 0 & 0 & A_2 \\ 0 & A_1^T & A_2^T & 0 \end{bmatrix} \begin{bmatrix} \hat{V}_2 \\ \hat{K}_1 \\ \hat{K}_2 \\ \hat{X} \end{bmatrix} + \begin{bmatrix} 0 \\ W_1 \\ W_2 \\ 0 \end{bmatrix} = 0 \ . \tag{9.24}$$

Eliminating $\hat{V}_2$ from Equation 9.24

$$\begin{bmatrix} -B_1P_1^{-1}B_1^T & 0 & A_1 \\ 0 & -B_2P_2^{-1}B_2^T & A_2 \\ A_1^T & A_2^T & 0 \end{bmatrix} \begin{bmatrix} \hat{K}_1 \\ \hat{K}_2 \\ \hat{X} \end{bmatrix} + \begin{bmatrix} W_1 \\ W_2 \\ 0 \end{bmatrix} = 0 \ . \tag{9.25}$$

Eliminating $\hat{K}_1$ from Equation 9.25

$$\begin{bmatrix} -B_2P_2^{-1}B_2^T & A_2 \\ A_2^T & A_1^T(B_1P_1^{-1}B_1^T)^{-1}A_1 \end{bmatrix} \begin{bmatrix} \hat{K}_2 \\ \hat{X} \end{bmatrix} + \begin{bmatrix} W_2 \\ A_1^T(B_1P_1^{-1}B_1^T)^{-1}W_1 \end{bmatrix} = 0 \ . \tag{9.26}$$

Eliminating $\hat{K}_2$ from Equation 9.26

$$\left(A_1^T(B_1P_1^{-1}B_1^T)^{-1}A_1 + A_2^T(B_2P_2^{-1}B_2^T)^{-1}A_2\right)\hat{X} + A_1^T(B_1P_1^{-1}B_1^T)^{-1}W_1 + A_2^T(B_2P_2^{-1}B_2^T)^{-1}W_2 = 0 \ ,$$

or

$$\hat{X} = -\left(A_1^T(B_1P_1^{-1}B_1^T)^{-1}A_1 + A_2^T(B_2P_2^{-1}B_2^T)^{-1}A_2\right)^{-1}\left(A_1^T(B_1P_1^{-1}B_1^T)^{-1}W_1 + A_2^T(B_2P_2^{-1}B_2^T)^{-1}W_2\right) \ . \tag{9.27}$$

From the first Equation of 9.26

$$\hat{K}_2 = (B_2P_2^{-1}B_2^T)^{-1}\ (A_2\hat{X} + W_2) \ . \tag{9.28}$$

From the first Equation of 9.25

$$\hat{K}_1 = (B_1P_1^{-1}B_1^T)^{-1}\ (A_1\hat{X} + W_1) \ . \tag{9.29}$$

From the first Equation of 9.23

$$\hat{V}_2 = -P_2^{-1}B_2^T\hat{K}_2 \ . \tag{9.30}$$

From the first Equation of 9.23

$$\hat{V}_1 = -P_1^{-1}B_1^T\hat{K}_1 \ . \tag{9.31}$$

The estimator for the total solution vector is

$$\hat{\bar{X}} = X^\circ + \hat{X} \quad . \tag{9.32}$$

9.3 ADDITIONAL CONSTRAINTS BETWEEN UNKNOWN PARAMETERS

Assume we have knowledge of some relationship between the unknown parameters <u>other</u> than that contained in the mathematical model

$$F(\bar{X}, \bar{L}) = 0$$

(for example we know that the satellite coordinates all fall on an elliptical trajectory). If this relationship is represented by a mathematical model

$$F(\bar{X}) = 0$$

then the complete relationship between $\bar{X}$ and $\bar{L}$ is specified

$$\begin{aligned} F_1(\bar{X}, \bar{L}) &= 0 \\ F_2(\bar{X}) &= 0 \quad , \end{aligned} \tag{9.33}$$

where

$$\bar{X} = X^\circ + X$$

$$\bar{L} = L + V$$

and the observations L have weight matrix $P = \sigma_o^2 \, \Sigma_L^{-1}$.

We linearize both $F_1(\bar{X}, \bar{L})$ and $F_2(\bar{X})$ by replacing them by their Taylor's series linear approximations, expanded about the initial approximation X° and the observed values L

$$F_1(\bar{X}, \bar{L}) = F_1(X^\circ, L) + \left.\frac{\partial F_1}{\partial \bar{X}}\right|_{X^\circ, L} X + \left.\frac{\partial F_1}{\partial \bar{L}}\right|_{X^\circ, L} V = 0$$

$$F_2(\bar{X}) = F_2(X^\circ) + \left.\frac{\partial F_2}{\partial \bar{X}}\right|_{X^\circ} X = 0 \quad ,$$

or

$$\begin{aligned} W_1 + A_1 X + BV &= 0 \\ W_2 + A_2 X &= 0 \ , \end{aligned} \tag{9.34}$$

where the misclosure vectors $W_1 = F_1(X^\circ, L)$ and $W_2 = F_2(X^\circ)$ and the design matrices

$$A_1 = \left.\frac{\partial F_1}{\partial \bar{X}}\right|_{X^\circ,L} , \quad A_2 = \left.\frac{\partial F_2}{\partial \bar{X}}\right|_{X^\circ} , \text{ and } B = \left.\frac{\partial F_1}{\partial \bar{L}}\right|_{X^\circ,L} .$$

This mathematical model is equivalent to partitioning the A matrix in the mathematical model

$$W + AX + BV = 0$$

so that

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \text{ and setting } B = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}$$

and W must be also partitioned, giving

$$\begin{bmatrix} W_1 \\ W_2 \end{bmatrix} + \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} X + \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} V = 0 \quad ,$$

which is the same as Equations 9.34

The normal equations are derived by setting

$$\hat{V}^T P \hat{V} = \text{minimum} \tag{9.35}$$

under the constraints

$$W_1 + A_1\hat{X} + B\hat{V} = 0$$
$$W_2 + A_2\hat{X} = 0 \quad . \tag{9.36}$$

The variation function is

$$\phi = \hat{V}^T P \hat{V} + 2\hat{K}_1^T(W_1 + A_1\hat{X} + B\hat{V}) + 2\hat{K}_2^T(W_2 + A_2\hat{X}) \quad ,$$

where $\hat{K} = \begin{bmatrix} \hat{K}_1 \\ \hat{K}_2 \end{bmatrix}$ is the estimator for the Lagrange multipliers.

Then

$$\frac{\partial \phi}{\partial \hat{V}} = 2\hat{V}^T P + 2\hat{K}_1^T B = 0$$

or

$$P\hat{V} + B^T\hat{K}_1 = 0 \quad , \tag{9.37}$$

$$\frac{\partial \phi}{\partial \hat{X}} = 2\hat{K}_1^T A_1 + 2\hat{K}_2^T A_2 = 0$$

or

$$A_1^T\hat{K}_1 + A_2^T\hat{K}_2 = 0 \quad . \tag{9.38}$$

The normal equations are

$$\begin{bmatrix} P & B^T & 0 & 0 \\ B & 0 & A_1 & 0 \\ 0 & A_1^T & 0 & A_2^T \\ 0 & 0 & A_2 & 0 \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{K}_1 \\ \hat{X} \\ \hat{K}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ W_1 \\ 0 \\ W_2 \end{bmatrix} = 0 \quad , \tag{9.39}$$

which are a partitioned version of the normal equation for the combined

method of Chapter 7.

Eliminating $\hat{V}$ from Equation 9.39

$$\begin{bmatrix} -BP^{-1}B^T & A_1 & 0 \\ A_1^T & 0 & A_2^T \\ 0 & A_2 & 0 \end{bmatrix} \begin{bmatrix} \hat{K}_1 \\ \hat{X} \\ \hat{K}_2 \end{bmatrix} + \begin{bmatrix} W_1 \\ 0 \\ W_2 \end{bmatrix} = 0 \ . \tag{9.40}$$

Eliminating $\hat{K}_1$ from Equation 9.40

$$\begin{bmatrix} A_1^T(BP^{-1}B^T)^{-1}A_1 & A_2^T \\ A_2 & 0 \end{bmatrix} \begin{bmatrix} \hat{X} \\ \hat{K}_2 \end{bmatrix} + \begin{bmatrix} A_1^T(BP^{-1}B^T)^{-1}W_1 \\ W_2 \end{bmatrix} = 0 \ . \tag{9.41}$$

Eliminating $\hat{X}$ from Equation 9.41

$$- A_2(A_1^T(BP^{-1}B^T)^{-1}A_1)^{-1}A_2^T\hat{K}_2 +$$

$$(W_2 - A_2(A_1^T(BP^{-1}B^T)^{-1}A_1)^{-1}A_1^T(BP^{-1}B^T)^{-1}W_1) = 0 \ ,$$

or

$$\hat{K}_2 = \left(A_2(A_1^T(BP^{-1}B^T)^{-1}A_1)^{-1}A_2^T\right)^{-1}\left(W_2 - A_2(A_1^T(BP^{-1}B^T)^{-1}A_1)^{-1}A_1^T)BP^{-1}B^T)^{-1}W_1\right). \tag{9.42}$$

From the first Equation in 9.41

$$\hat{X} = - \left(A_1^T(BP^{-1}B^T)^{-1}A_1\right)^{-1}(A_2^T\hat{K}_2 + A_1^T(BP^{-1}B^T)^{-1}W_1) \ . \tag{9.43}$$

From the first Equation in 9.40

$$\hat{K}_1 = (BP^{-1}B^T)^{-1}(A_1\hat{X} + W_1) \ . \tag{9.44}$$

From the first Equation in 9.39

$$\hat{V} = - P^{-1}B^T\hat{K}_1 \ . \tag{9.45}$$

The estimator for the total solution vector is

$$\hat{\bar{X}} = X^o + \hat{X} \ . \tag{9.46}$$

9.4 WEIGHTING UNKNOWN PARAMETERS

One tacit assumption which we have made all through these notes is that the mathematical model itself is complete - that is it perfectly represents the relationship between the unknown parameters and the observations. In many simpler and well established applications of least squares estimation this is no doubt true. However for very complex relationships, and particularly in new applications where the relationships may not yet be fully known, this is not a valid assumption to make. For example in the previous section we assumed that satellites follow an elliptical trajectory. In fact the trajectory is only approximated by an ellipse, so that in this case our mathematical model was incomplete.

One way in which to incorporate this uncertainty into the least squares estimation process is to treat the unknown parameters not as completely unknown, but as partially known. That is our "initial approximation" X^o is no longer the arbitrary thing it used to be, but is now a "preliminary estimate". And rather than having infinite a priori variances, we assign a priori variances to this preliminary estimate X^o, which reflect the uncertainties we feel might exist in the mathematical model. What this accomplishes in effect is to assert that the resulting least squares estimates for the unknown parameters must fall within the limits specified by these variances of the preliminary estimate. In other words the unknowns have become

"pseudo-measurements".

In practice this means that our mathematical model

$$F(\bar{X}, \bar{L}) = 0 \ , \tag{9.47}$$

where

$$\bar{X} = X^{\circ} + X$$

$$\bar{L} = L \ + V$$

will have a linearized form

$$W + AX + BV = 0 \ , \tag{9.48}$$

but that the "residual vector" is now

$$\begin{bmatrix} V \\ X \end{bmatrix}$$

and the weight matrix is now

$$P = \begin{bmatrix} P_V & 0 \\ 0 & P_X \end{bmatrix}$$

where

$$P_V = \sigma_o^2 \, \Sigma_L^{-1} \ ,$$

and

$$P_X = \sigma_o^2 \, \Sigma_{X^{\circ}}^{-1}$$

($\Sigma_{X^{\circ}}$ being the a priori covariance matrix of the unknowns).

This mathematical model is equivalent to partitioning the B matrix in the condition method mathematical model

$$W + BV = 0$$

so that

$$B = [B \ \ A]$$

and

$$W + [B \quad A] \begin{bmatrix} V \\ X \end{bmatrix} = 0 \quad , \quad P = \begin{bmatrix} P_V & 0 \\ 0 & P_X \end{bmatrix} .$$

The normal equations are derived by setting

$$\begin{bmatrix} \hat{V} \\ \hat{X} \end{bmatrix}^T P \begin{bmatrix} \hat{V} \\ \hat{X} \end{bmatrix} = \hat{V}^T P_V \hat{V} + \hat{X}^T P_X \hat{X} = \text{minimum} \tag{9.49}$$

under the constraint

$$W + A\hat{X} + B\hat{V} = 0 \quad . \tag{9.50}$$

The variation function is

$$\phi = \hat{V}^T P_V \hat{V} + \hat{X}^T P_X \hat{X} + 2\hat{K}^T (W + A\hat{X} + B\hat{V}) \ ,$$

where $\hat{K}$ is the estimator for the Lagrange multipliers.

Then

$$\frac{\partial \phi}{\partial \hat{V}} = 2\hat{V}^T P_V + 2\hat{K}^T B = 0$$

or

$$P_V \hat{V} + B^T \hat{K} = 0 \quad , \tag{9.51}$$

$$\frac{\partial \phi}{\partial \hat{X}} = 2\hat{X}^T P_X + 2\hat{K}^T A = 0$$

or

$$P_X \hat{X} + A^T \hat{K} = 0 \quad . \tag{9.52}$$

The normal equations are

$$\begin{bmatrix} P_V & 0 & B^T \\ 0 & P_X & A^T \\ B & A & 0 \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{X} \\ \hat{K} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ W \end{bmatrix} = 0 \ , \tag{9.53}$$

which are a partitioned version of the normal equation for the condition method of Chapter 7.

Eliminating $\hat{V}$ from Equation 9.53

$$\begin{bmatrix} P_X & A^T \\ A & -BP_V^{-1}B^T \end{bmatrix} \begin{bmatrix} \hat{X} \\ \hat{K} \end{bmatrix} + \begin{bmatrix} 0 \\ W \end{bmatrix} = 0 \ . \tag{9.54}$$

Eliminating $\hat{X}$ from Equation 9.54

$$- (BP_V^{-1}B^T + AP_X^{-1}A^T)\hat{K} + W = 0$$

or

$$\hat{K} = (BP_V^{-1}B^T + AP_X^{-1}A^T)^{-1}W \ . \tag{9.55}$$

From the first of Equation 9.54

$$\hat{X} = - P_X^{-1}A^T\hat{K} \ . \tag{9.56}$$

From the first of Equation 9.53

$$\hat{V} = - P_V^{-1}B^T\hat{K} \ . \tag{9.57}$$

These expressions are unsatisfactory however, because they should reduce to the simpler expressions for unweighted parameters by merely setting $P_X = 0$.

However in that case P_X is singular and P_X^{-1} does not exist. Therefore we will reformulate the normal equations in such a way that our

expressions will not include P_X^{-1}, that is

$$\begin{bmatrix} P_V & B^T & 0 \\ B & 0 & A \\ 0 & A^T & P_X \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{K} \\ \hat{X} \end{bmatrix} + \begin{bmatrix} 0 \\ W \\ 0 \end{bmatrix} = 0 \quad . \tag{9.58}$$

Eliminating $\hat{V}$ from Equation 9.58

$$\begin{bmatrix} -BP_V^{-1}B^T & A \\ A^T & P_X \end{bmatrix} \begin{bmatrix} \hat{K} \\ \hat{X} \end{bmatrix} + \begin{bmatrix} W \\ 0 \end{bmatrix} = 0 \quad . \tag{9.59}$$

Eliminating $\hat{K}$ from Equation 9.59

$$[P_X + A^T(BP_V^{-1}B^T)^{-1}A]\hat{X} + A^T(BP_V^{-1}B^T)^{-1}W = 0 \ ,$$

or

$$\hat{X} = - \left(P_X + A^T(BP_V^{-1}B^T)^{-1}A\right)^{-1} A^T(BP_V^{-1}B^T)^{-1}W \tag{9.60}$$

and

$$\hat{K} = (BP_V^{-1}B^T)^{-1} (A\hat{X} + W) \tag{9.61}$$

$$\hat{V} = - P_V^{-1}B^T\hat{K} \tag{9.62}$$

and the final estimator for the weighted solution vector is

$$\hat{\bar{X}} = X^o + \hat{X} \quad . \tag{9.63}$$

10. STEP BY STEP LEAST SQUARES ESTIMATION

As mentioned in Chapter 1, even the largest fastest digital computer is incapable of simultaneously solving systems which may incorporate several thousand equations. In this Chapter, the problem of chopping a large system into smaller systems is discussed. We will take it for granted that any chopping scheme must yield the same final result as would have been obtained from a simultaneous solution.

Step by step least squares estimation is not a new concept [see for example Tobey 1930; Tienstra 1956; Schmid and Schmid 1965; and Kalman 1960]. It has gone under many names, some of which are differential adjustment, phased adjustment, sequential adjustment and Kalman filtering. There are differences in detail between some of these methods, but basically they all involve the derivation of expressions for the current least squares estimate in terms of the previous estimate plus a "correction" term, using the rules of matrix partitioning.

In this Chapter, we will not attempt to be exhaustive, but will derive a sequential expression for the solution vector X following Krakiwsky [1968]. We will then show this result to be equivalent to the Kalman filter equations for the case where the unknown parameters are not time variable.

10.1 SEQUENTIAL LEAST SQUARES EXPRESSIONS

We will consider the parametric mathematical model

$$F(\bar{X}) - \bar{L} = 0 \quad ,$$

which when linearized becomes

$$AX - V + W = 0 \quad .$$

Applying the least squares criterion

$$\hat{V}^T P \hat{V} = \text{minimum},$$

we obtain the normal equations

$$\begin{bmatrix} P & -I & 0 \\ -I & 0 & A \\ 0 & A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{K} \\ \hat{X} \end{bmatrix} + \begin{bmatrix} 0 \\ W \\ 0 \end{bmatrix} = 0 \qquad 10\text{-}1$$

We now partition the system into the previous set of equations (quantities subscripted k-1) and those that have just been added to obtain the current estimate (quantities subscripted k). Note that P, A, W, $\hat{V}$ and $\hat{K}$ are partitioned because there are new observations and thus new equations, however $\hat{X}$ is not partitioned because it is assumed that the new observations are related to the same parameters.

$$\left[\begin{array}{cc|cc|c} P_{k-1} & 0 & -I & 0 & 0 \\ 0 & P_k & 0 & -I & 0 \\ \hline -I & 0 & 0 & 0 & A_{k-1} \\ 0 & -I & 0 & 0 & A_k \\ \hline 0 & 0 & A_{k-1}^T & A_K^T & 0 \end{array}\right] \left[\begin{array}{c} \hat{V}_{k-1} \\ \hat{V}_k \\ \hline \hat{K}_{k-1} \\ \hat{K}_k \\ \hline \hat{X} \end{array}\right] + \left[\begin{array}{c} 0 \\ 0 \\ \hline W_{k-1} \\ W_k \\ \hline 0 \end{array}\right] = 0 \qquad 10\text{-}2$$

We will denote by $\hat{X}_{k-1}$ the estimate obtained only when the previous observations are used. We will denote by $\hat{X}_k$ the estimate obtained when

all observations are used. The problem now is to find a sequential expression for $\hat{X}_k$.

The first step is to find the previous solution by setting $P_k = A_k = W_k = 0$. The normal equations then reduce to

$$\begin{bmatrix} P_{k-1} & -I & 0 \\ -I & 0 & A_{k-1} \\ 0 & A_{k-1}^T & 0 \end{bmatrix} \begin{bmatrix} \hat{V}_{k-1} \\ \hat{K}_{k-1} \\ \hat{X}_{k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ W_{k-1} \\ 0 \end{bmatrix} = 0 \qquad 10\text{-}3$$

and the solution is

$$\hat{X}_{k-1} = -N_{k-1}^{-1} U_{k-1} \qquad 10\text{-}4a$$

$$\hat{K}_{k-1} = P_{k-1}(A_{k-1} \hat{X}_{k-1} + W_{k-1}) \qquad 10\text{-}4b$$

$$\hat{V}_{k-1} = A_{k-1} \hat{X}_{k-1} + W_{k-1} \quad , \qquad 10\text{-}4c$$

where

$$N_{k-1}^{-1} = (A_{k-1}^T P_{k-1} A_{k-1})^{-1} \qquad 10\text{-}4d$$

$$U_{k-1} = A_{k-1}^T P_{k-1} W_{k-1} \quad . \qquad 10\text{-}4e$$

Next we rearrange the current system, equation 10-2

$$\begin{bmatrix} P_{k-1} & 0 & -I & 0 & 0 \\ 0 & P_k & 0 & 0 & -I \\ -I & 0 & 0 & A_{k-1} & 0 \\ 0 & 0 & A_{k-1}^T & 0 & A_k^T \\ 0 & -I & 0 & A_k & 0 \end{bmatrix} \begin{bmatrix} \hat{V}_{k-1} \\ \hat{V}_k \\ \hat{K}_{k-1} \\ \hat{X}_k \\ \hat{K}_k \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ W_{k-1} \\ 0 \\ W_k \end{bmatrix} = 0 \qquad 10\text{-}5$$

Eliminating $\hat{V}_{k-1}$ and $\hat{V}_k$ from equation 10-5

$$\begin{bmatrix} -P_{k-1}^{-1} & A_{k-1} & 0 \\ A_{k-1}^T & 0 & A_k^T \\ 0 & A_k & -P_k^{-1} \end{bmatrix} \begin{bmatrix} \hat{K}_{k-1} \\ \hat{X}_k \\ \hat{K}_k \end{bmatrix} + \begin{bmatrix} W_{k-1} \\ 0 \\ W_k \end{bmatrix} = 0 \qquad 10\text{-}6$$

Eliminating $\hat{K}_{k-1}$ from 10-6

$$\begin{bmatrix} N_{k-1} & A_k^T \\ A_k & -P_k^{-1} \end{bmatrix} \begin{bmatrix} \hat{X}_k \\ \hat{K}_k \end{bmatrix} + \begin{bmatrix} U_{k-1} \\ W_k \end{bmatrix} = 0 \quad . \qquad 10\text{-}7$$

Eliminating $\hat{X}_k$ from 10-7

$$-(P_k^{-1} + A_k N_{k-1}^{-1} A_k^T) \hat{K}_k + W_k - A_k N_{k-1}^{-1} U_{k-1} = 0$$

but from equation 10-4a

$$-N_{k-1}^{-1} U_{k-1} = \hat{X}_{k-1}$$

Therefore

$$\boxed{\hat{K}_k = (P_k^{-1} + A_k N_{k-1}^{-1} A_k^T)^{-1}(A_k \hat{X}_{k-1} + W_k) \; .} \qquad 10\text{-}8$$

From the first of equations 10-7

$$N_{k-1} \hat{X}_k + A_k^T \hat{K}_k + U_{k-1} = 0$$

or

$$\hat{X}_k = -N_{k-1}^{-1} U_{k-1} - N_{k-1}^{-1} A_k^T \hat{K}_k$$

or

$$\boxed{\hat{X}_k = \hat{X}_{k-1} - N_{k-1}^{-1} A_k^T \hat{K}_k \; .} \qquad 10\text{-}9$$

From the second of equations 10-5

$$P_k \hat{V}_k - \hat{K}_k = 0$$

or

$$\boxed{\hat{V}_k = P_k^{-1} \hat{K}_k \; .} \qquad 10\text{-}10$$

To find a sequential expression for the covariance matrix of the estimated solution vector, we have from equations 10-8 and 10-9

$$\hat{X}_k = [C_1 \quad C_2] \begin{bmatrix} \hat{X}_{k-1} \\ W_k \end{bmatrix}$$

where

$$C_1 = I - N_{k-1}^{-1} A_k^T (P_k^{-1} + A_k N_{k-1}^{-1} A_k^T)^{-1} A_k$$

$$C_2 = - N_{k-1}^{-1} A_k^T (P_k^{-1} + A_k N_{k-1}^{-1} A_k^T)^{-1} .$$

By the covariance law

$$N_k^{-1} = \Sigma_{\hat{x}_k} = [C_1 \quad C_2] \begin{bmatrix} N_{k-1}^{-1} & 0 \\ 0 & P_k^{-1} \end{bmatrix} \begin{bmatrix} C_1^T \\ C_2^T \end{bmatrix}$$

$$= C_1 N_{k-1}^{-1} C_1^T + C_2 P_k^{-1} C_2^T \quad ,$$

where we have ignored the variance factor and set

$$N_{k-1}^{-1} = \Sigma_{\hat{x}_{k-1}}$$

$$P_k^{-1} = \Sigma_{w_k} .$$

Multiplying out

$$\boxed{N_k^{-1} = N_{k-1}^{-1} - N_{k-1}^{-1} A_k^T (P_k^{-1} + A_k N_{k-1}^{-1} A_k^T)^{-1} A_k N_{k-1}^{-1} .} \qquad 10\text{-}11$$

And lastly from equation 10-10

$$\boxed{\hat{V}_k^T P_k \hat{V}_k = \hat{K}_k^T P_k^{-1} \hat{K}_k .} \qquad 10\text{-}12$$

We will now compare equations 10-8, 10-9 and 10-11 with the Kalman Filter equations, as given for example by Sorenson [1970].

10.2 THE KALMAN FILTER EQUATIONS

For a broad class of problems in electrical engineering, usually called "optimal control problems", not only does the estimate of the "state-space vector" (analogous to the unknown vector X in these least squares estimation notes) change as new data becomes available for least squares estimation, but also the actual value of this vector itself changes with time. Therefore, in optimal control problems there are two time dependent factors - the actual value of the state-space vector is changing continuously, and new observed data is being accumulated continuously from which new estimates of the new value of the state-space vector can be made.

During the past ten years, the big news in optimal control has been the Kalman filter [Kalman 1960].

We will now review Kalman's equations (using optimal control notation) and then simplify the equations by dropping the time dependence and rewriting in our notation.

The time dependence of the state vector is expressed by the mathematical model

$$x_{k+1} = \Phi_{k+1,\,k}\, x_k + w_k \ , \qquad 10\text{-}13$$

where x is the state-space vector (solution vector),

$\Phi_{k+1,\,k}$ is the transfer function between kth and (k+1)th states (plant model),

$\{w_k\}$ is the plant white noise sequence (residual vector).

The linearized mathematical model between the state vector and the measurement data is expressed by

$$z_k = H_k\, x_k + v_k \ , \qquad 10\text{-}14$$

where z_k is the measurement vector (misclosure vector),

H_k is the design matrix,

$\{V_k\}$ is the measurement white noise sequence (residual vector).

The covariance matrices of W_k and V_k are denoted Q_k and R_k respectively. The Kalman least squared estimation problem can now be stated as follows: defining the estimate (not necessarily least squares) of the state x_i found by using all data z_0, z_1 . . . z_j as $\hat{x}_{i/j}$, the problem is

a) to find the least squares estimate $\hat{x}_{k/k}$ of the current state x_k using all data up to and including the current set z_k,

b) to express this estimate only in terms of the current measurement z_k and the previous best estimate $\hat{x}_{k-1/k-1}$,

c) to ensure that this solution is as rigorous as that obtained by processing all the data z_0, z_1 . . . z_k simultaneously.

In the absence of new data the <u>predicted estimate</u> is

$$\hat{x}_{k/k-1} = \Phi_{k,\,k-1}\,\hat{x}_{k-1/k-1} \qquad 10\text{-}15$$

The Kalman equations can now be stated as

$$\hat{x}_{k/k} = \Phi_{k,\,k-1}\,\hat{x}_{k-1/k-1} + K_k\,[z_k - H_k\,\Phi_{k,\,k-1}\,\hat{x}_{k-1/k-1}] \qquad 10\text{-}16$$

$$K_k = P_{k/k-1}\,H_k^T\,(H_k\,P_{k/k-1}\,H_k^T + R_k)^{-1} \qquad 10\text{-}17$$

$$P_{k/k-1} = \Phi_{k,\,k-1}\,P_{k-1/k-1}\,\Phi_{k,\,k-1}^T + Q_{k-1} \qquad 10\text{-}18$$

$$P_{k/k} = P_{k/k-1} - K_k\,H_k\,P_{k/k-1}\ , \qquad 10\text{-}19$$

where K_k is the "gain matrix",

$P_{k/k-1}$ is the covariance matrix of the error in the <u>predicted</u> estimate $\hat{x}_{k/k-1}$,

$P_{k/k}$ is the covariance matrix of the error in the <u>least squares</u>

estimate $\hat{x}_{k/k}$.

Dropping the time variability of the state vector means we can ignore equations 10-13, 10-15, and 10-18 above. For the rest, the notation conversion is

Kalman	These notes	Kalman	These notes
z_k	$-W_k$	$\hat{x}_{k/k}$	$\hat{x}_k$
H_k	A_k	$P_{k/k}$	N_k^{-1}
x_k	X	$P_{k/k-1}$	N_{k-1}^{-1}
v_k	$-\hat{V}_k$	K_k	no equivalent
R_k	P_k^{-1}	no equivalent	$\hat{K}_k$

Rewriting equations 10-16, 10-17 and 10-19 in our notation:

$$\hat{X}_k = \hat{X}_{k-1} - K_k (W_k + A_k \hat{X}_{k-1}) \qquad 10\text{-}20$$

$$K_k = N_{k-1}^{-1} A_k^T (A_k N_{k-1}^{-1} A_k^T + P_k^{-1})^{-1} \qquad 10\text{-}21$$

$$N_k^{-1} = N_{k-1}^{-1} - K_k A_k N_{k-1}^{-1} \qquad 10\text{-}22$$

which can be seen to be equivalent to equations 10-8, 10-9 and 10-11, although the definition of K_k is different in the two cases.

REFERENCES

Carnahan, B., H.A. Luther and J.O. Wilkes (1969). Applied Numerical Methods, Wiley, Toronto.

Gauss, K.F. (1963). Theory of Motion of the Heavenly Bodies, Dover, New York.

Hamilton, W.C. (1964). Statistics in Physical Science, Ronald Press, New York.

Hamming, R.W. (1962). Numerical Methods for Scientists and Engineers, McGraw Hill, Toronto (page 225).

Hogg, R.V. and A.T. Craig (1965). Introduction to Mathematical Statistics , McMillan, New York.

Kalman, R.E. (1960). "A New Approach to Linear Filtering and Prediction," Journal of Basic Engineering, ASME, Vol. 82D, pages 35-45, 1960.

Krakiwsky, E.J. (1968). "Sequential Least Squares Adjustment of Satellite Triangulation and Trilateration in Combination with Terrestrial Data", Reports of OSU Dept. of Geodetic Science, Report No. 114, pages 117-134.

Kendall, M.G. and W.R. Buckland (1957). A Dictionary of Statistical Terms, Oliver and Boyd, Edinburgh.

Kendall, M.G. and A. Stuary (1969). The Advanced Theory of Statistics Vol. I: Distribution Theory, 3rd. ed., Griffin, London.

Schmid, H.H. and E. Schmid (1965). "A Generalized Least Squares Solution for Hybrid Measuring Systems," The Canadian Surveyor, Vol. XIX, No. 1, pages 27-41.

Schwartz, C.R. (1967). "Optimal Filter and Least Squares Estimation", U.S. Army Map Service Geodetic Memorandum No. 1610, Washington.

Sorenson, H.W. (1970). "Least Squares Estimation: from Gauss to Kalman," IEEE Spectrum, Vol. 7 No. 6 , pages 63-68, July 1970.

Spiegel, M.R. (1962). Theory and Problems of Statistics, Schaum's Outline Series, McGraw-Hill, Toronto.

Tienstra, J.M. (1956). Theory of the Adjustment of Normally Distributed Observations , Argus, Amsterdam.

Tobey, W.M. (1930). "The Differential Adjustment of Observations", Geodetic Survey Publication No. 27, Ottawa.

Wells, D.E. (1971). "Matrices", Department of Surveying Engineering, Lecture Notes No. 15, University of New Brunswick.

APPENDIX A

NUMERICAL EXAMPLE*

A.1 Statement of the Problem:

An observer on a plane surface measures the directions to distant landmarks whose coordinates are known. The observer is close to the origin of the coordinate system. The values of the measurements made are tabulated in Figure A-1. All observations have a standard deviation of 0.1 degrees. The mathematical model relating the measurements to the unknown parameters is

$$\tan(\alpha_i + z) = \frac{y_i - y_o}{x_i - x_o}$$

where (x_o, y_o) are the unknown observers coordinates, (x_i, y_i) are the known landmark coordinates, α_i are the measured directions from observer to landmarks, and z is an unknown orientation angle between the measurement coordinate system and the coordinate system to which the landmark and observers coordinates are referred. See inset in Figure A-1.

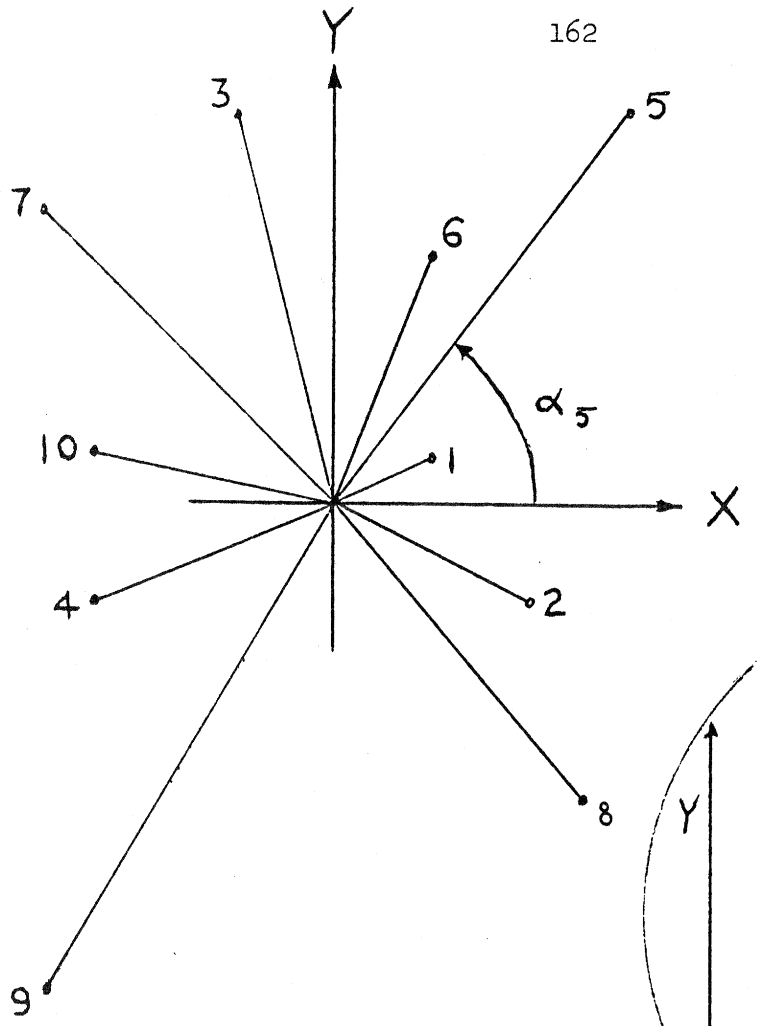
We will use units of radians and meters in our example. Then $\bar{L} = L + V$ where

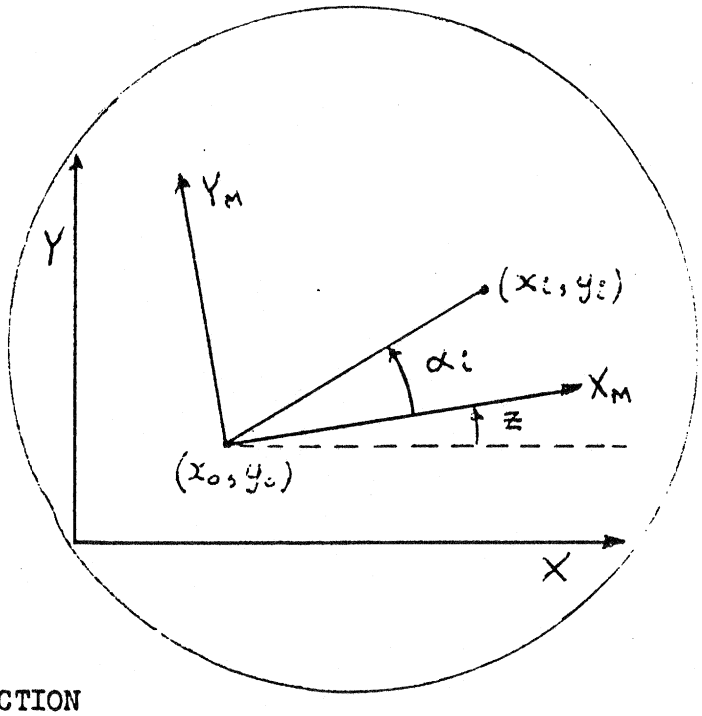
$$L = \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \alpha_{10} \end{bmatrix} = \begin{bmatrix} 26.7047 * \pi/180 \\ \cdot \\ \cdot \\ 168.6730 * \pi/180 \end{bmatrix} \text{ in radians}$$

and

---

*data taken from Bennet, J.E. and J.C. Hung (1970). "Application of Statistical Techniques to Landmark Navigation". Navigation Vol. 17, No. 4, page 349, Winter.





	LANDMARK		MEASURED DIRECTION
No.	x(km)	y(km)	(degrees)
1	2	1	26.7047
2	4	-2	-26.5795
3	-2	8	103.8340
4	-5	-2	-157.9978
5	6	8	53.1393
6	2	5	68.2511
7	-6	6	135.0673
8	5	-6	-50.3190
9	-6	-10	-120.8734
10	-5	1	168.6730

ILLUSTRATIVE EXAMPLE

Figure A-1

$$V = \begin{bmatrix} v_1 \\ \cdot \\ \cdot \\ v_{10} \end{bmatrix} \text{ where the } v_i \text{ are in radians}$$

and we will choose

$$X^o = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ so that } \bar{X} = X = \begin{bmatrix} x_o \\ y_o \\ z \end{bmatrix}$$

where (x_o, y_o) are in meters and z is in radians. Finally $\Sigma_L = \sigma^2 I$ where $\sigma^2 = (0.1 * \pi/180)^2$ in radians^2, since we have assumed all observations to have the same standard deviation (0.1 degrees).

A.2 Linearization of the Mathematical Model:

We can use either the combined or parametric method. Using the combined method, the ith equation is

$$f_i(\bar{X},\bar{L}) = \tan(\bar{\alpha}_i + \bar{z}) - \frac{y_i - \bar{y}_o}{x_i - \bar{x}_o} = 0$$

where the (x_i, y_i) are constants and thus unbarred.

After linearization this becomes $AX + BV + W = 0$ where the ith rows of A, B and W have the form

$$A_i = \left[\frac{\partial f_i}{\partial \bar{x}_o} ; \frac{\partial f_i}{\partial \bar{y}_o} ; \frac{\partial f_i}{\partial \bar{z}} \right] \Bigg|_{X^o,L} = \left[\frac{-y_i}{x_i^2} ; \frac{1}{x_i} ; \sec^2 \alpha_i \right]$$

$$B_i = \left[\frac{\partial f_i}{\partial \bar{\alpha}_i} ; \ldots \frac{\partial f_i}{\partial \bar{\alpha}_{10}} \right] \Bigg|_{X^o,L} = \left[0\ 0\ 0\ \sec^2\alpha_i\ 0\ 0\ 0 \right]$$

(ith column.)

$$W_i = f_i(X^o, L) = \tan \alpha_i - \frac{y_i}{x_i}$$

Using the parametric method we rewrite the model as

$$\bar{\alpha}_i = \arctan\left(\frac{y_i - \bar{y}_o}{x_i - \bar{x}_o}\right) - \bar{z}$$

from which

$$f_i(\bar{X}) - \bar{L}_i = \arctan\left(\frac{y_i - \bar{y}_o}{x_i - \bar{x}_o}\right) - \bar{z} - \bar{\alpha}_i = 0$$

and after linearization we have $AX - V + W = 0$ where the ith rows of A and W have the form

$$A_i = \left[\frac{\partial f_i}{\partial \bar{x}_o} ; \frac{\partial f_i}{\partial \bar{y}_o} ; \frac{\partial f_i}{\partial \bar{z}}\right]\Bigg|_{X^o} = \left[\frac{y_i}{x_i^2+y_i^2} ; \frac{-x_i}{x_i^2+y_i^2} ; -1\right]$$

$$W_i = f_i(X^o) - L_i = \arctan\left(\frac{y_i}{x_i}\right) - \alpha_i$$

The parametric method was used to produce the computer printouts which follow.

A.3 Solution

The nonlinear mathematical model

$$F(\bar{X}) - \bar{L} = 0$$

has been linearized about the initial approximation X^o and observed values L to give the linearized model

$$_{10}A_{3}\ {}_{3}X_{1} - {}_{10}V_{1} + {}_{10}W_{1} = 0$$

with weight matrix of the observations (assuming $\sigma_o^2 = 1$)

$$_{10}P_{10} = \sigma_o^2\ \Sigma_L^{-1} = \Sigma_L^{-1}$$

Matrices X^o, L, A, W and P are shown in Table A-1.

The least squares estimators are:

$$_{3}\hat{X}_{1} = -\ (A^TPA)^{-1}\ A^TPW$$

$$_{3}\hat{\bar{X}}_{1} = X^o + \hat{X}$$

$$_{10}\hat{V}_{1} = A\hat{X} + W$$

$$_{10}\hat{\bar{L}}_{1} = L + \hat{V}$$

$$\hat{\sigma}_o^2 = \frac{\hat{V}^TP\hat{V}}{n - u} = \frac{\hat{V}^TP\hat{V}}{7}$$

$$\hat{\Sigma}_{\hat{X}} = \hat{\sigma}_o^2\ (A^TPA)^{-1}$$

Matrices $\hat{X}$, $\hat{\bar{X}}$, $\hat{V}$, $\hat{\bar{L}}$, and $\hat{\Sigma}_{\hat{X}}$ are shown in Table A-2 .

These results may be summarized as follows:

a) The solution vector is

$$_{3}\hat{X}_{1} = \begin{bmatrix} x_o \\ y_o \\ z \end{bmatrix} = \begin{bmatrix} 0.63 \text{ meters} \\ -0.25 \text{ meters} \\ -0.32 \times 10^{-3} \text{ radians} = -1.1 \text{ arcminute} \end{bmatrix}$$

b) The variances (diagonal elements of $\hat{\Sigma}_{\hat{X}}$) are

$3X_1^0$

		1
RCW	1	0.0
RCW	2	0.0
RCW	3	0.0

$10L_1$

		1
RCW	1	0.4660849400D 00
RCW	2	-0.4638997885D 00
RCW	3	0.1812245176D 01
RCW	4	-0.2757581821D 01
RCW	5	0.9274557472D 00
RCW	6	0.1191206413D 01
RCW	7	0.2357369097D 01
RCW	8	-0.8782322263D 00
RCW	9	-0.2109638808D 01
RCW	10	0.2943899209D 01

$10A_3$

		1	2	3
RCW	1	0.2000000000D-03	-0.4000000000D-03	-0.1000000000D 01
RCW	2	-0.1000000000D-03	-0.2000000000D-03	-0.1000000000D 01
RCW	3	0.1176470588D-03	0.2941176471D-04	-0.1000000000D 01
RCW	4	-0.6896551724D-04	0.1724137931D-03	-0.1000000000D 01
RCW	5	0.8000000000D-04	-0.6000000000D-04	-0.1000000000D 01
RCW	6	0.1724137931D-03	-0.6896551724D-04	-0.1000000000D 01
RCW	7	0.8333333333D-04	0.8333333333D-04	-0.1000000000D 01
RCW	8	-0.9836065574D-04	-0.8196721311D-04	-0.1000000000D 01
RCW	9	-0.7352941176D-04	0.4411764706D-04	-0.1000000000D 01
RCW	10	0.3846153846D-04	0.1923076923D-03	-0.1000000000D 01

$10W_1$

		1
RCW	1	-0.2437331757D-02
RCW	2	0.2521795330D-03
RCW	3	0.3529814406D-02
RCW	4	-0.3504455570D-02
RCW	5	-0.1605292034D-03
RCW	6	-0.9164634254D-03
RCW	7	-0.1174606587D-02
RCW	8	0.2174178713D-02
RCW	9	-0.1577018985D-02
RCW	10	0.2978645235D-03

$P = 328\ 280.635\ {}_{10}I_{10}$

TABLE A-1

$3\hat{X}_1$

		1
RCW	1	0.6331934126D C0
RCW	2	-0.2523943427D 00
RCW	3	-0.3221C7C452D-C3

$3\hat{\bar{X}}_1$

		1
RCW	1	0.6331934126D C0
RCW	2	-0.2523943427D C0
RCW	3	-0.3221C7C452D-C3

$10\hat{V}_1$

		1
RCW	1	-0.1887628292D-02
RCW	2	0.5614461C55D-C3
RCW	3	0.3918991431D-C2
RCW	4	-0.3269533C2D-C2
RCW	5	0.2273769753D-C3
RCW	6	-0.4677785958D-C3
RCW	7	-0.82C7662856D-C3
RCW	8	0.2454689S0D-02
RCW	9	-0.1312EC5324D-02
RCW	10	0.5958C77879D-C3

$10\hat{L}_1$

		1
RCW	1	0.4641973125D CC
RCW	2	-0.4633383424D C0
RCW	3	0.1810104107D 01
RCW	4	-0.2760851354D C1
RCW	5	0.9276831242D C0
RCW	6	0.1190738635D 01
RCW	7	0.2356548330D 01
RCW	8	-0.8797775368D CC
RCW	9	-0.2110951413D C1
RCW	10	0.2944495C17D C1

$$\hat{\sigma}_0^2 = \frac{\hat{V}^T P \hat{V}}{\nu}$$

7	
0.1279450754D C2	ν
0.1827786791D C1	

$3\hat{\Sigma}_{\hat{X}_3}$

		1	2	3
RCW	1	0.53C7141924D 02	C.1045912246D C2	C.1560174413D-02
RCW	2	0.1045912246D 02	C.2172367921D 02	-C.2614560551D-03
RCW	3	0.1560174413D-02	-C.2614560551D-C3	C.6191030316D-06

TABLE A-2

$$\hat{\sigma}^2_{x_o} = 53 \text{ meters}^2$$

$$\hat{\sigma}^2_{y_o} = 21 \text{ meters}^2$$

$$\hat{\sigma}^2_{z} = 0.62 \times 10^{-6} \text{ radians}^2$$

c) The standard deviations are

$$\hat{\sigma}_{x_0} = 7.3 \text{ meters}$$

$$\hat{\sigma}_{y_o} = 4.6 \text{ meters}$$

$$\hat{\sigma}_{z} = 2.6 \text{ arcminutes}$$

d) The 95% confidence interval for the a priori variance factor is

$$\left[\frac{\hat{V}^T P \hat{V}}{\chi^2_{0.975}} \leq \sigma^2_o \leq \frac{\hat{V}^T P \hat{V}}{\chi^2_{0.025}} \right]$$

or

$$\left[\frac{12.79}{16.01} \leq \sigma^2_o \leq \frac{12.79}{1.69} \right]$$

or

$$\left[0.80 \leq \sigma^2_o \leq 7.6 \right].$$

Therefore the hypothesis

$$H_o: \sigma^2_o = 1$$

is not rejected.

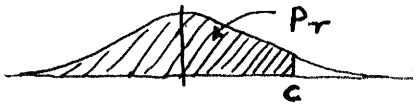
APPENDIX B

STATISTICAL TABLES

---

taken from Natrella, M.G. (1966). "Experimental Statistics", U.S. National Bureau of Standards Handbook 91.

TABLE B-1. CUMULATIVE NORMAL DISTRIBUTION - VALUES OF Pr

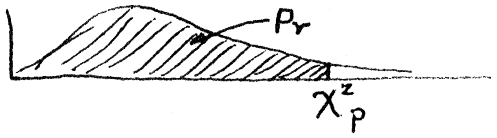


Values of Pr corresponding to c for the normal curve.

The value of Pr for (-c) equals one minus the value of Pr for (+c).

c	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

TABLE B-2. PERCENTILES OF THE χ^2 DISTRIBUTION



Values of χ^2_p corresponding to Pr.

ν	$\chi^2_{.005}$	$\chi^2_{.01}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.10}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$
1	.000039	.00016	.00098	.0039	.0158	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.1026	.2107	4.61	5.99	7.38	9.21	10.60
3	.0717	.115	.216	.352	.584	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.064	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64

For large degrees of freedom,

$$\chi^2_P = \tfrac{1}{2}(z_P + \sqrt{2\nu - 1})^2 \text{ approximately,}$$

where ν = degrees of freedom and z_P is given in Table A-2.

TABLE B-2 cont'd. PERCENTILES OF THE X^2 DISTRIBUTION

PROBABILITY FUNCTIONS

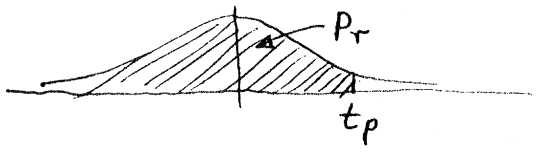
PERCENTAGE POINTS OF THE x^2-DISTRIBUTION—VALUES OF x^2 IN TERMS OF Q AND ν

ν \ Q	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25
1	(-5)3.92704	(-4)1.57088	(-4)9.82069	(-3)3.93214	0.0157908	0.101531	0.454937	1.32330
2	(-2)1.00251	(-2)2.01007	(-2)5.06356	0.102587	0.210720	0.575364	1.38629	2.77259
3	(-2)7.17212	0.114832	0.215795	0.351846	0.584375	1.212534	2.36597	4.10835
4	0.206990	0.297110	0.484419	0.710721	1.063623	1.92255	3.35670	5.38527
5	0.411740	0.554300	0.831211	1.145476	1.61031	2.67460	4.35146	6.62568
6	0.675727	0.872085	1.237347	1.63539	2.20413	3.45460	5.34812	7.84080
7	0.989265	1.239043	1.68987	2.16735	2.83311	4.25485	6.34581	9.03715
8	1.344419	1.646482	2.17973	2.73264	3.48954	5.07064	7.34412	10.2188
9	1.734926	2.087912	2.70039	3.32511	4.16816	5.89883	8.34283	11.3887
10	2.15585	2.55821	3.24697	3.94030	4.86518	6.73720	9.34182	12.5489
11	2.60321	3.05347	3.81575	4.57481	5.57779	7.58412	10.3410	13.7007
12	3.07382	3.57056	4.40379	5.22603	6.30380	8.43842	11.3403	14.8454
13	3.56503	4.10691	5.00874	5.89186	7.04150	9.29906	12.3398	15.9839
14	4.07468	4.66043	5.62872	6.57063	7.78953	10.1653	13.3393	17.1170
15	4.60094	5.22935	6.26214	7.26094	8.54675	11.0365	14.3389	18.2451
16	5.14224	5.81221	6.90766	7.96164	9.31223	11.9122	15.3385	19.3688
17	5.69724	6.40776	7.56418	8.67176	10.0852	12.7919	16.3381	20.4887
18	6.26481	7.01491	8.23075	9.39046	10.8649	13.6753	17.3379	21.6049
19	6.84398	7.63273	8.90655	10.1170	11.6509	14.5620	18.3376	22.7178
20	7.43386	8.26040	9.59083	10.8508	12.4426	15.4518	19.3374	23.8277
21	8.03366	8.89720	10.28293	11.5913	13.2396	16.3444	20.3372	24.9348
22	8.64272	9.54249	10.9823	12.3380	14.0415	17.2396	21.3370	26.0393
23	9.26042	10.19567	11.6885	13.0905	14.8479	18.1373	22.3369	27.1413
24	9.88623	10.8564	12.4011	13.8484	15.6587	19.0372	23.3367	28.2412
25	10.5197	11.5240	13.1197	14.6114	16.4734	19.9393	24.3366	29.3389
26	11.1603	12.1981	13.8439	15.3791	17.2919	20.8434	25.3364	30.4345
27	11.8076	12.8786	14.5733	16.1513	18.1138	21.7494	26.3363	31.5284
28	12.4613	13.5648	15.3079	16.9279	18.9392	22.6572	27.3363	32.6205
29	13.1211	14.2565	16.0471	17.7083	19.7677	23.5666	28.3362	33.7109
30	13.7867	14.9535	16.7908	18.4926	20.5992	24.4776	29.3360	34.7998
40	20.7065	22.1643	24.4331	26.5093	29.0505	33.6603	39.3354	45.6160
50	27.9907	29.7067	32.3574	34.7642	37.6886	42.9421	49.3349	56.3336
60	35.5346	37.4848	40.4817	43.1879	46.4589	52.2938	59.3347	66.9814
70	43.2752	45.4418	48.7576	51.7393	55.3290	61.6983	69.3344	77.5766
80	51.1720	53.5400	57.1532	60.3915	64.2778	71.1445	79.3343	88.1303
90	59.1963	61.7541	65.6466	69.1260	73.2912	80.6247	89.3342	98.6499
100	67.3276	70.0648	74.2219	77.9295	82.3581	90.1332	99.3341	109.141
X	-2.5758	-2.3263	-1.9600	-1.6449	-1.2816	-0.6745	0.0000	0.6745

$$Q(x^2|\nu) = \left[2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} \int_{x^2}^{\infty} e^{-\frac{t}{2}} t^{\frac{\nu}{2}-1}\, dt$$

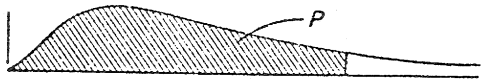
From E. S. Pearson and H. O. Hartley (editors), Biometrika tables for statisticians, vol. I. Cambridge Univ. Press, Cambridge, England, 1954 (with permission) for $Q > 0.0005$.

TABLE B-3. PERCENTILES OF THE t DISTRIBUTION



ν	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	.325	.727	1.376	3.078	6.314	12.706	31.821	63.657
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925
3	.277	.584	.978	1.638	2.353	3.182	4.541	5.841
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032
6	.265	.553	.906	1.440	1.943	2.447	3.143	3.707
7	.263	.549	.896	1.415	1.895	2.365	2.998	3.499
8	.262	.546	.889	1.397	1.860	2.306	2.896	3.355
9	.261	.543	.883	1.383	1.833	2.262	2.821	3.250
10	.260	.542	.879	1.372	1.812	2.228	2.764	3.169
11	.260	.540	.876	1.363	1.796	2.201	2.718	3.106
12	.259	.539	.873	1.356	1.782	2.179	2.681	3.055
13	.259	.538	.870	1.350	1.771	2.160	2.650	3.012
14	.258	.537	.868	1.345	1.761	2.145	2.624	2.977
15	.258	.536	.866	1.341	1.753	2.131	2.602	2.947
16	.258	.535	.865	1.337	1.746	2.120	2.583	2.921
17	.257	.534	.863	1.333	1.740	2.110	2.567	2.898
18	.257	.534	.862	1.330	1.734	2.101	2.552	2.878
19	.257	.533	.861	1.328	1.729	2.093	2.539	2.861
20	.257	.533	.860	1.325	1.725	2.086	2.528	2.845
21	.257	.532	.859	1.323	1.721	2.080	2.518	2.831
22	.256	.532	.858	1.321	1.717	2.074	2.508	2.819
23	.256	.532	.858	1.319	1.714	2.069	2.500	2.807
24	.256	.531	.857	1.318	1.711	2.064	2.492	2.797
25	.256	.531	.856	1.316	1.708	2.060	2.485	2.787
26	.256	.531	.856	1.315	1.706	2.056	2.479	2.779
27	.256	.531	.855	1.314	1.703	2.052	2.473	2.771
28	.256	.530	.855	1.313	1.701	2.048	2.467	2.763
29	.256	.530	.854	1.311	1.699	2.045	2.462	2.756
30	.256	.530	.854	1.310	1.697	2.042	2.457	2.750
40	.255	.529	.851	1.303	1.684	2.021	2.423	2.704
60	.254	.527	.848	1.296	1.671	2.000	2.390	2.660
120	.254	.526	.845	1.289	1.658	1.980	2.358	2.617
∞	.253	.524	.842	1.282	1.645	1.960	2.326	2.576

TABLE B-4 PERCENTILES OF THE F DISTRIBUTION



$F_{.90}(n_1, n_2)$

n_1 = degrees of freedom for numerator

n_2 = degrees of freedom for denominator

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.50	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

TABLE B-4 (Continued). PERCENTILES OF THE F DISTRIBUTION

$F_{.95}(n_1, n_2)$

n_1 = degrees of freedom for numerator

n_2 = degrees of freedom for denominator

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

TABLE B–4 (Continued). PERCENTILES OF THE F DISTRIBUTION

$F_{.975}(n_1, n_2)$

n_1 = degrees of freedom for numerator

n_2 = degrees of freedom for denominator

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

TABLE B–4 (Continued). PERCENTILES OF THE F DISTRIBUTION

$F_{.99}(n_1, n_2)$

n_1 = degrees of freedom for numerator

n_2 = degrees of freedom for denominator

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

APPENDIX C

PROPERTIES OF EXPECTED VALUES

Definitions:

Given the random variable x with probability density function ϕ, then the expected value of a function f of x is

$$E\,[f(x)] = \int_{-\infty}^{\infty} f(x)\ \phi(x)\ dx \quad \text{for x continuous}$$

$$E\,[f(x)] = \Sigma\ f(x)\ \phi(x) \quad \text{for x discrete.}$$

Given the random variables x, y with joint probability density function ϕ, then the expected value of a function f of x and y is

$$E\,[f(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ \phi\ (x,y)\ dx\ dy \quad \text{for x, y continuous}$$

$$E\,[f(x,y)] = \Sigma\ \Sigma\ f(x,y)\ \phi(x,y) \quad \text{for x, y discrete.}$$

Properties (Theorems):

Given a constant k

$$E\,[k] = k$$

$$E\,[k\ f(x)] = k\ E\,[f(x)]$$

$$E\,[E[f(x)]] = E\,[f(x)]$$

$$E\,[\Sigma\ f_i(x)] = \Sigma\ E\,[f_i(x)]$$

APPENDIX D

PROPERTIES OF MOMENT GENERATING FUNCTIONS

Definition:

Given the random variable x with probability density function ϕ, the moment generating function of ϕ is

$$M_x(t) = E\,[\exp\,(xt)]$$

Properties (Theorems):

Given a constant k

$$M_{x+k}(t) = \exp\,(kt)\; M_x(t)$$

$$M_{kx}(t) = M_x(kt)$$

Given constants k_i and statistically independent random variables x_i

$$M_{\Sigma k_i c_i}(t) = \prod_i M_{x_i}(k_i t).$$

APPENDIX E

PROPERTIES OF MATRIX TRACES*

Definitions:

Given a square matrix A, its trace is the sum of its diagonal elements

$$\text{Trace } A = \text{Tr}(A) = \sum_i a_{ii} \ .$$

Given a matrix A and a square matrix F which is a product of matrices including A, the partial derivative of the trace of F with respect to the matrix A is a matrix whose elements are the partial derivatives of the trace of F with respect to the corresponding elements of A, that is if

$$A = [a_{ij}]$$

then

$$\frac{\partial \text{Tr}(F)}{\partial A} = [\frac{\partial \ \text{Tr}(F)}{\partial a_{ij}}] \ .$$

Properties (Theorems):

$$\text{Tr}(A^T) = \text{Tr}(A)$$

Given a constant k

$$\text{Tr}(kA) = k \ \text{Tr}(A)$$

Given two matrices A and B conformable under addition

$$\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B)$$

Given two matrices A and B conformable under both multiplications AB and BA

$$\text{Tr}(A \ B) = \text{Tr} \ (B \ A)$$

---

* A complete discussion of these properties of traces is found in Blaha, G. (1971). "Inner Adjustment Constraints With Emphasis on Range Observations", Reports of the O.S.U. Department of Geodetic Science, Report No. 148.

Given two matrices A and B conformable under both multiplications $A^T B$ and AB^T

$$\text{Tr } (A^T B) = \text{Tr } (A B^T)$$

From the above properties it is evident that similar matrices have the same trace, that is for any nonsingular matrix R, and any matrix A of same order as R

$$\text{Tr } (R^{-1} A R) = \text{Tr } (A)$$

and in particular if R is the orthogonal matrix which diagonalizes A we have

$$\text{Tr } (A) = \sum_i \lambda_i$$

where λ_i are the eigenvalues of A.

A property of the derivative of Tr (F) is

$$\frac{\partial \text{ Tr}(F)}{\partial A^T} = \left[\frac{\partial \text{ Tr}(F)}{\partial A} \right]^T$$

For specific forms of F we have

$$F = AB \qquad \frac{\partial \text{ Tr}(A B)}{\partial A} = \frac{\partial \text{ Tr } (B A)}{\partial A} = B^T$$

$$F = ABA^T \qquad \frac{\partial \text{Tr } (A B A^T)}{\partial A} = A(B + B^T)$$

$$F = A^T BA \qquad \frac{\partial \text{ Tr}(A^T BA)}{\partial A} = (B + B^T)A$$

$$F = ABA^T C \qquad \frac{\partial \text{ Tr } (A B A^T C)}{\partial A} = C^T AB^T + CAB$$