

Observations and Least squares

The IEP Series in Civil Engineering

Series Editor
Russell C. Brinker
New Mexico State University

Brinker and Wolf	<i>ELEMENTARY SURVEYING</i>, 6th edition
Clark, Viessman, and Hammer	<i>WATER SUPPLY AND POLLUTION CONTROL</i>, 3rd edition
Ghali and Neville	<i>STRUCTURAL ANALYSIS: A UNIFIED CLASSICAL AND MATRIX APPROACH</i>
Hill	<i>FUNDAMENTALS OF STRUCTURAL DESIGN: STEEL, CONCRETE, AND TIMBER</i>
Jumikis	<i>FOUNDATION ENGINEERING</i>
McCormac	<i>STRUCTURAL ANALYSIS</i>, 3rd edition
McCormac	<i>STRUCTURAL STEEL DESIGN</i>, 2nd edition
Meyer	<i>ROUTE SURVEYING AND DESIGN</i>, 4th edition
Moffitt	<i>PHOTOGRAMMETRY</i>, 2nd edition
Moffitt and Bouchard	<i>SURVEYING</i>, 6th edition
Salmon and Johnson	<i>STEEL STRUCTURES: DESIGN AND BEHAVIOR</i>
Spangler and Handy	<i>SOIL ENGINEERING</i>, 3rd edition
Ural	<i>FINITE ELEMENT METHOD: Basic Concepts and Applications</i>
Viessman, Harbaugh, Knapp, and Lewis	<i>INTRODUCTION TO HYDROLOGY</i>, 2nd edition
Wang	<i>COMPUTER METHODS IN ADVANCED STRUCTURAL ANALYSIS</i>
Wang	<i>MATRIX METHODS OF STRUCTURAL ANALYSIS</i>, 2nd edition
Wang and Salmon	<i>REINFORCED CONCRETE DESIGN</i>, 2nd edition
Winfrey	<i>ECONOMIC ANALYSIS FOR HIGHWAYS</i>

OBSERVATIONS AND LEAST SQUARES

Edward M. Mikhail

Purdue University

with contributions by

F. Ackermann

Stuttgart University

Copyright © 1976 by Thomas Y. Crowell Company, Inc.

All Rights Reserved

Except for use in a review, the reproduction or utilization of this work in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, and in any information storage and retrieval system is forbidden without the written permission of the publisher.

Published simultaneously in Canada by Fitzhenry & Whiteside, Ltd., Toronto.

Library of Congress Cataloging in Publication Data

Mikhail, Edward M.

Observations and least squares.

Bibliography: p.

Includes index.

1. Least squares. 2. Errors, Theory of. I. Ackermann, Friedrich E., 1929- joint author. II. Title.

QA275.M62 1976 519.5 76-10277

ISBN 0-7002-2481-5

Tables D1, D2, D3, D4(a), D4(b), and D4(c) adapted with permission of Addison-Wesley, from *Handbook of Statistical Tables* by Donald B. Owen. © 1962 by Addison-Wesley, Reading Mass. courtesy of the U.S. Energy Research and Development Adm. Adapted and reproduced with permission of John Wiley & Sons, Inc. from *Random Data: Analysis and Measurement* by Julius S. Bendat and Allan G. Piersol, © 1971 by John Wiley & Sons, Inc.

IEP—A Dun-Donnelley Publisher
666 Fifth Avenue
New York, New York 10019

Manufactured in the United States of America

Contents

Preface	ix
PART I	1
Chapter 1. Concepts of Observation and the Model	3
1.1. Introduction	3
1.2. Observations (Measurements)	
1.3. The Mathematical Model	5
Chapter 2. Review of Statistical Concepts	9
2.1. Introduction	9
2.2. Probability, Distributions, and Density Functions	10
2.3. Multidimensional Distributions, Marginal and Conditional Distributions, Independence	13
2.4. Expectations, Moments, and Correlation	16
2.5. Some Often Used Distributions	22
2.6. Multinormal Distribution	27
2.7. Sampling, Estimation, and Confidence Measures	35
2.8. Statistical Tests	53

Chapter 3. Error Properties of Observations	60
3.1. Introduction	60
3.2. Random Errors	61
3.3. Precision, Accuracy, Cofactors, and Weights	64
3.4. Blunders	66
3.5. Systematic Effects (Errors)	67
Chapter 4. Principle and Techniques of Propagation	72
4.1. Introduction	72
4.2. Propagation of Distributions	73
4.3. Propagation of Means	74
4.4. Propagation of Variances and Covariances	76
4.5. Propagation of Systematic and “True” Errors	89
Problems for Part I	91
PART II	99
Chapter 5. Introduction to Least Squares Adjustment	101
5.1. General	101
5.2. The Least Squares Principle	103
5.3. The Techniques of Least Squares	105
5.4. Linear and Nonlinear Functions in the Model	108
Chapter 6. Adjustment with Conditions Only—General Case	110
6.1. Introduction	110
6.2. Derivation	112
6.3. Adjustment with Maximum Number of Independent Parameters (Adjustment with n Condition Equations)	128
6.4. Geometric Interpretation of the Least Squares Principle	131
6.5. Summary of Equations	134
Chapter 7. Adjustment with Conditions Only—Special Cases	136
7.1. Introduction	136
7.2. Adjustment of Observations Only	137
7.3. Adjustment of Indirect Observations	159
7.4. Summary of Equations and Concluding Remarks	173
Chapter 8. Examples and General Discussion on Adjustment with Conditions Only	175
8.1. General	175
8.2. Coordinate Transformations	184

Chapter 9. Least Squares Adjustment with Conditions and Constraints	213
9.1. Introduction	213
9.2. General Case for Adjustment with Conditions and Constraints	214
9.3. Special Cases	228
9.4. Constraints with Added Parameters	233
9.5. Summary of Equations	252
Chapter 10. Adjustment with Derived Observations and Adjustment in Steps (of Observations Only)	256
10.1. Introduction	256
10.2. Adjustment with Derived Observations	257
10.3. Adjustment in Steps	266
Chapter 11. Numerical and Statistical Considerations in Adjustment	274
11.1. Nonlinearity of the Equations	274
11.2. Approximate Values for Model Variables	281
11.3. A Posteriori Estimate of the Reference Variance ($\hat{\sigma}_0^2$)	285
11.4. Iteration Termination with Linearized Conditions	289
11.5. A Posteriori Statistical Analysis	294
11.6. Computational and Numerical Considerations	303
Problems for Part II	312
PART III	331
Chapter 12. A United Approach to Least Squares Adjustment	333
12.1. Introduction	333
12.2. Formulation	335
12.3. The Unified Approach and Nonlinear Functions	341
12.4. The Unified Approach and Parameter Constraints	343
12.5. Precision Estimation	349
12.6. The Reference Variance	353
12.7. Review Examples	354
12.8. Summary of Equations and Concluding Remarks	357
Chapter 13. Sequential Data Processing	360
13.1. Introduction	360
13.2. Mathematical Derivation	362
13.3. Sequential Processing and the Unified Approach	375
13.4. Sequential Reduction, the Unified Approach, and Parameter Constraints	384

13.5. Precision Estimation	387
13.6. Sequential Processing and Nonlinear Equations	387
13.7. Summary of Equations and Concluding Remarks	390
Chapter 14. An Introduction to Least Squares Interpolation, Filtering, and Collocation	393
14.1. Introduction	393
14.2. Linear Interpolation of Stationary Functions	399
14.3. Linear Filtering of Stationary Functions	401
14.4. Extension to Multidimensional Cases	402
14.5. Trend Surfaces, Covariance Functions, and Experimental Results	403
14.6. Examples of Filtering Applications	409
14.7. Least Squares Collocation	418
Problems for Part III	427
Appendix A. Review of Matrix Theory	431
A1. Definitions	431
A2. Types of Matrices	432
A3. Matrix Operations	434
A4. Rank of a Matrix	447
A5. Trace of a Matrix	451
A6. The Eigenvalue Problem	451
A7. Bilinear and Quadratic Forms	455
A8. Matrix Differentiation	457
A9. Some Useful Matrix Relationships	460
Appendix B. Linearization by Series Expansion	462
B1. One Function of One Variable	462
B2. One Function of Two Variables	463
B3. Two Functions of One Variable	464
B4. Two Functions of Two Variables Each	464
B5. General Case of m Functions of n Variables	465
Appendix C. Ranks of a Posteriori Cofactor (Covariance) Matrices	467
C1. Adjustment of Observations Only	467
C2. Adjustment of Indirect Observations	470
C3. Adjustment of Combined Observations and Parameters	471
Appendix D. Pertinent Statistical Tables	475
Bibliography	485
Index	493

Preface

This book is an outgrowth of lecture material that evolved over many years of teaching the subject of least squares at the undergraduate and the graduate levels by the primary and contributing authors. The presentation is such that the beginner may acquire a thorough understanding of the basic concepts. There is also sufficient material to challenge the advanced student. The book is therefore suitable for a first course and for one or two advanced courses on observations and least squares. In addition, the numerous examples worked out in detail together with complete derivations make it equally useful for self-study or as a reference.

The book is divided into three parts: Part I deals with fundamental concepts, Part II covers in detail the common techniques of least squares adjustment, and Part III discusses advanced topics. The appendices following the main text provide necessary supplementary material.

Chapter 1, which introduces the concepts of observation and the mathematical model, should be read first. Chapter 2 includes a review of statistical concepts that are pertinent to the main theme of the book. Since it is not intended to substitute for statistical literature, it contains very few theoretical derivations. As a review, however, it is sufficiently complete, covering both one- and multidimensional cases. The beginner need not read this chapter in its entirety before proceeding. Chapters 3 and 4

are concerned with "errors" and their propagation. "Error," a commonly used term, is related to the more appropriate term "observation" in the context of statistical notions. For propagation the reader needs familiarity with matrix manipulation. Prior review of Appendix A is necessary before reading Chapter 4.

An introductory undergraduate course on the subject might include Chapter 1, parts of Chapter 2, Chapters 3, 4 in Part I; Chapter 5, 7 and parts of Chapters 8 and 11 in Part II. If the first course is dual-level undergraduate-graduate, Chapter 6 should be added to the list. Thus the student would be exposed to basic statistics, propagation, and two or three techniques of least squares adjustment using condition equations. The term "condition equations" has been generalized in this book to mean all equations involving observations.

A second course might include Chapter 2; a review of Chapters 4 and 6, Chapters 9, 10 and 11 in Part II; and parts of Chapters 12 and 13 in Part III. At the end of such a course the student would have learned techniques of adjustment with constraints (equations involving only parameters); unified adjustment where all variables in the model are treated as observations; adjustment in steps; and sequential data processing.

For more advanced study, the latter parts of Chapters 12 and 13, as well as 14 in Part III provide the reader with further material. In particular, Chapter 14 introduces the concepts of interpolation (prediction), filtering, and collocation. After reading this chapter, the student will be better prepared to further his knowledge in these advanced subjects by studying the technical literature.

The development of the topics is systematic in that each technique is progressively shown to be an extension of a preceding simpler technique. Essential to the following of the development is familiarity with matrix algebra, a review of which is given in Appendix A. Although generalized matrix algebra may be used instead, it is not considered necessary to the topics covered in this book.

The presentation is complete in itself. However, a bibliographical list is given at the end of the book to provide a source of supplementary reading. The reader should consult this list when seeking additional information on specific subjects.

The terms used are usually defined when introduced. However, some common statistical and adjustment terms may have been employed without specific definition. Readers unfamiliar with such common terms may consult a dictionary of statistical terms or the *International Society for Photogrammetry Multi-Lingual Dictionary*, 1961.

There are many colleagues whose encouragement and assistance made the completion of this book possible and are gratefully acknowledged: Mr. D. C. Brown reviewed a major part of the manuscript, made valuable

recommendations, and allowed the use of some paragraphs and figures and adaptation of a section on recursive partitioning; Dr. E. Stark programmed most of the numerical examples; Dr. G. W. Marks edited Chapter 2 and computed some of its numerical examples; Professor K. Kraus contributed examples and illustrations to Chapter 14 which was reviewed by Dr. F. Leberl; Professor P.-L. Baetslé assisted with the material in Appendix C; Drs. B. Makarovič and P. Stefanovič read parts of the manuscript; Dr. D. L. Gifford had many discussions with the primary author and made significant recommendations; Professor K. Linkwitz allowed the use of Example 9.4; Professor H. Moritz permitted the use of the example in Chapter 14; and Mr. R. S. Johnson who assisted in reviewing the galley proofs. The primary author also wishes to acknowledge his past and present graduate students who taught him more than he is willing to admit and who continuously encouraged him to get this book published; and to thank his wife LaVerne without whose faithful support this book would not have become a reality.

**Observations
and
Least squares**

PART I

Concepts of Observation and the Model

1.1. INTRODUCTION

Adjustment is meaningful only in those cases in which the data available exceed the minimum necessary for a unique determination (redundant data). Since the data are usually obtained from observations (or measurements), which are subject to probabilistic fluctuations (classically known as errors), redundant data are usually inconsistent in the sense that each sufficient subset yields different results from another subset. To obtain a unique solution, an additional criterion (such as that of least squares) is applied.

In order that we may understand the details of adjustment techniques, a number of basic concepts and their interrelations are introduced in this chapter.

1.2. OBSERVATIONS (MEASUREMENTS)

Both the terms *measurement* and *observation* will be used alternatively hereafter as having the same meaning. The term “observation” (or measurement) is often used in practice to refer to both the operation or process itself, as well as the actual outcome of such operation. With regard to adjustment

the outcomes, and in particular *numerical* outcomes, will be designated as the “observations.” Such numerical observational data are fundamental to science and engineering because they supply the instrument for analysis and manipulation.

Casual consideration of the concept of measurement is not problematic, since we can easily accept such seemingly simple notions as measuring distances, angles, temperature, or speed. However, when we take a closer look at the operation called “measurement,” matters do not turn out to be so easy. As a matter of fact, even the measuring of a distance with a tape is a relatively complex operation, the analysis of which leads to the following fundamental properties of measurement:

1. To measure always means to perform a physical operation; and the process of measurement often consists of several more elementary operations, such as preparations, either instrument setup or calibration (or both), pointing, matching, and comparing.
2. The numerical reading or result obtained from the process is considered to represent the measurement. Thus the figure obtained as the measurement carries with it the circumstances and relevant historical data of how it originated.
3. Measurements are almost always performed with the help of instruments, no matter how simple, or with the aid of physical events. (An exception would be simple counting of a certain event.)
4. Measurements refer to standards, which are established by convention and which are rather arbitrary. Thus, in essence, to measure is to compare with a standard. Consequently, one is concerned with units and dimensions because the results of measurement always have a dimension of some sort.
5. Deeper consideration of measurements reveals that they refer to rather theoretical concepts, such as the geometrical abstractions used for distances and angles that have no real, direct equivalent in physical nature. We choose such concepts in order to describe certain elements of nature, such as location, area, or extension, that may be of interest to us.
6. Although performing measurements is an operation or a process, the results obtained from such processes take their meaning of a measurement only by association with the theoretical concepts to which they refer and on which they are based.

The theoretical abstractions to which measurements refer are called the *model*. In science and engineering, which are the subjects considered in this book, the model is almost always mathematical. Although it may appear that such a model concept is rather theoretical, it is in fact of basic importance to the subject of adjustment. A thorough understanding of the concept of the model considerably facilitates the task of adjustment.

1.3. THE MATHEMATICAL MODEL

Since most of the activities treated in this book pertain to quantitative problems, consideration will always be given to the *mathematical model*. It is defined here as a theoretical system or an abstract concept by which one describes a physical situation or a set of events. Such a description is not necessarily meant to be complete or exhaustive, but to relate only to those aspects or properties that are under consideration. Since a model serves a particular purpose, its setup can vary widely from one point of view to another. Thus the same physical system may be described by more than one model. The model then replaces the physical situation for the purpose of assessing it.

The mathematical model is often thought of as being composed of two parts: the functional model and the stochastic model. The *functional model* will in general describe the deterministic properties of the physical situation or event under consideration. On the other hand, the *stochastic model* designates and describes the nondeterministic or stochastic (probabilistic) properties of the variables involved, particularly those representing the observations. Both functional and stochastic models must be considered together at all times as there may be several possible combinations, each representing a possible mathematical model.

1.3.1. The Functional Model Whenever measurements are planned, some functional model is usually chosen to represent either a physical or a fictitious system with which the measurements are associated. In fact measurements are usually made in order to assess values for some or all of the parameters of the functional model. In photogrammetry, geodesy, and surveying we generally deal with geometric time-independent models and occasionally with dynamic models, such as the following:

1. Geometric model in surveying: A plane triangle, in Euclidian space, characterized by three angles, three corner points, three sides, and perhaps also orientation with respect to a coordinate system.
2. Geometric model in photogrammetry: Aerial photographs considered to be perspective images of (geometrical) terrain points.
3. Dynamic model in geodesy: Gravity field of the earth.
4. Dynamic models in photogrammetry: Orbital photography and time-dependent scanning photography.

Because the geometric models used are rather simple and easy to visualize, the elements of the model and the physical elements to which they refer are often not clearly distinguished. It must be recognized, however, that there are in nature no such objects as points, angles, distances, or coordinates. These are only elements of the functional model that are used to

describe corresponding features of natural objects or their relationships of location.

Functional models are very often not stated explicitly, as they may refer to conventional models which are mostly defined by implication. If, for instance, a surveyor says that he measured a distance, then he refers to two objects that are abstracted and considered as two geometrical points. Even then he may not refer to the distance in its direct geometrical meaning, but rather to its projection onto a plane or even on an ellipsoid. Similarly, an angle CAB from point C to B about another point A is usually not the angle in the plane ABC , but its projection in another plane such as the horizontal (XY) plane.

It is perhaps the role of education in a given field to expound different workable models for different tasks. It is part of the technical skill of a scientist or engineer to know in which cases to operate with certain models and in which cases to construct new ones.

In general the functional model should conform to the physical reality with sufficient accuracy for the intended purpose.

1.3.2. Relating Observations to the Model A functional model is a completely fictitious construction used to describe physical events by an intelligible system suitable for analysis. It is linked to physical realities by measurements or observations that are themselves physical operations. In simpler situations, measurements refer directly to at least some elements of the functional model. However, it is not necessary, and often not practical, that all elements of the model be observable. In more involved situations, measurements do not actually relate directly to the elements of the model in question. For example, electronic distance measurement is actually the measurement of time, or time differences, not of distances. Here, a lot more theory becomes involved that must be incorporated into the model when the measurements are linked to the model. In fact, additional variables are necessarily included and new functional relationships added, thus expanding the model concept beyond the apparently simple task of measuring a distance. Consequently, in order to associate the results of the measurements with the elements of the model, the model must be extended.

The augmentation of the functional model due to measurements and their properties is of basic importance. The evaluation of observations depends on how and by what instruments and methods they have been acquired. The "measured" length of a line depends to a great extent on the measuring process, on whether the calibration of the instrument is considered to be known, on which reductions are to be applied, and so on. The same applies to the case of measuring angular directions. For instance, we must consider the zero direction and determine its place in the model, whether it is to be taken as known or unknown.

It may be a relatively long process, therefore, until the outcome of a measurement can be associated with an element of the model. Very often the direct readings from the measuring operation have to be “reduced” or preprocessed (which implements a correction model) before they can be considered appropriate.

In linking the observations to the model many of their features are often discarded for the sake of simplicity. For example, the special circumstances under which the actual measurements are acquired are seldom taken into account. In the light of the extension of the functional model, the modeling of this part is simplified. However, whenever a change is made in the functional model, the stochastic model must be modified accordingly.

1.3.3. The Stochastic Model—Statistical Properties of Measurements As the practitioner knows from experience, measurements are always subject to some unaccountable influences. They may be subject to physical influences that cannot be completely controlled, resulting in a certain variability of outcomes when observations are repeated. The variability of the results of measurements may or may not be attributed to some distinct physical causes. Such statistical variations, either due to neglecting physical or time-dependent influences or due to more inherent qualities of physical processes, are a basic property of measurements.

In the past these variations were said to be due to observational errors. At present, however, we accept variability or randomness of the results of measurements as a principal property of the observations and refer to statistical concepts in order to account for it.

From a practical standpoint it is rather difficult to assess the statistical properties of the observations. One way is to obtain repeated observations and derive the required properties, but this is usually demanding. Another way, which is often used, is to assume the statistical properties on the basis of a general reference to similar observations that were performed under similar circumstances in the past. Therefore when measurements are made, all relevant physical and environmental circumstances should be recorded during the measuring period in order to be able to judge the results properly. Actually, we often accept in practice rather coarse approximations for the statistical properties of the observations. For example, in geodesy the observations are usually considered (statistically) independent and often of equal precision (weight). The same applies in photogrammetry to image coordinates, although it is rather well known that film shrinkage causes (physical) correlation between image points. The assumption of independent image coordinates is made mainly because of its simplicity and because of the practical difficulty experienced so far in determining the correlation.

The totality of the assumptions on the statistical properties of the variables involved is called the *stochastic model*. It includes all model variables and designates those that are considered fixed (that is, constant during the

adjustment, either known or a priori determined) and those that are considered free (that is, the parameters to be determined in the adjustment).

The classical theory of least squares adjustment did not explicitly specify the concept of stochastic model. Instead, the terms *observational errors* or *error properties of observations* were used. To the present practitioner, however, the term “measurement” or “observation” is not ambiguous and is used with very general meaning. In this book the term “observation” will refer to any quantity that is considered as a stochastic (random) variable and for which an estimate is available a priori. Such an estimate may be derived from direct or previous measurements or otherwise.

EXAMPLES OF VARIABLE DESIGNATIONS

1. The refraction coefficient in trigonometric height observation may be treated as a constant, as a free parameter to be determined from the adjustment, or as a (stochastic) variable (to which a “standard deviation” is associated). In the latter case it is considered as an observation and treated accordingly.
2. In fitting polynomials to a set of data one or more coefficients can be treated as observations (with “expectation,” or mean, zero and a given standard deviation). This practice can often alleviate the possibility of an indeterminate solution.

As has been alluded to above, in the theory of observation and adjustment we deal with random variables whose treatment requires knowledge of statistical concepts. Therefore the remainder of Part I, particularly Chapter 2, is devoted to those aspects of statistics that are useful in adjustment. The treatment of such subjects will by necessity be concise and the reader is advised to refer to statistical textbooks for more detailed study and deeper insight (see the Bibliography at the end of the book).

Chapter 3 of Part I deals with an elaboration on the properties of observations. It attempts to relate classical theory of errors and modern theory of observations. Chapter 4 covers the principle and techniques of propagation, from one set of random variables to another.

2

Review of Statistical Concepts

2.1. INTRODUCTION

This chapter summarizes briefly a number of statistical concepts that are needed for the understanding of the theory of observations as will be treated in this book. These concepts are also necessary for the introductory foundation of the method of least squares as well as its assessment and practical application. The concepts to be introduced will be extracted from the two related mathematical fields of probability and statistics. In simple terms probability involves the laws of chance regarding the outcome of experiments. Its theories deal with averages of mass phenomena that may occur sequentially or simultaneously. They attempt to describe and predict those averages that would have been realized if the experiments were repeated many times. Statistics, on the other hand, deals with applying the laws of probability either to obtain estimates or to make inferences based on a given set of observational data. It is therefore concerned with methods of collecting, analyzing, and interpreting data. Both of these disciplines are used to assist the experimenter in making proper judgments in the face of uncertainty. They yield criteria on the basis of which decisions may be made. However, they never actually lead directly to the decisions to be made.

2.2. PROBABILITY, DISTRIBUTIONS, AND DENSITY FUNCTIONS

2.2.1. Probability Classical definitions relate the concept of probability of an event to the frequency of its occurrence when repeating an experiment; this repetition may be real or hypothetical. This leads to defining probability as the limit of the frequency of occurrence when the number of repetitions approaches infinity. If, for instance, the relative frequency of throwing a 6 with a die tends toward $\frac{1}{6}$ (with number of throws n approaching ∞), then the probability of throwing a 6 with this die is considered to be $\frac{1}{6}$ or 0.166.

In modern statistics this concept of probability as expected frequency is no longer in common use. Instead, probability is now considered as an independent basic concept associated with statistical events, and its properties are being founded axiomatically. In order to appreciate this concept, we first introduce the idea of the random variable.

2.2.2. Random Variable Probabilities are associated with statistical events, whether real or hypothetical, and an event is the outcome of a statistical experiment (such as throwing dice, measuring angles, or counting defective units of a product). If a statistical event has several possible outcomes, we associate with that event a *stochastic or random variable*, \tilde{x} . A random variable may be defined as a variable that takes on several possible values to each of which is associated a probability.

In probability we normally seek the behavior of a system on the basis of a known mathematical model with specified parameters. The totality of elements to be studied and about which information regarding its behavior is sought is called the *population*. In the theoretical sense the population is assumed to contain an infinite number of observations, whereas in practice it may have a very large, but finite, number of observations. The population includes *all* possible values of the random variables under consideration. Said another way, the population is the total of all possible outcomes of the statistical event associated with the random variable. Because of the very large size of the population, it is either impossible or entirely impractical to study each and every one of its elements to evaluate its characteristics. Therefore we select only a certain number of observations from the population, which we call a *sample*, and study it. From the results of sample study, we may draw inferences and make statements regarding the population from which the sample was taken.

The method of drawing a sample, as well as its size, will influence the conclusions that may be made. Obviously, the larger the sample size, to a certain extent, the more confident we will be regarding the results. As far as the method of drawing the sample is concerned, we must be careful that it does not follow a regular pattern. When this does occur, we run the risk of having the elements of the sample exhibit systematic effects, and the extension of the sample results to the population may not be entirely valid. To

alleviate such a difficulty it must be ascertained that the sample is drawn *at random*, which means that each element of the population has an equal chance of being chosen into the sample. This is equivalent to saying that the selection of each element of the sample is *independent* of each other selection.

The total set of possible values of a random variable, \tilde{x} , together with their probabilities, constitute what is termed a *probability distribution* associated with the random variable. Consequently, a probability distribution describes the various probabilities related to the possible values of a random variable. These distributions are specified as two types: *cumulative distribution functions* and *probability density functions*.

2.2.3. The Cumulative Distribution Function The cumulative distribution function $F(x)$ for the random variable \tilde{x} is defined by the relationship

$$P(\tilde{x} \leq x) = F(x) \quad (2.1)$$

which says: The probability that the random variable \tilde{x} takes values less than or equal to x (x being a running variable) defines $F(x)$ or the cumulative distribution function. Because probabilities are, by axiomatic definition, limited to values between 0 and +1 (that is, $0 \leq P \leq +1$), the distribution function $F(x)$ satisfies the following marginal conditions:

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \lim_{x \rightarrow \infty} F(x) = +1 \quad (2.2)$$

Although the basic definition in equation (2.1) holds for discrete as well as continuous functions, we shall, in the context of this book, be mainly concerned with continuous distribution functions.

2.2.4. The Probability Density Function The probability density function $f(x)$ is formulated in a manner analogous to the concept of density in physics. It can be defined as the probability for an interval Δx . Assuming continuous differentiable functions, the relationship between $F(x)$ and $f(x)$ is

$$F(x) = \int_{-\infty}^x f(r) dr \quad (2.3)$$

and

$$f(x) = \frac{\partial F(x)}{\partial x} \quad (2.4)$$

From these two relationships we can easily deduce the following statements:

1. The probability that a random variable \tilde{x} takes values in the interval between x_1 and x_2 , where $x_2 > x_1$ is given by

$$P(x_1 < \tilde{x} < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(r) dr \quad (2.5)$$

which is depicted by the shaded area in Figure 2.1.

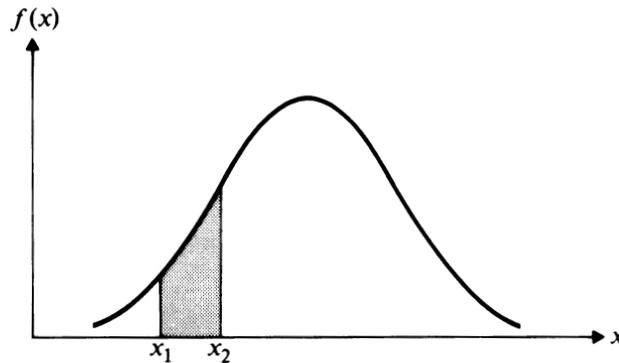


Figure 2.1

2. The probability that a random variable \tilde{x} takes values smaller than x_1 is given by

$$\begin{aligned} P(-\infty < \tilde{x} < x_1) &= F(x_1) - F(-\infty) = F(x_1) \\ &= \int_{-\infty}^{x_1} f(r) dr \end{aligned} \quad (2.6)$$

Note that $F(-\infty) = 0$ according to equation (2.2).

3. Similar to (2) we can write for the probability that \tilde{x} takes on values larger than x_1 .

$$\begin{aligned} P(x_1 < \tilde{x} < \infty) &= F(\infty) - F(x_1) = 1 - F(x_1) \\ &= \int_{x_1}^{\infty} f(r) dr \end{aligned} \quad (2.7)$$

Note again that $F(\infty) = 1$ according to equation (2.2).

From the preceding development two conditions stand out to characterize $f(x)$ as a probability density function:

1. $f(x) \geq 0$ for all values of x .
2. $\int_{-\infty}^{\infty} f(x) dx = 1$, since infinite values of x have been implied and since the total value of the probability, being in this case the entire area under the density curve, must equal 1.

The two criteria (1) and (2) given above must be strictly satisfied in order that a function may be considered as a probability density function.

Before addressing cases involving more than one random variable (variate), it should be mentioned that the probability density function expresses the whole *population*. It is usually represented by a number of variables which are called *parameters* in statistical terminology. Knowing the parameters totally specifies the density function. By contrast, the *sample*

values drawn for a population are used to compute quantities that are known as *statistics*. Each statistic usually represents an estimate of the corresponding population parameter.

2.3. MULTIDIMENSIONAL DISTRIBUTIONS, MARGINAL AND CONDITIONAL DISTRIBUTIONS, INDEPENDENCE

The concepts introduced in Section 2.2 involved only one random variable. In many practical situations we often encounter cases with more than one random variable. For example, the case of two random variables \tilde{x} and \tilde{y} has a two-dimensional cumulative distribution function $F(x, y)$. The joint probability of \tilde{x} taking values smaller than x and of \tilde{y} taking values smaller than y (x and y are both running variables) is given by

$$F(x, y) = P(\tilde{x} < x, \tilde{y} < y) \quad (2.8)$$

In a manner similar to the one-dimensional case, if $f(x, y)$ refers to the *two-dimensional probability density function*, and assuming continuous differentiable functions, then

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \quad (2.9)$$

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv \quad (2.10)$$

As before, since $0 \leq P \leq 1$ is the allowed region of values of probability, therefore

$$\lim_{\substack{x \rightarrow -\infty \\ y \rightarrow -\infty}} F(x, y) = 0 \quad \lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F(x, y) = +1 \quad (2.11)$$

Finally, the probability that the two random variables \tilde{x} and \tilde{y} take values between x_1 and x_2 , and y_1 and y_2 , respectively, is given by

$$P(x_1 < \tilde{x} < x_2; y_1 < \tilde{y} < y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(u, v) du dv \quad (2.12)$$

Geometric representation for two-dimensional joint probability distribution can be effected in a way similar to the one-dimensional case (Figure 2.2). A probability within the interval $x_1 \rightarrow x_2$ and $y_1 \rightarrow y_2$ is equal to the volume that is bounded by the surface $f(x, y)$, the $x - y$ plane and the four vertical planes, $x = x_1$, $x = x_2$, $y = y_1$, and $y = y_2$.

For the n -dimensional case, the n random variables are collected in a *random vector* $\tilde{\mathbf{x}}$, or

$$\tilde{\mathbf{x}} = [\tilde{x}_1 \tilde{x}_2 \cdots \tilde{x}_n]^t \quad (2.13)$$

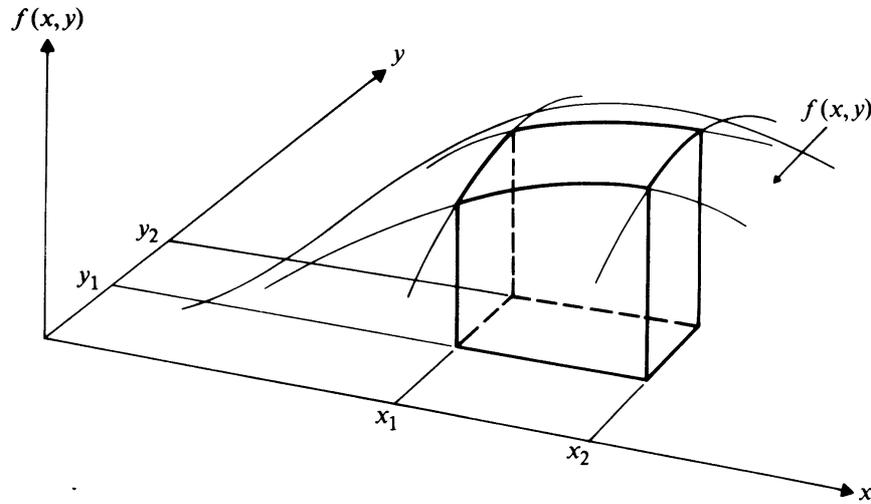


Figure 2.2

For that random vector, an n -dimensional joint density distribution function $f(\mathbf{x})$ is

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n) \quad (2.14)$$

and a corresponding cumulative distribution function $F(\mathbf{x})$, is

$$F(\mathbf{x}) = F(x_1, x_2, \dots, x_n) \quad (2.15)$$

The probability that the random vector $\tilde{\mathbf{x}}$ takes values smaller than \mathbf{x} is given by

$$\begin{aligned} F(\mathbf{x}) &= P(\tilde{\mathbf{x}} < \mathbf{x}) \\ &= P(\tilde{x}_1 < x_1, \tilde{x}_2 < x_2, \dots, \tilde{x}_n < x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \\ &\quad \dots \int_{-\infty}^{x_n} f(u_1, u_2, \dots, u_n) du_1 du_2 \dots du_n \end{aligned} \quad (2.16)$$

Although these relations show a relatively straightforward extension of the one-dimensional case, the two-dimensional and multidimensional cases are sufficiently different to warrant a few new concepts. These concepts include *marginal distribution*, *conditional distribution*, *independence (and correlation)*, all of which are only relevant when speaking of situations other than one-dimensional cases.

2.3.1. Marginal Distributions A marginal distribution is obtained from an n -dimensional distribution ($n \geq 2$) by disregarding the distribution of one or more of the components of the random vector $\tilde{\mathbf{x}}$. For example, the two-dimensional joint distribution for \tilde{x} and \tilde{y} may be reduced to the one-dimensional marginal distribution for \tilde{x} by disregarding its relation, if any,

with \tilde{y} . This is accomplished technically by selecting for \tilde{y} the upper limit value of $y = +\infty$ in the cumulative distribution function of equation (2.8) or (2.10). Consequently, the *marginal cumulative distribution function* of the random variable \tilde{x} is given as the joint probability of \tilde{x} taking values smaller than x_1 and \tilde{y} taking values less than $+\infty$, or

$$F_m(x) = F(x, \infty) = P(\tilde{x} < x, \tilde{y} < \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(u, v) du dv \quad (2.17)$$

The *joint density distribution function* $f(x, y)$ is reduced to the marginal form

$$f_m(x) = \int_{-\infty}^{\infty} f(u, v) dv \quad (2.18)$$

which leads to the marginal cumulative distribution function for \tilde{x} as

$$F_m(x) = \int_{-\infty}^x f_m(x) dx = P(\tilde{x} < x) \quad (2.19)$$

Extending to the n -dimensional case, we can readily see that marginal distributions exist for all possible combinations of the random variables in groups of 1, 2, ..., and $(n - 1)$.

2.3.2. Independence Let $F(x, y)$ denote the joint two-dimensional cumulative distribution function of the two random variables \tilde{x} and \tilde{y} , and $F(x)$ and $F(y)$ denote the marginal cumulative distribution functions for \tilde{x} and \tilde{y} , respectively. The two random variables \tilde{x} and \tilde{y} are said to be *independent* if

$$F(x, y) = F(x) \cdot F(y) \quad (2.20)$$

Equation (2.20) also leads to a relation for density functions, provided they are continuous. Thus for two *independent* random variables

$$f(x, y) = f(x) \cdot f(y) \quad (2.21)$$

In fact, the relation in equation (2.21) can be directly extended to the case of n -dimensional density functions, or

$$f(x_1, x_2, \dots, x_n) = f(x_1) \cdot f(x_2) \cdots f(x_n) \quad (2.22)$$

The concept of *independence* of two or more random variables is a consequence of the concept of independence of statistical events. It may be recalled that two events are said to be statistically independent if the probability of the joint event is equal to the product of the probabilities of the separate events, that is,

$$P(ab) = P(a) \cdot P(b) \quad (2.23)$$

We should distinguish between statistical or stochastic independence and functional independence of variables. Two variables may be functionally independent but may not necessarily be stochastically independent.

2.3.3. Conditional Distribution The concept of conditional distribution is defined in a manner similar to that of conditional probability. For example, $P(b|a)$, which reads probability of b given a , is the probability of the statistical event b on condition that the statistical event a has happened. Analogously, the *conditional* distribution of $(\tilde{y}|\tilde{x})$ (this reads the distribution of \tilde{y} given \tilde{x}) is the distribution of \tilde{y} for a given value of \tilde{x} from a two-dimensional distribution of the random variables \tilde{x} and \tilde{y} .

The conditional density function of \tilde{y} given \tilde{x} is

$$f(y|x) = \frac{f(x, y)}{f_m(x)} \quad (2.24)$$

The two concepts of independence and conditional distribution are directly related to each other. If two random variables \tilde{x} and \tilde{y} are independent, then the conditional distribution of \tilde{y} given \tilde{x} is the same for any value of \tilde{x} and vice versa. In terms of the relationships already given

$$f(y|x) = \frac{f(x, y)}{f_m(x)} = \frac{f_m(x) \cdot f_m(y)}{f_m(x)} = f_m(y) = f(y) \quad (2.25)$$

Thus the values of \tilde{x} have no influence on the conditional distributions of \tilde{y} that are then all equal and identical to the marginal distribution of \tilde{y} .

The concepts of marginal and conditional distributions and of correlation are illustrated for the two-dimensional case in Section 2.7.1 on sampling by a numerical example on the stereogram (bidirectional histogram). Therefore the reader may consult that section at this point in order to gain a better understanding of these otherwise apparently abstract topics.

2.4. EXPECTATIONS, MOMENTS, AND CORRELATION

The distribution and density functions of random variables are characterized (and determined) by a number of parameters which are useful for the understanding of their behavior. The first of these expresses the intuitive concept of the mean and is referred to by one of several terms, such as expectation, expected value, mean, or average.

2.4.1. Expectation The expectation $E(\tilde{x})$ of a random variable \tilde{x} , if it exists, is defined as the average value μ_x of the variable over all possible values. It is computed by taking the sum of all possible values x_i of \tilde{x} each multiplied by its corresponding probability $P(x_i)$. Thus

$$E(\tilde{x}) = \mu_x = \sum_{i=1}^n x_i \cdot P(x_i) \quad (2.26)$$

For a random variable \tilde{x} with a continuous density function $f(x)$, the expectation is given by

$$E(\tilde{x}) = \mu_x = \int_{-\infty}^{\infty} xf(x) dx \quad (2.27)$$

where $f(x)$ is the density function of \tilde{x} . The relation in equation (2.27) may be extended to a more general form when we consider the expectation of a function $g(\tilde{x})$ of a random variable \tilde{x} whose density function is $f(x)$. In this case

$$E(g(\tilde{x})) = \int_{-\infty}^{\infty} g(x)f(x) dx \quad (2.28)$$

The expression of equation (2.26) can be used to show how the name of mean of \tilde{x} comes about. If there are n possible values x_i of \tilde{x} , each having equal probability $P(x_i) = 1/n (= \text{constant})$, then the computed expectation μ_x from equation (2.26) becomes identical to the arithmetic mean of the n values x_i . If the probabilities $P(x_i)$ are not equal, it can be easily shown that equation (2.26) would provide the weighted mean of the n values, taking the weights as proportional to the probabilities of occurrence of each value x_i .

Equation (2.28) may be extended to cases of two or more dimensions. If $\tilde{\mathbf{x}} = [\tilde{x}_1 \tilde{x}_2]^t$ is a two-dimensional random vector with the corresponding joint density function $f(x_1, x_2)$, which is assumed to be continuous, then for any function $g(\tilde{x}_1, \tilde{x}_2)$ the expectation is defined as

$$E(g(\tilde{x}_1, \tilde{x}_2)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) \cdot f(x_1, x_2) dx_1 dx_2 \quad (2.29)$$

Similarly, for the general case of n dimensions

$$E(g(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_n) \cdot f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n \quad (2.30)$$

When working with expectations, several simple rules can be obtained as a consequence of equations (2.26) through (2.30). These rules are given here without proofs.

$$E(E(\tilde{x})) = E(\tilde{x}) \quad (2.31)$$

$$E(\tilde{x} + \tilde{y}) = E(\tilde{x}) + E(\tilde{y}) \quad (2.32)$$

$$E(c) = c \quad (2.33)$$

where c is constant,

$$E(c\tilde{x}) = cE(\tilde{x}) \quad (2.34)$$

where c is constant.

$$E(\tilde{x} \cdot \tilde{y}) = E(\tilde{x}) \cdot E(\tilde{y}) \quad (2.35)$$

if and only if \tilde{x} and \tilde{y} are *independent* random variables.

$$E(\tilde{x}^2) \neq (E(\tilde{x}))^2 \quad (2.36)$$

in general.

Example 2.1. If \tilde{x} , \tilde{y} , and \tilde{z} are three independent random variables and $\tilde{w} = 3\tilde{x} + 5\tilde{y}\tilde{z} - 2$, what is the expectation of \tilde{w} if we are given the expectations μ_x , μ_y , and μ_z .

Solution: Applying the relations above, we can then write

$$E(\tilde{w}) = \mu_w = 3\mu_x + 5\mu_y\mu_z - 2$$

Before leaving this section it is perhaps worth pointing out that the combined equations (2.31) through (2.34) express the so-called laws of propagation of the mean as will be explained later (see Section 4.3.2).

2.4.2. Variance Let $g(\tilde{x})$ be defined as

$$g(\tilde{x}) = (\tilde{x} - E(\tilde{x}))^2 = (\tilde{x} - \mu_x)^2 \quad (2.37)$$

The expectation of this special function $g(\tilde{x})$ is called the *variance* of the random variable \tilde{x} , and is given by

$$\text{var}(\tilde{x}) = \sigma_x^2 = E(g(\tilde{x})) = E[(\tilde{x} - E(\tilde{x}))^2] = \int_{-\infty}^{\infty} (x - \mu_x)^2 \cdot f(x) dx \quad (2.38)$$

in which $f(x)$ is the density function of \tilde{x} , assumed to be continuous. The square root of σ_x^2 , denoted by σ_x , is usually given the name *standard deviation*. The variance (or the standard deviation) is a measure of dispersion.

In view of the few rules of expectation given previously, an alternative form for the variance is possible, as follows:

$$\begin{aligned} \text{var}(\tilde{x}) &= \sigma_x^2 = E[(\tilde{x} - E(\tilde{x}))^2] = E[(\tilde{x} - \mu_x)^2] \\ &= E[\tilde{x}^2 - 2\tilde{x}\mu_x + \mu_x^2] \end{aligned}$$

$$\text{var}(\tilde{x}) = E(\tilde{x}^2) - 2\mu_x E(\tilde{x}) + E(\mu_x^2)$$

or

$$\sigma_x^2 = E(\tilde{x}^2) - \mu_x^2 = E(\tilde{x}^2) - (E(\tilde{x}))^2 \quad (2.38a)$$

2.4.3. Covariance and Correlation In a manner similar to defining the variance of one random variable, the covariance may be introduced when considering two random variables. Let the two random variables \tilde{x} and \tilde{y} have the joint density function $f(x, y)$, and define the function

$$h(\tilde{x}, \tilde{y}) = [(\tilde{x} - E(\tilde{x}))(\tilde{y} - E(\tilde{y}))] = [(\tilde{x} - \mu_x)(\tilde{y} - \mu_y)] \quad (2.39)$$

The covariance between \tilde{x} and \tilde{y} is defined as the expectation of the function $h(\tilde{x}, \tilde{y})$, or

$$\begin{aligned} \text{cov}(\tilde{x}, \tilde{y}) &= \sigma_{xy} = E[h(\tilde{x}, \tilde{y})] \\ &= E[(\tilde{x} - \mu_x)(\tilde{y} - \mu_y)] \end{aligned}$$

or

$$\sigma_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y)f(x, y) dx dy \quad (2.40)$$

provided that the distributions are continuous. Although the variance of a random variable expresses the variation of its distribution, the covariance describes the mutual variation of two random variables. It reflects their interrelationship or *mutual correlation* for which a correlation coefficient ρ is constructed as

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = E \left\{ \frac{(\tilde{x} - E(\tilde{x}))}{\sigma_x} \cdot \frac{(\tilde{y} - E(\tilde{y}))}{\sigma_y} \right\} \quad (2.41)$$

where σ_x and σ_y designate the standard deviations of the marginal distributions of \tilde{x} and \tilde{y} , respectively. The correlation between two random variables describes some interdependence between them. It should not, however, be confused with stochastic dependence or independence that is defined by the concept of conditional distributions [see equation (2.25)]. Correlation and statistical dependence are not the same, although both concepts are often used synonymously. It can be shown that the covariance σ_{xy} is always zero when the random variables \tilde{x} and \tilde{y} are statistically independent. But the reverse is not true in general. Zero covariance does not necessarily imply statistical independence. Nevertheless for multivariate *normal* probability distribution, zero covariance (lack of correlation) is a sufficient condition for statistical independence (see Hamilton, p. 31; in the bibliography at the back of the book).

2.4.4. Moments The concepts of mean (expectation), variance, and covariance are special cases of a more general concept termed *statistical moments*. The expected value of the general function $g(\tilde{x}) = (\tilde{x} - c)^k$, where c is a constant that may or may not be zero, is called the *statistical moment of the k th order* of the random variable \tilde{x} (or, in short, the k th moment of \tilde{x}) about c . Thus

$$\bar{m}_k = E((\tilde{x} - c)^k) \quad (2.42)$$

Using the two definitions of expectation given by equations (2.26) and (2.27), we get for the case of $c = 0$

$$\bar{m}_k = E(\tilde{x}^k) = \sum_{i=1}^n x_i^k P(x_i) \quad (2.43)$$

for a discrete random variable, and

$$\bar{m}_k = E(\tilde{x}^k) = \int_{-\infty}^{\infty} x^k f(x) dx \quad (2.44)$$

for the continuous case. The class of moments given by equation (2.42) includes a group called *central moments* which is of particular importance in practice. For this group the constant c takes the value

$$c = \mu_x = E(\tilde{x}) \quad (2.45)$$

hence central moments are expectations with respect to the mean, or

$$m_k = E[(\tilde{x} - E(\tilde{x}))^k] \quad (2.46)$$

If we now refer back to the definition of the variance in equation (2.38), we find that it is a special case of equation (2.46) in which $k = 2$. This is to say that the variance σ_x^2 of the *random variable* \tilde{x} is its *second central moment*, or $\sigma_x^2 = m_2$.

The different moments of a random variable \tilde{x} reflect the properties of its density function. When the density function is symmetric with respect to the first moment $\bar{m}_1 = E(\tilde{x}) = \mu_x$, all central moments of odd order vanish; that is, $m_k = 0$ for k odd. Conversely, when such odd-order central moments are not zero, their values reflect the degree of asymmetry or skewness of the density function.

As with the other concepts, moments may be extended to multi-dimensional random variables. Suppose that (\tilde{x}, \tilde{y}) is a two-dimensional random vector. The expected value of the general function $g(\tilde{x}^l, \tilde{y}^n)$ is called the moment of $(l + n)$ order. The corresponding central moment of the $(l + n)$ order is given by

$$\begin{aligned} m_{ln} &= E[(\tilde{x} - E(\tilde{x}))^l (\tilde{y} - E(\tilde{y}))^n] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(\tilde{x}))^l (y - E(\tilde{y}))^n f(x, y) dx dy \end{aligned} \quad (2.47)$$

for continuous distributions. For such a two-dimensional random vector there are three central moments of order two. Let \bar{m}_{10} and \bar{m}_{01} denote the first moments (that is, expectations) of the variables \tilde{x} and \tilde{y} , respectively. Then the following are the three central moments of second order

$$\begin{aligned} m_{20} &= E\{(\tilde{x} - \bar{m}_{10})^2\} = E((\tilde{x} - \mu_x)^2) = \sigma_x^2 = \text{variance of } \tilde{x} \\ m_{02} &= E\{(\tilde{y} - \bar{m}_{01})^2\} = E((\tilde{y} - \mu_y)^2) = \sigma_y^2 = \text{variance of } \tilde{y} \\ m_{11} &= E\{(\tilde{x} - \bar{m}_{10})(\tilde{y} - \bar{m}_{01})\} = E((\tilde{x} - \mu_x)(\tilde{y} - \mu_y)) = \sigma_{xy} \\ &= \text{covariance of } \tilde{x} \text{ and } \tilde{y} \end{aligned} \quad (2.48)$$

The central moments of order two, which according to equations (2.38) and (2.40) are identical to the variance and covariance, are of special importance for the statistical assessment of observations and for least squares adjustment.

In the general case of an n -dimensional random vector $\tilde{\mathbf{x}}$, second-order central moments can be set up in all combinations between the elements of the vector. The totality of these moments can be arranged in a matrix array, which is termed *the matrix of second central moments*, *the variance-covariance matrix*, or simply the *covariance matrix*. Its elements may be written symbolically as

$$\begin{aligned} M_{xx} &= \begin{bmatrix} m_{x_1x_1} & m_{x_1x_2} & \cdots & m_{x_1x_n} \\ m_{x_2x_1} & m_{x_2x_2} & \cdots & m_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{x_nx_1} & m_{x_nx_2} & \cdots & m_{x_nx_n} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_n} \\ \sigma_{x_2x_1} & \sigma_{x_2}^2 & \cdots & \sigma_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_nx_1} & \sigma_{x_nx_2} & \cdots & \sigma_{x_n}^2 \end{bmatrix} \end{aligned} \quad (2.49)$$

which is a square symmetric matrix, because

$$\begin{aligned} m_{x_ix_j} &= E\{(\tilde{x}_i - E(\tilde{x}_i))(\tilde{x}_j - E(\tilde{x}_j))\} \\ &= E\{(\tilde{x}_j - E(\tilde{x}_j))(\tilde{x}_i - E(\tilde{x}_i))\} = m_{x_jx_i} \end{aligned} \quad (2.50)$$

It should be mentioned, however, that \mathbf{M}_{xx} is square and symmetric only because it relates to the one n -dimensional random vector $\tilde{\mathbf{x}}$. If, on the other hand, we construct the cross-covariance matrix \mathbf{M}_{xy} for the n -dimensional vector $\tilde{\mathbf{x}}$ and m -dimensional vector $\tilde{\mathbf{y}}$, it will be an $n \times m$ rectangular matrix. All its elements will be covariances; that is, there will be no variances included. The symbol \mathbf{M} is used to denote the covariance matrix, to be consistent with the symbol m used for the individual elements. Later it will be convenient to use Σ instead, in keeping with the lower case σ being used for variances (σ^2) and covariances.

Example 2.2. (a) Derive the relationship between the central moments m_1 , m_2 , m_3 , and the moments \bar{m}_i (about zero).

Solution: By definition

$$\begin{aligned} m_1 &= E(\tilde{x} - E(\tilde{x})) = E(\tilde{x}) - E(E(\tilde{x})) = E(\tilde{x}) - E(\tilde{x}) = 0 \\ m_2 &= E[(\tilde{x} - E(\tilde{x}))^2] = E[(\tilde{x} - \bar{m}_1)^2] \\ &= E(\tilde{x}^2) - 2\bar{m}_1 E(\tilde{x}) + \bar{m}_1^2 \\ &= \bar{m}_2 - 2\bar{m}_1^2 + \bar{m}_1^2 \\ &= \bar{m}_2 - \bar{m}_1^2 \end{aligned}$$

This result is identical to that derived in equation (2.38a) (see Section 2.4.2).

$$\begin{aligned}
 m_3 &= E[(\tilde{x} - E(\tilde{x}))^3] = E[(\tilde{x} - \bar{m}_1)^3] \\
 &= \acute{E}(\tilde{x}^3) + 3\bar{m}_1^2 E(\tilde{x}) - 3\bar{m}_1 E(\tilde{x}^2) - \bar{m}_1^3 \\
 &= \bar{m}_3 + 3\bar{m}_1^3 - 3\bar{m}_1 \bar{m}_2 - \bar{m}_1^3 \\
 &= \bar{m}_3 - 3\bar{m}_1 \bar{m}_2 + 2\bar{m}_1^3
 \end{aligned}$$

(b) Derive the relationship between the central moment m_{11} and the moments \bar{m}_{01} , \bar{m}_{10} , and \bar{m}_{11} (about zero) for a case of two random variables.

Solution: Again, by definition

$$\begin{aligned}
 m_{11} &= E[(\tilde{x} - E(\tilde{x}))(\tilde{y} - E(\tilde{y}))] \\
 &= E[\tilde{x}\tilde{y} - \tilde{x}\bar{m}_{01} - \tilde{y}\bar{m}_{10} + \bar{m}_{01}\bar{m}_{10}] \\
 &= E(\tilde{x}\tilde{y}) - E(\tilde{x})\bar{m}_{01} - E(\tilde{y})\bar{m}_{10} + \bar{m}_{01}\bar{m}_{10} \\
 &= \bar{m}_{11} - \bar{m}_{10}\bar{m}_{01} - \bar{m}_{01}\bar{m}_{10} + \bar{m}_{01}\bar{m}_{10} \\
 &= \bar{m}_{11} - \bar{m}_{01}\bar{m}_{10}
 \end{aligned}$$

2.5. SOME OFTEN USED DISTRIBUTIONS

In connection with the theory of errors of observation and least squares adjustment there are a few (one-dimensional) distributions that are often used. Only continuous distributions are discussed, particularly those used for statistical testing.

2.5.1. The Gaussian or Normal Distribution The one-dimensional normal distribution is the most frequently used distribution in statistical theory and application. Its density function is given by

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu_x)^2}{2\sigma_x^2} \right\} \quad (2.51)$$

In equation (2.51) there are two parameters that specify the distribution (Figure 2.3):

μ_x = expectation or mean of \tilde{x}

and

σ_x = standard deviation of $\tilde{x} = +\sqrt{\text{var}(\tilde{x})} = \sqrt{\sigma_x^2}$

The cumulative normal distribution function of the standardized random variable,

$$z = \frac{x - \mu_x}{\sigma_x} \quad (2.52)$$

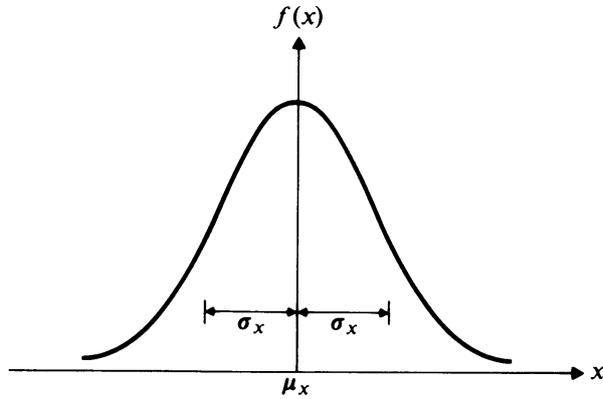


Figure 2.3. Normal Distribution Density Function

(with zero mean and unit standard deviation) is given by

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{u^2}{2}\right) du \quad (2.53)$$

Appendix Table D.1 gives the values of $[1 - F(z)]$. A few remarks may point out the essential features of the normal distribution.

1. The normal density function $f(x)$ is symmetric about the mean μ_x . Therefore all odd central moments are zero. Also the median[†] and the mode,[‡] which are two parameters of location sometimes used in practice, are equal to the mean μ_x .
2. The maximum density value for the standardized variable is 0.399.
3. The density function approaches zero asymptotically as x goes to $\pm\infty$.
4. The density function has two points of inflection at $x = \mu_x \pm \sigma_x$.
5. The probability for \tilde{x} taking values within x_1 and x_2 is given by the area between the x axis, the density function curve, and the boundaries of the interval $x = x_1, x = x_2$. In particular the probabilities for deviation from the mean within some multiples of σ are as follows:

$$\begin{aligned} P[-\sigma_x < x - \mu_x < +\sigma_x] &= 0.6827 \\ P[-2\sigma_x < x - \mu_x < +2\sigma_x] &= 0.9545 \\ P[-3\sigma_x < x - \mu_x < +3\sigma_x] &= 0.9973 \end{aligned} \quad (2.54)$$

[†] The median x_m is defined for the continuous distribution such that

$$\int_{-\infty}^{x_m} f(x) dx = \frac{1}{2}$$

Thus x_m is the abscissa through which a vertical line divides the area under the density distribution into two equal parts.

[‡] The mode, another measure of location or central tendency, is for continuous distributions the value of the random variable at which the probability density function has a relative maximum.

6. The abscissae associated with intervals covering probabilities of 0.90, 0.95, and 0.99 are

$$\begin{aligned} P[-1.645\sigma_x < x - \mu_x < +1.645\sigma_x] &= 0.90 \\ P[-1.960\sigma_x < x - \mu_x < +1.960\sigma_x] &= 0.95 \\ P[-2.576\sigma_x < x - \mu_x < +2.576\sigma_x] &= 0.99 \end{aligned} \quad (2.55)$$

7. The probability that x takes on values on either side of μ_x (that is, either larger or smaller than μ_x) is equal to 0.5.

The theoretical and practical importance of the normal distribution is due to the "central limit theorem" which states that the sum $\sum_{i=1}^n \tilde{x}_i$ of n independent variables $\tilde{x}_1, \dots, \tilde{x}_n$ will be asymptotically normally distributed as $n \rightarrow \infty$. In practical applications, normal distributions are encountered very often. In particular, random variables that represent measurements in photogrammetry, geodesy, or surveying are often nearly normally distributed.

2.5.2. The t (Student) Distribution The t distribution is used in connection with sampling (testing hypotheses using sample statistics instead of population parameters).

Let $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ be n independent stochastic (random) variables of identical normal distribution with mean μ and standard deviation σ . Then the random variable \tilde{t} is defined by

$$t = \frac{\bar{x} - \mu}{s} \sqrt{n} \quad (2.56)$$

where

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k \quad \text{and} \quad s^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$$

The statistics \bar{x} and s refer to sample mean and sample standard deviation, respectively (see Equation (2.89) and (2.92)). The density function of t is

$$f(t) = \frac{1}{\sqrt{(n-1)\pi}} \frac{\Gamma(n/2)}{\Gamma[(n-1)/2]} \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}} \quad (2.57)$$

Γ = gamma function

$$\Gamma(n) = \int_0^{\infty} e^{-t} t^{n-1} dt$$

$$\Gamma(n+1) = n \cdot \Gamma(n) \quad \text{and} \quad \Gamma(1) = 1 \quad (2.58)$$

For whole, positive numbers n ,

$$\Gamma(n) = (n - 1)! \quad \text{or} \quad \Gamma(n + 1) = n!$$

Appendix D Table D.2 lists the values of $[1 - F(t)]$. The t distribution, as defined in equation (2.57) is said to have $(n - 1)$ degrees of freedom. The density function is symmetrical with respect to the mean. With increasing degrees of freedom the t distribution approaches the normal distribution with mean 0 and variance 1. For $n \geq 30$ the t distribution can be replaced by the normal distribution.

2.5.3. The χ^2 Distribution Let $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ be n independent stochastic (random) variables, each of which is normally distributed with mean 0 and variance 1. Then the sum of their squares yields a random variable for which the symbol χ^2 is customarily used

$$\chi_n^2 = \tilde{x}_1^2 + \tilde{x}_2^2 + \dots + \tilde{x}_n^2 \quad (2.59)$$

This random variable has a χ^2 distribution (chi square) of n “degrees of freedom.”

The density function of the random variable χ^2 is

$$f(\chi_n^2) = f(x) = c_n \cdot x^{(n-2)/2} \cdot e^{-x/2} \quad \text{for } x > 0 \quad (2.60)$$

and

$$f(x) = 0 \quad \text{for } x \leq 0$$

where

$$c_n = \frac{1}{2^{n/2} \cdot \Gamma\left(\frac{n}{2}\right)}$$

$n =$ degrees of freedom

The χ^2 distribution has particularly simple central moments as follows:

Mean:

$$E(\chi_n^2) = n \quad (2.61)$$

Variance:

$$\text{var}(\chi_n^2) = \sigma^2(\chi_n^2) = 2n \quad (2.62)$$

In Appendix Table D.3 the values of $[1 - F(\chi^2)]$ are given for several degrees of freedom. The graphical representations in Figure 2.4 show that χ^2 density functions are markedly skew for small degrees of freedom. Beyond $n \sim 10$, however, they approach more and more a normal distribution. The asymptotic convergence for $n \rightarrow \infty$ approaches normal density function with mean n and variance equal to $2n$.

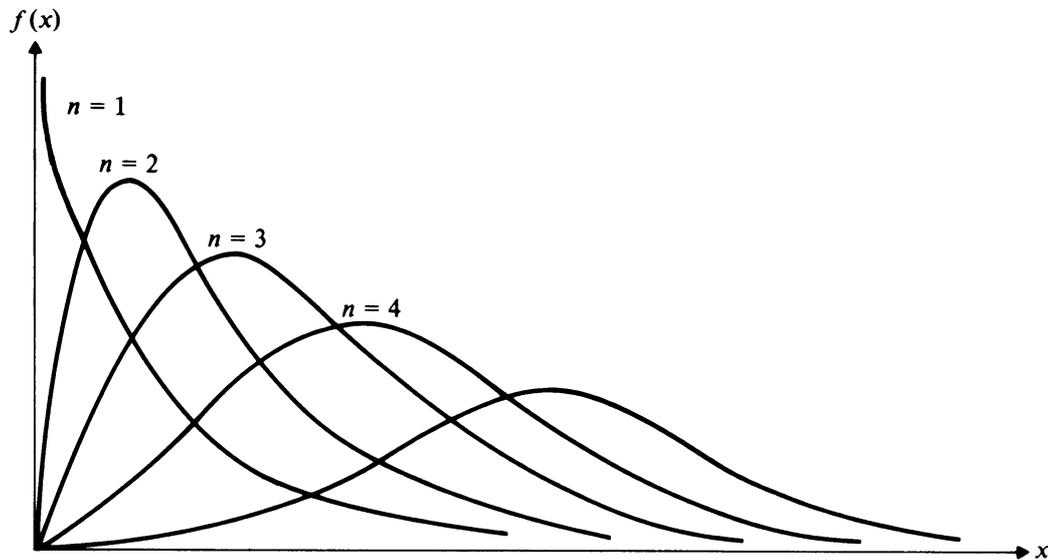


Figure 2.4

2.5.4. The F (Fisher) Distribution Let $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_m$ and $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n$ be two sets of independent, normally distributed, normalized random variables, each with mean 0 and variance 1. Consider the sum of the squares of each set,

$$\tilde{\chi}_m^2 = \tilde{x}_1^2 + \tilde{x}_2^2 + \dots + \tilde{x}_m^2 = \sum_{i=1}^m \tilde{x}_i^2 \quad (2.63)$$

$$\tilde{\chi}_n^2 = \tilde{y}_1^2 + \tilde{y}_2^2 + \dots + \tilde{y}_n^2 = \sum_{j=1}^n \tilde{y}_j^2 \quad (2.64)$$

The random variable F ,

$$F_{m,n} = \frac{\tilde{\chi}_m^2/m}{\tilde{\chi}_n^2/n} \quad (2.65)$$

is said to have an F distribution of m and n degrees of freedom. The density function $f_{m,n}(u)$ and the cumulative distribution function $F_{m,n}(u)$ are, for $u > 0$, given by

$$f_{m,n}(u) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2) \cdot \Gamma(n/2)} \cdot \left(\frac{m}{n}\right)^{m/2} \frac{u^{(m-2)/2}}{[1 + (m/n)u]^{(m+n)/2}} \quad (2.66)$$

$$F_{m,n}(u) = \int_0^u f_{m,n}(t) dt \quad (2.67)$$

The value of $[1 - F]$ (where F is a cumulative function) for different degrees of freedom are given in Appendix Tables D.4(a) through D.4(c). The values

of the expectation (mean) and the variance of an F -distributed random variable, with m, n degrees of freedom are

$$\begin{aligned} \text{mean} &= \frac{n}{n-2} \quad \text{for } n > 2 \\ \text{variance} &= \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad \text{for } n > 4 \end{aligned} \quad (2.68)$$

The F distribution with m, n degrees of freedom approaches asymptotically for $n \rightarrow \infty$ the χ^2 distribution with m degrees of freedom.

The practical application of the F distribution in least squares adjustment is concerned with the comparison of variances such as those obtained from two adjustments. In some particular cases the comparison may be between a variance obtained from the adjustment and an a priori given reference variance. This case refers to the $F_{m, \infty} = \chi_m^2$ distribution.

2.6. MULTINORMAL DISTRIBUTION

Although least squares theory of adjustment does not require a specified distribution, most of the statistical testing following the adjustment is concerned with multinormal distribution of random vectors with density function

$$\begin{aligned} f(x_1, \dots, x_n) = f(\mathbf{x}) &= \left[\frac{1}{(2\pi)^{n/2} \sqrt{|\Sigma|}} \right] \\ &\times \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_x)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_x) \right] \end{aligned} \quad (2.69)$$

with mean vector $\boldsymbol{\mu}_x$ and covariance matrix Σ . From equation (2.69) the expanded form for the two-dimensional normal distribution of a random vector (\tilde{x}, \tilde{y}) is defined by the density function,

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi \sqrt{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2}} \exp \left\{ -\frac{\sigma_1^2 \sigma_2^2}{2(\sigma_1^2 \sigma_2^2 - \sigma_{12}^2)} \right. \\ &\quad \left. \left[\frac{(x - \mu_1)^2}{\sigma_1^2} - 2\sigma_{12} \frac{(x - \mu_1)(y - \mu_2)}{\sigma_1^2 \sigma_2^2} + \frac{(y - \mu_2)^2}{\sigma_2^2} \right] \right\} \end{aligned} \quad (2.70)$$

where μ_1, μ_2 are the means, σ_1^2, σ_2^2 are the variances for \tilde{x} and \tilde{y} , respectively, and σ_{12} is the covariance.

The graphical representation of the two-dimensional normal density function in Figure 2.5 demonstrates that its intersection with any vertical plane $y = ax + b$ will be a one-dimensional Gauss function, that is, a one-dimensional normal density function. Any plane parallel to the x, y plane and cutting the bell-shaped surface intersects it in an elliptical curve.

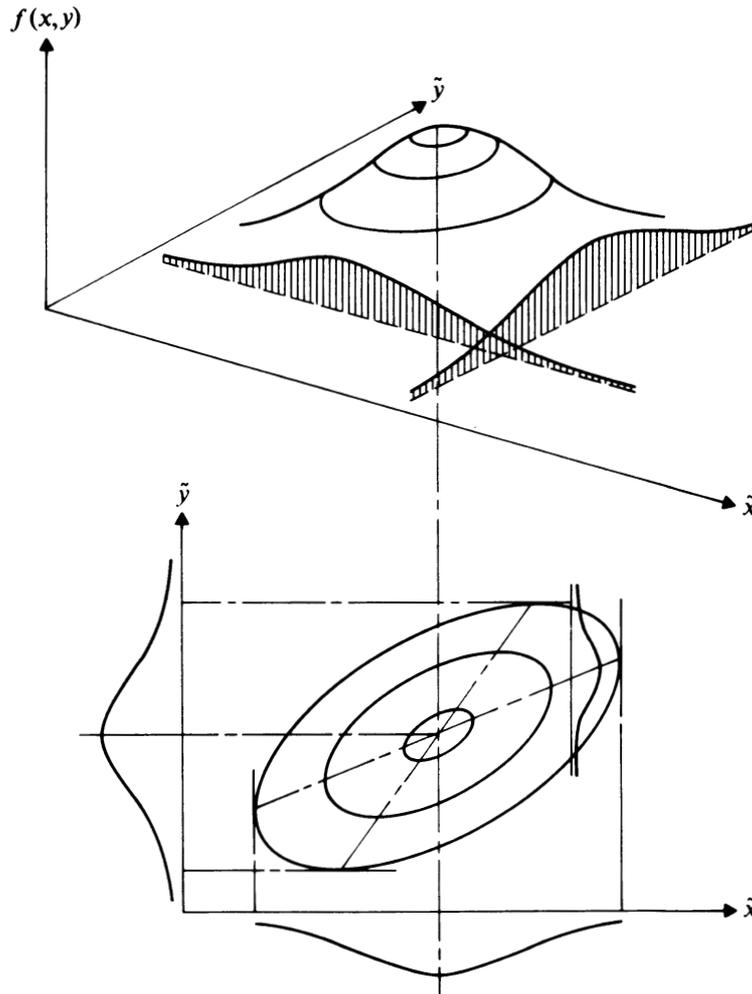


Figure 2.5

The conditional density function of y given x can be considered as that of a one-dimensional normal distribution

$$f(y|x) = \frac{1}{\bar{\sigma}\sqrt{2\pi}} \exp \left\{ -\frac{(y - \mu(x))^2}{2\bar{\sigma}^2} \right\} \quad (2.71)$$

the mean of which is a function of x ,

$$\mu(x) = \mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1) = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(x - \mu_1) \quad (2.72)$$

The variance in equation (2.71) can be seen to be

$$\bar{\sigma}^2 = \sigma_2^2(1 - \rho^2) = \frac{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}{\sigma_1^2} \quad (2.73)$$

Thus the mean (or expectation) of the conditional distribution of y for given values of x is a linear function of x . This function $\mu(x)$ is called the *regression*

line of y on x . This provides for an interpretation of correlation and of regression between two random variables. In case of correlation the regression line is the functional dependence of the expectation of the conditional distribution of y on x and similarly for the other regression line.

2.6.1. Ellipses of Constant Probability Referring to equation (2.70) of the two-dimensional normal distribution, the properties of the density function $f(x, y)$ can also be geometrically interpreted by intersecting the surface $f(x, y)$ by horizontal planes $f(x, y)$ equal to constant. The resulting curves of intersection lines form a family of ellipses (because $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < \rho < 1$), the equation of which is

$$h(x, y) = \frac{1}{(1 - \rho^2)} \left[\frac{(x - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x - \mu_1)(y - \mu_2)}{\sigma_1 \sigma_2} + \frac{(y - \mu_2)^2}{\sigma_2^2} \right] = k^2 \quad (2.74)$$

The common center of this family of ellipses is defined by the expectations μ_1 and μ_2 (of the marginal distributions of \tilde{x} and \tilde{y}) as coordinates. For simplicity in the analysis to follow, we shift the origin of the coordinate system to the point of common ellipses' center (μ_1, μ_2) .

The equation $h(x, y) = k^2$ in equation (2.74), for a specific value of k , is an ellipse bounded by a rectangle having the dimensions $2k\sigma_1$ in the direction of x , and $2k\sigma_2$ in the direction of y , as shown in Figure 2.6. Such an ellipse is

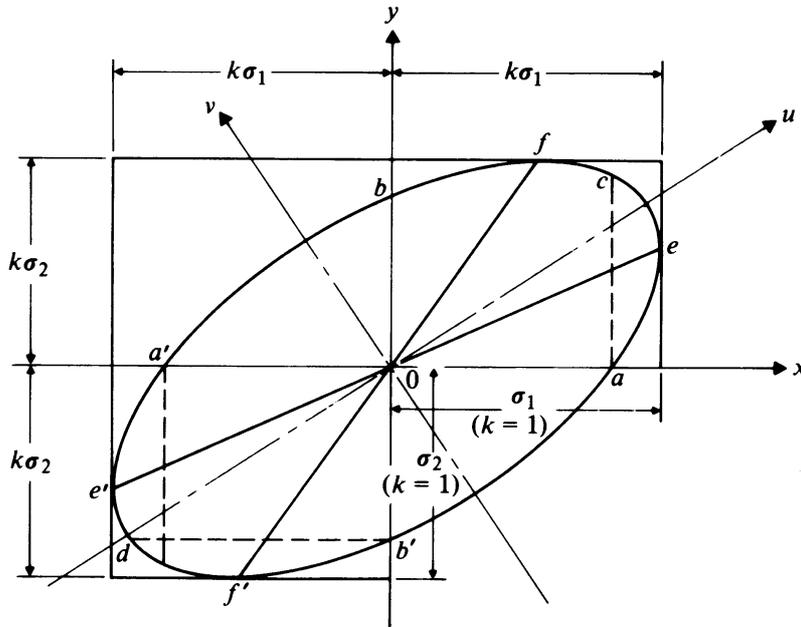


Figure 2.6. The Standard Ellipse. Relations for $k = 1$: $x_a = -x_{a'} = \sigma_1 \sqrt{1 - \rho^2} = \sigma(x|y)$; $y_b = -y_{b'} = \sigma_2 \sqrt{1 - \rho^2} = \sigma(y|x)$; $y_c = 2\rho\sigma_2 \sqrt{1 - \rho^2}$; $x_d = 2\rho\sigma_1 \sqrt{1 - \rho^2}$; $x_e = -x_{e'} = \sigma_1 y_e = -y_{e'} = \rho\sigma_2$; $x_f = x_{f'} = \rho\sigma_1 y_f = -y_{f'} = \sigma_2$

known as an *ellipse of constant probability*; the value of such probability depends on the selected value of k .

In the shifted coordinate system considering the ellipse with $k = 1$ as a representative ellipse, its equation becomes

$$\left(\frac{x}{\sigma_1}\right)^2 - 2\rho\left(\frac{x}{\sigma_1}\right)\left(\frac{y}{\sigma_2}\right) + \left(\frac{y}{\sigma_2}\right)^2 = (1 - \rho^2) \quad (2.75)$$

This particular ellipse is named the *standard ellipse* and its shape is determined by σ_1 , σ_2 , and ρ (see Figure 2.6).

The points of intersection with the x axis are a and a' with $x = \pm\sigma_1\sqrt{1-\rho^2}$, and with the y axis they are b and b' with $y = \pm\sigma_2\sqrt{1-\rho^2}$. These values can be easily obtained from equation (2.75) by setting $y = 0$ and $x = 0$, respectively. Consequently, they represent the square roots of the corresponding conditional second moments $m(x|y=0)$ and $m(y|x=0)$. The vertical tangents to the ellipse are at points e and e' with coordinates $(\sigma_1, \rho\sigma_2)$ and $(-\sigma_1, -\rho\sigma_2)$, whereas the horizontal tangents are at points f and f' with corresponding coordinates $(\rho\sigma_1, \sigma_2)$ and $(-\rho\sigma_1, -\sigma_2)$, respectively.

From equation (2.75) and well-established relations for the geometry of an ellipse, the semimajor and semiminor axes may be computed from

$$a^2 = \frac{1}{2}(\sigma_1^2 + \sigma_2^2) + \sqrt{\frac{1}{4}(\sigma_1^2 - \sigma_2^2)^2 + \sigma_{12}^2} \quad (2.76)$$

$$b^2 = \frac{1}{2}(\sigma_1^2 + \sigma_2^2) - \sqrt{\frac{1}{4}(\sigma_1^2 - \sigma_2^2)^2 + \sigma_{12}^2} \quad (2.77)$$

where ρ in equation (2.75) was replaced by its value $\sigma_{12}/\sigma_1\sigma_2$ in terms of the covariance σ_{12} . The values of a and b obtained from equations (2.76) and (2.77) can be readily shown to be the square roots of the eigenvalues of the covariance matrix,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

The characteristic polynomial of Σ is evaluated as

$$(-\lambda)^2 + \text{tr}(\Sigma)(-\lambda) + |\Sigma| = 0$$

where

$$\text{tr}(\Sigma) = \text{trace of the covariance matrix} = \sigma_1^2 + \sigma_2^2$$

$$|\Sigma| = \text{determinant of the covariance matrix} = \sigma_1^2\sigma_2^2 - \sigma_{12}^2$$

Thus the polynomial becomes

$$\lambda^2 - (\sigma_1^2 + \sigma_2^2)\lambda + (\sigma_1^2\sigma_2^2 - \sigma_{12}^2) = 0 \quad (2.78)$$

The roots of equation (2.78) are those given by equations (2.76) and (2.77), which verifies the relationship between the eigenvalues of Σ and the semimajor and semiminor axes of the standard ellipse.

The angle γ between the semimajor axis of the ellipse and the x axis is obtained from the following relationship:

$$\tan 2\gamma = \frac{2\sigma_{12}}{\sigma_1^2 - \sigma_2^2} \quad (2.79)$$

The proper quadrant for 2γ is determined such that $\sin 2\gamma$ has the same sign as σ_{12} and $\cos 2\gamma$ has the same sign as $(\sigma_1^2 - \sigma_2^2)$.

Referring to Figure 2.6, we denote by u and v a rotated coordinate system, the axes of which coincide with the semimajor and semiminor axes of the ellipse. The system u, v is obtained from the original system x, y by a simple rotation through the angle γ . With respect to this new system, u, v , we may refer two new random variables \tilde{u}, \tilde{v} with the marginal standard deviations $\bar{\sigma}_1, \bar{\sigma}_2$. From the foregoing discussion and Figure 2.6 it becomes clear that $\bar{\sigma}_1 = a, \bar{\sigma}_2 = b$, and that the new random variables \tilde{u}, \tilde{v} are independent (uncorrelated), that is, $\bar{\sigma}_{12} = 0$. Consequently, through that transformation from x, y to u, v , it was possible to replace a pair of correlated random variables by another pair that is not correlated. This possibility is in fact general; it is always possible to replace a set (of any dimension) of correlated random variables by another set (of the same dimension) of uncorrelated random variables. The transformation matrix can be constructed by determining the normalized eigenvectors of the covariance matrix of the original set of random variables and using them as columns. This is equivalent to diagonalizing the original covariance matrix since the resulting covariance matrix of the new uncorrelated set of random variables is always in diagonal form. The elements of this diagonal matrix are the eigenvalues of the original covariance matrix. This concept is explained further in Appendix A (Section A.6).

Considering further the case of no correlation ($\rho = 0$), the density function of equation (2.70) can be written in the form,

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp \left\{ -\frac{1}{2} \left[\frac{(x - \mu_1)^2}{\sigma_1^2} + \frac{(y - \mu_2)^2}{\sigma_2^2} \right] \right\} \\ &= \left\{ \frac{1}{\sigma_1\sqrt{2\pi}} \exp \left[-\frac{(x - \mu_1)^2}{2\sigma_1^2} \right] \right\} \\ &\quad \times \left\{ \frac{1}{\sigma_2\sqrt{2\pi}} \exp \left[-\frac{(y - \mu_2)^2}{2\sigma_2^2} \right] \right\} \\ &= f_1(x) \cdot f_2(y) \end{aligned}$$

where $f_1(x)$ and $f_2(y)$ are one-dimensional normal density functions with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 , respectively. Thus for no correlation, or when $\rho = 0$, the joint density function $f(x, y)$ can be written as the product of the marginal density functions for x and y . Referring to equation (2.21) we can directly conclude that for a normal distribution, the case of no

correlation is identical to the case of statistical independence. Perhaps it is this fact that sometimes leads to the use of the terms “uncorrelated” and “independent” synonymously. This is obviously true *only* when the random variables in question *are normally distributed*.

Having defined the standard ellipse [see equation (2.75)], it is now of interest to find out the probability that the random vector (\tilde{x}, \tilde{y}) takes values within the ellipse [applying in principle equation (2.12)]. The derivation of that probability is simplified if we work with (\tilde{u}, \tilde{v}) , instead of (\tilde{x}, \tilde{y}) , as they are uncorrelated and lie along the principal axes of the ellipse. If the normalized variables $\tilde{u}/\bar{\sigma}_1$ and $\tilde{v}/\bar{\sigma}_2$ are used, the sum of their squares has, according to equation (2.59), a χ^2 distribution with two degrees of freedom. The general expression for the probability of being within or on an ellipse with axes $k\bar{\sigma}_1$ and $k\bar{\sigma}_2$ is

$$P\left\{\left[\frac{u^2}{\bar{\sigma}_1^2} + \frac{v^2}{\bar{\sigma}_2^2}\right] < k^2\right\} = P\{\chi_2^2 < k^2\} = 1 - \alpha \quad (2.80)$$

For the standard ellipse $k = 1$, and the value of $(1 - \alpha)$ is 0.3935, so the probability of falling inside the standard ellipse is 0.3935. In order to establish confidence regions (see Section 11.5.2), we select the α level and compute the multiplier k . For example, for $\alpha = 0.05$

$$P\{\chi_2^2 < \chi_{0.05, 2}\} = P\{\chi_2^2 < 5.99\} = 0.95$$

from which $k = \sqrt{5.99} = 2.447$. Consequently, the 95% ellipse has semiaxes of $a = 2.447\bar{\sigma}_1$ and $b = 2.447\bar{\sigma}_2$. Table 2.1 gives other typical values.

TABLE 2.1

P	0.394	0.500	0.900	0.950	0.990
k	1.000	1.177	2.146	2.447	3.035

The reader should note that for each one-dimensional marginal normal distribution, the probability that each variable (\tilde{x} or \tilde{y}) lies in the region within plus and minus one standard deviation ($\pm\sigma_1$ or σ_2) is 0.683 [see equation (2.54)]. By contrast the probability for the joint event, which is falling within the standard ellipse, is considerably less, being only 0.394.

2.6.2. Radial Standard Deviation in Two Dimensions In many cases we are interested in one measure (for example, standard deviation) instead of the two, σ_x and σ_y , in the two-dimensional estimation. If \tilde{x} , \tilde{y} are two *independent* random variables, a new variable $\tilde{r} = \sqrt{\tilde{x}^2 + \tilde{y}^2}$ can be defined and its distribution derived from the joint distribution of \tilde{x} and \tilde{y} . The value r that \tilde{r} takes can be determined given the selected values for \tilde{x} and \tilde{y} . For example, considering the case of standard ellipse for which the probability is 0.394, the

value of r obtained from $P[\tilde{r} < r] = 0.394$ is called “circular standard error” and is given the symbol σ_c . In another case $P[\tilde{r} < r] = 0.500$, r is called the “circular probable error” (or CPE). The values of both σ_c and the circular probable error depend on the relative magnitudes of σ_x and σ_y . For practical reasons we may use the following approximation for σ_c :

$$\sigma_c \cong \frac{1}{2}(\sigma_x + \sigma_y) \quad (2.81)$$

particularly for $0.5 \leq (\sigma_y/\sigma_x) \leq 1.0$ (with $\sigma_x \geq \sigma_y$). Similarly, an approximation for the circular probable error is

$$\text{CPE} \cong 0.589(\sigma_x + \sigma_y) \quad (2.82)$$

especially for $0.2 \leq (\sigma_y/\sigma_x) \leq 1.0$ (when $\sigma_x \geq \sigma_y$).

Another circular measure of precision is termed the “mean square positional error” (MSPE) evaluated from

$$\text{MSPE} = \sqrt{\sigma_x^2 + \sigma_y^2} \quad (2.83)$$

$P[\tilde{r} < \text{MSPE}]$ varies with the ratio (σ_y/σ_x) . When $\sigma_x = \sigma_y$, then $\text{MSPE} = \sqrt{2}\sigma_x$ and $P[\tilde{r} \leq \text{MSPE}] = 0.632$. This precision index (the mean square positional error) is not recommended for use because of the variation in probability.

2.6.3. Ellipsoids of Constant Probability Referring to equation (2.69), the function $(\mathbf{x} - \boldsymbol{\mu}_x)' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x) = k^2$ (which is a positive definite quadratic form) represents a family of hyperellipsoids of constant probability. The case of three dimensions is important because it is often encountered when applying statistical concepts to the determination of point coordinates (x, y, z) through photogrammetric or geodetic techniques. In this case the ellipsoid equation becomes (assuming for simplicity $\boldsymbol{\mu}_x = \mathbf{0}$)

$$h(x, y, z) = [x \quad y \quad z] \boldsymbol{\Sigma}^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = k^2 \quad (2.84)$$

and when $k = 1$, it is called the *standard ellipsoid*. The semiaxes of the ellipsoid (a, b, c) are determined by diagonalizing $\boldsymbol{\Sigma}$ by writing

$$\begin{bmatrix} \sigma_u^2 & & \\ & \sigma_v^2 & \\ & & \sigma_w^2 \end{bmatrix} = \begin{bmatrix} a^2 & & \\ & b^2 & \\ & & c^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} = \mathbf{T}' \boldsymbol{\Sigma} \mathbf{T} \quad (2.85)$$

where \mathbf{T} is an orthogonal matrix whose columns are the normalized eigenvectors of $\boldsymbol{\Sigma}$; λ_1 , λ_2 , and λ_3 are the eigenvalues of $\boldsymbol{\Sigma}$; and the u, v, w is a rotated coordinate system such that the random variables in the directions of its axes are uncorrelated. In a manner similar to the two-dimensional case, the statistical significance of the ellipsoids may be derived. Thus for the

probability of the point falling inside or on the ellipsoid defined by $a = k\sigma_u$, $b = k\sigma_v$, $c = k\sigma_w$ the expression is

$$P\left[\left(\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} + \frac{w^2}{\sigma_w^2}\right) < k^2\right] = P[\chi_3^2 < k^2] = 1 - \alpha$$

For the standard ellipsoid $(1 - \alpha) = 0.199$, which is obtained from χ^2 with three degrees of freedom. Furthermore, Table 2.2 gives different values of the probability and k .

TABLE 2.2

P	0.199	0.500	0.900	0.950	0.990
k	1.000	1.538	2.500	2.700	3.368

The reader should note that for one dimension ($n = 1$) the probability for the standard region is 0.683, for two dimensions it is 0.394, and for three dimensions it is 0.199. Table 2.3 gives the values for several other dimensions.

TABLE 2.3

n	1	2	3	4	5	6
P	0.683	0.394	0.199	0.090	0.037	0.014

When seeking a single measure for three-dimensional cases, the spherical standard error σ_s is defined as $P[\tilde{s} < \sigma_s] = 0.199$ and is approximated by

$$\sigma_s \cong \frac{1}{3}(\sigma_x + \sigma_y + \sigma_z) \quad (2.86)$$

for $0.35 \leq (\sigma_z/\sigma_x) \leq 1.0$ ($\sigma_x \geq \sigma_y \geq \sigma_z$). Similarly, spherical probable error (SPE) is given by $P[\tilde{s} < \text{SPE}] = 0.500$ and is approximated by

$$\text{SPE} \cong 0.513(\sigma_x + \sigma_y + \sigma_z) \quad (2.87)$$

for $0.35 \leq (\sigma_z/\sigma_x) \leq 1.0$.

Finally, the mean radial spherical error (MRSE) is

$$\text{MRSE} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} \quad (2.88)$$

and if $\sigma_x = \sigma_y = \sigma_z$, then

$$\text{MRSE} = \sqrt{3}\sigma_x \quad \text{and} \quad P[\tilde{s} \leq \text{MRSE}] = 0.608$$

Since the probability varies with the different values of σ_x , σ_y , and σ_z , the mean radial spherical error is not suitable for use in practice.

Summarizing the discussion of the properties of two- and more-dimensional normal distributions, the basic concept of correlation should be

emphasized. Correlation describes a certain dependence between x and y values of the random variables \tilde{x} and \tilde{y} . The dependence is not strict (functional), however. It holds strictly (functionally) only for the expectations of both random variables. Their mutual dependence is given by regression lines. There are always two of them, since the expectation of y as a function of x is different from the expectation of x as a function of y .

Thus, in general terms, correlation for actual values of x and y expresses only a trend relationship, which is not strictly valid for individual pairs of values. The higher the correlation, however, the closer would be the tendency to have the regression relationship realized.

2.7. SAMPLING, ESTIMATION, AND CONFIDENCE MEASURES

In statistics a physical situation or process is assessed conceptually by a (hypothetical) functional model. The elements of that model are considered as random variables, associated with joint probability distributions. Statistical events (for instance, measurements) are *considered* realizations of the random variables. The realizations, whether repeated or not, are always limited in extent. They are called *samples* from the probability distributions.

From the sample values, the probability distributions of the random variables, or at least some of their parameters, are to be assessed or estimated. The method of least squares, for instance, is a method for estimating expectations of random variables from samples (measurements).

Related to estimation is the task of determining the reliability or accuracy with which the estimates are obtained (confidence measures). Furthermore, statistical testing treats the question of whether the results of estimations are in agreement with the initial assumptions (hypotheses).

Those concepts are increasingly linked to the application of the least squares method. Therefore introductory essentials to these topics are presented next.

2.7.1. Sampling Any set of measurements is considered the realization of the sample (x_1, x_2, \dots, x_n) of the random vector \tilde{x} . The purpose of sampling is (in this context) to derive estimates for the probability distributions or some of their parameters. These distributions may be of the random variables concerned, or of some other random variables derived from the ones to which the measurements refer directly.

When repeated observations are taken, some factors that are commonly used to assess the sample data are as follows:

1. Frequency distributions.
2. Sample statistics for location (mean, median, mode, or midrange).
3. Sample statistics for dispersion (variances and covariances).
4. Other sample moments.

FREQUENCY DISTRIBUTIONS Establishing frequency distributions makes sense only if the size of the sample is relatively large, in particular for multidimensional frequency distributions. They are obtained by classifying data according to some appropriate subdivision and counting the events of each class, thus yielding class frequencies. In practice, relative frequencies are generally used. The table of class frequencies gives the frequency distribution, which can be graphically displayed by either a bar diagram or more frequently by a histogram.

BAR DIAGRAM AND HISTOGRAM As an example of a one-dimensional case, we shall consider measuring a distance (which is approximately 50 meters), say, 100 times, and recording the values to the nearest 4 mm. We shall regard any measurement whose last two digits fall between, say, 62 and 66 as 64 mm, that between 66 and 70 as 68 mm, and so on. (The actual values of the length would be 49,964 mm and 49,968 mm, but the digits 499 are the same for all readings and will be dropped). We may represent this sample of measurements by a bar diagram using the values 64, 68, 72, and so on, as the points along the horizontal axis. The number of measurements belonging to each of these values and the ratio between such numbers and the total number of measurements (100) which is the relative frequency are given in Table 2.4. The graph itself is shown in Figure 2.7, where $f(x)$ represents the values of the relative frequency of occurrence of a certain measurement value. In this presentation the approximation was made to take the middle value of an interval as representing a certain value of the observation. If we eliminate this approximation, we can replace each of the vertical lines by a rectangle that covers the appropriate interval and whose height is proportional to the relative frequency. In this case the probability of a certain *interval*, which is estimated as the ratio of measurements within the interval and the total number of measurements, is equal to the *area* of the rectangle constructed over it. In this manner the sum of the areas of all the rectangles will equal to 1, thus satisfying the requirement of having all probabilities adding up to 1. Such a frequency distribution is called a *histogram*. The

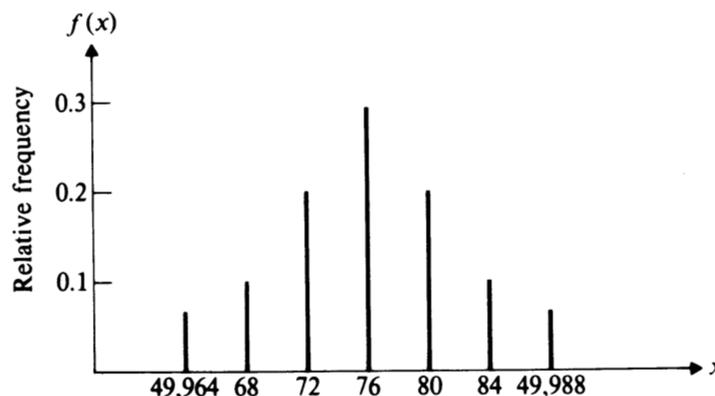


Figure 2.7

TABLE 2.4

Nominal Values	64	68	72	76	80	84	88	
Interval Range	62-66	66-70	70-74	74-78	78-82	82-86	86-90	
Number in Each Interval	6	10	20	28	20	10	6	$\Sigma = 100$
Relative Frequency	0.06	0.10	0.20	0.28	0.20	0.10	0.06	$\Sigma = 1.0$
Height of Rectangles in Histogram	0.015	0.025	0.05	0.07	0.05	0.025	0.015	

heights of the rectangles are given in the fifth line of Table 2.4 (which are one fourth of the values in the fourth row because the interval width is 4 mm) and the histogram itself is depicted in Figure 2.8.

BIDIRECTIONAL HISTOGRAM (STEREOGRAM) As an example for a two-dimensional case, we shall consider an example of a target chart grouping. Every point on the target chart represents one shot and its position may be determined by a pair of cartesian coordinates x and y with origin at the center of the target board. The behavior of the position of the points follows the laws of probability, in this case, a *bivariate* distribution with \tilde{x} and \tilde{y} as the random variables or variates. We shall assume that one person has taken 400 shots at the practice board. A grid made up of 3-cm squares is then superimposed on the target sheet and the number of points in each square are counted and entered in the upper half of the square as shown in Table 2.5. This table represents a two-dimensional frequency distribution of the rifle shots.

Table 2.5 can also serve as the base for constructing a stereogram (a bidirectional histogram) by erecting square columns over each square. The volume of each column is equal to the relative frequency of the number of

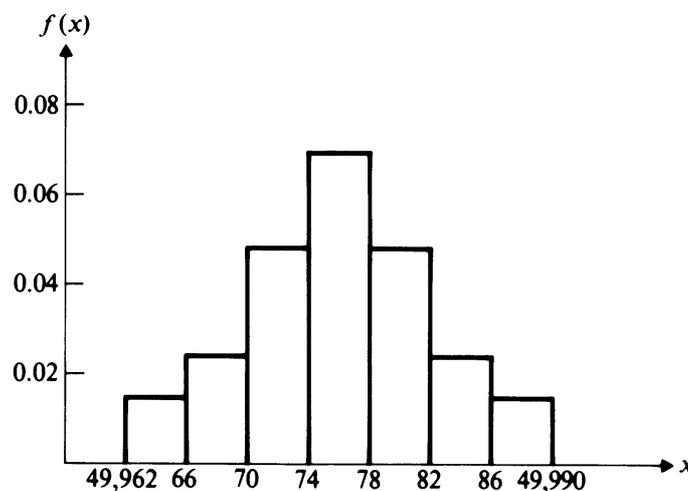


Figure 2.8

TABLE 2.5

		THE X DIRECTION (cm)															MARGINAL y FREQUENCY	
		-18	-15	-12	-9	-6	-3	0	3	6	9	12	15	18				
18																		2
15																		5
12																		12
9																		21
6																		42
3																		67
0																		114
-3																		54
-6																		38
-9																		23
-12																		14
-15																		6
-18																		2
MARGINAL x FREQUENCY	1	6	13	15	28	41	77	62	54	48	30	17	8				Σ 400	
MARGINAL x RELATIVE FREQUENCY	.002	.015	.033	.038	.070	.103	.192	.155	.135	.120	.075	.042	.020				Σ 1.000	
Σ OF $P(y x)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00					

shots on which the column is erected. For example, the volume of the column over the square centered about $(6, 6)$ is equal to $8/400 = 0.02$, where 8 is the number of shots in the square and 400 is the total number of shots. Similarly, the relative frequency and hence the volume of the column over $(-3, 6)$ is $4/400 = 0.01$. The relative frequency represents the probability of occurrence and therefore $P(6, 6) = 0.02$ and $P(-3, 6) = 0.01$.

In the foregoing discussion consideration was given to the probability of shots falling in a particular square on the target board, that is, the probability involving both random variables \tilde{x} and \tilde{y} . Suppose, however, that we are interested in the aiming ability of the individual in the x direction only, without regard to the y direction. In other words, we wish to find out how the shots are distributed about the center of the target to the right and to the left of it. This is then a case of one random variable, \tilde{x} , and the values in the third row from the bottom in Table 2.5 show the frequency of shots relative to the x intervals of 3 cm each. These values are simply obtained by summing up the values in each column. Similarly, the distributions in the y direction only are obtained by adding up the numbers along each row and are given in the last column of Table 2.5. These two sets of numbers are called *marginal frequencies* as each represents the frequencies along one direction or margin. If the marginal frequencies are divided by the total number of shots, we would get the marginal relative frequencies. This is shown in the second to last row of Table 2.5 for the x direction. From these relative frequencies we would then also talk about *marginal probability* as being the probability that *only one random variable* takes on a certain value, or falls within a certain interval, without regard to other random variables. Thus the probability that the x coordinate of the shots falls within the interval $+7.5$ to $+10.5$ cm is 0.120, whereas it is 0.192 for the interval -1.5 to $+1.5$. The same can be computed for the marginal probabilities of the y coordinates. These ideas demonstrate the theoretical concepts given in Section 2.3.1.

Another univariate probability distribution is also possible for one variable as related to a particular interval along the other variable. As an example, the distribution of y given a value $x = -6$ is represented by the numbers (relative frequencies) 0.04, 0.11, 0.25, 0.25, 0.20, 0.11, and 0.04 in the column under $x = -6$. Similarly, each horizontal row in Table 2.5 gives the distribution of x for a given interval of y . Such distributions are called *conditional frequencies* or *conditional distributions*. If we are interested in the conditional distribution of x given intervals of y , we should study the relative frequencies along different rows. It can be seen that the relative frequencies differ from one row to another. This means that as one variable changes (in this case y), the conditional distribution of the other (x) also changes. Such an influence between the two random variables is an indication of *correlation*. As was said previously, statistical correlation or dependence should not be confused with functional (algebraic) dependence. If

the latter existed between two variables, the value of one would completely determine the other. The example under consideration shows clearly that x and y are not functionally related; given x , the variate y may take several values, and vice versa.

Correlation between random variables may be strong or weak. Strong correlation would be indicated by the fact that relative frequencies of the conditional distribution of one random variable differ appreciably as the value of the other variable changes. On the other hand weak correlation would be reflected by having conditional relative frequencies that vary little when the other variable takes on different values. In the limit when correlation is zero, relative frequencies of the conditional distribution would be the same regardless of the value of the second variable. Such a case would be reflected in Table 2.5 by having equal relative frequencies from row to row and from column to column. This discussion of correlation and conditional distribution using numerical values from Table 2.5 amplifies on the theoretical treatment of Sections 2.3.2 and 2.3.3.

From frequency distributions, conclusions may be drawn as to the associated probability distribution. In applications to engineering, the assessment of probability distributions through frequency distributions is not often done simply because large sample size is not practical. In most cases the type of probability distribution associated with the random variable at hand is assumed to be known a priori. It is then some parameters of the distribution that are estimated from the sample. For such parameter estimation the sample size need not be as large as for estimation of the distribution itself.

The most common sample statistics used to estimate distribution parameters are discussed next.

SAMPLE STATISTICS FOR POSITION MEASURES

Sample Mean: From a sample of a size n (x_1, x_2, \dots, x_n) of the stochastic variable \tilde{x} , the sample mean is calculated as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.89)$$

which is the same as the arithmetic mean. Since \tilde{x} is a random variable and its expected value is the population mean, that is, $E(\tilde{x}) = \mu$, the expectation of the mean is

$$\begin{aligned} E(\bar{x}) &= E\left[\frac{1}{n} \sum_i x_i\right] = \frac{1}{n} E(x_1 + x_2 + \dots + x_n) \\ &= \frac{1}{n} [E(x_1) + E(x_2) + \dots + E(x_n)] = \frac{1}{n} \cdot n\mu. \end{aligned}$$

$$\text{or } E(\bar{x}) = \mu. \quad (2.90)$$

Thus the arithmetic mean of a set of independent observations is an estimate of the mean of the population describing the random variable for which observations are made.

The Median: A sample median is obtained by arranging the values in the sample in their order of magnitude. The median is the value in the middle if the number of observations n is odd, and it is the mean of the middle two values if n is even. There will therefore be an equal number of observations of larger magnitude as well as of smaller magnitude than the median.

It should be noted that for a continuous distribution the mean and median will be equal if the distribution is symmetric. This will also be nearly so for a sample of large size and approximately symmetrical values.

The Mode: The sample mode is the value that occurs most often in the sample. Therefore if a histogram is constructed, it is the value at which the highest rectangle is located.

Midrange: If the observation of smallest magnitude is subtracted from that of the largest magnitude, we obtain a value that is called the *range*. The value of the observation that is midway along the range is called the *midrange* and may be used, though infrequently, as a measure of position for a given sample. It is simply the arithmetic mean of the largest and smallest observations.

SAMPLE STATISTICS FOR DISPERSION MEASURES

The Range: The simplest dispersion measure is the range as defined in the preceding section. It is not, however, as indicative a measure as are others.

THE MEAN (OR AVERAGE) DEVIATION The mean (or average) deviation is another measure of dispersion that is suited for coarse estimation and has been conventionally called the *average error*. It is the arithmetic mean of the absolute values of the deviations from any measure of position (mostly the mean). Thus the mean deviation from the mean for a sample of n observations would be given by

$$\text{mean deviation} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (2.91)$$

where \bar{x} is the arithmetic mean [equation (2.89)].

VARIANCE AND STANDARD DEVIATION The mean deviation, though useful in certain cases, does not reflect the dispersion or scatter of the measured values as does the standard deviation that was defined previously as the

square root of the variance. The variance of a sample is defined as (see also Section 2.5.2)

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.92)$$

where \bar{x} is the sample mean and n is the sample size. The reason for using $(n-1)$ instead of n in equation (2.92) can be justified if it is shown that the expectation of s_x^2 is equal to the population variance. First

$$\begin{aligned} \sum_{i=1}^n (\tilde{x}_i - \bar{x})^2 &= \sum [(\tilde{x}_i - \mu) - (\bar{x} - \mu)]^2 \\ &= \sum (\tilde{x}_i - \mu)^2 + n(\bar{x} - \mu)^2 - 2(\bar{x} - \mu) \sum (\tilde{x}_i - \mu) \\ &= \sum (\tilde{x}_i - \mu)^2 + n(\bar{x} - \mu)^2 - 2(\bar{x} - \mu)(n\bar{x} - n\mu) \end{aligned}$$

or

$$\sum (\tilde{x}_i - \bar{x})^2 = \sum (\tilde{x}_i - \mu)^2 - n(\bar{x} - \mu)^2 \quad (2.93)$$

The expectation of the last term in equation (2.93) is

$$\begin{aligned} E[n(\bar{x} - \mu)^2] &= nE\left[\frac{\sum \tilde{x}_i}{n} - \mu\right]^2 = nE\left[\frac{(\sum \tilde{x}_i - n\mu)^2}{n^2}\right] \\ &= \frac{1}{n}E[\sum (\tilde{x}_i - \mu)^2] = \frac{1}{n}\sum E(\tilde{x}_i - \mu)^2 \end{aligned}$$

or

$$E[n(\bar{x} - \mu)^2] = \frac{n\sigma^2}{n} = \sigma^2 \quad (2.94)$$

Now, taking the expectation of equation (2.92),

$$\begin{aligned} (n-1)E(s_x^2) &= \sum E(\tilde{x}_i - \mu)^2 - \sigma^2 = n\sigma^2 - \sigma^2 \\ E(s_x^2) &= \sigma^2 \end{aligned} \quad (2.95)$$

Equation (2.95) is a very important result as regards to estimating parameters through calculation of statistics. It expresses the fact that the statistic s_x^2 , as computed by equation (2.92), has a probability distribution whose expected value or mean is σ^2 . If we used n instead of $(n-1)$ in equation (2.92), we would find that the expected value is $[(n-1)/n]\sigma^2$ instead of σ^2 . In such a case the sample variance computed using n would be a *biased* estimate because its expectation is different from the population parameter, as will be discussed in Section 2.7.2.

SAMPLE COVARIANCE Given a set of n pairs of values $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ of the random vector (\tilde{x}, \tilde{y}) , we can compute in addition to the marginal variances s_x^2 and s_y^2 , the sample covariance as

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (2.96)$$

where \bar{x} and \bar{y} are the sample means of \tilde{x} and \tilde{y} .

2.7.2. Estimation Drawing inferences regarding parameters of probability distributions on the basis of sample statistics is referred to as *estimation*. The sample statistic used for estimating the corresponding parameter is called an *estimator*, whereas the computed value is referred to as an *estimate* or, more precisely, a *point estimate*. These terms refer equally to single as well as multiple dimension cases.

As examples of estimators, the sample mean \bar{x} , (2.89), the sample variance s_x^2 , (2.92), and the sample covariance s_{xy} , (2.96), are estimators for the distribution parameters μ , σ_x^2 , and σ_{xy} , respectively. From the statistical point of view all values derived from observations are estimates for the distribution parameters in question. Because estimates can be obtained in different ways, certain criteria are needed to judge the quality of estimation. There are four different criteria.

1. *Consistence*: An estimator is called “consistent” if for $n \rightarrow \infty$ the probability for the estimator \hat{p} to approach the parameter p converges toward 1. Thus for any small $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|\hat{p} - p| < \varepsilon) = 1 \quad (2.97)$$

The estimator is said to converge in probability to the parameter. As an example, the sample mean \bar{x} given by equation (2.89) is a consistent estimator of the distribution mean μ_x of the random variable \tilde{x} .

2. *Unbiased Estimation*: Although consistence is a property related to the limit case of $n \rightarrow \infty$ it does not say anything about the property for small sample size. Among several possibilities, an estimator should not be biased. That means the expectation of the sample statistic \hat{p} should be identical to the parameter p itself for any sample size n ,

$$E(\hat{p}) = p \quad (2.98)$$

If this property holds only for $n \rightarrow \infty$, the estimator is said to be asymptotically unbiased.

As an example, the sample variance s_x^2 of equation (2.92) is an unbiased estimator of the population variance σ_x^2 . As was shown in the preceding section if instead of using $(n-1)$ we used n in equation (2.92), the computed

sample variance would be biased (too small). It is, however, asymptotically unbiased since

$$\lim_{n \rightarrow \infty} \frac{1}{n} \rightarrow \lim_{n \rightarrow \infty} \frac{1}{n-1}$$

Figure 2.9 shows the distribution of three different estimators of p . The first two, \hat{p}_1 and \hat{p}_2 , are unbiased but the third is biased. The bias is defined as the difference between the value of the parameter and the expectation of the estimator,

$$\text{bias} = E(\hat{p}) - p \quad (2.99)$$

The bias is equivalent to the term “systematic effect” or “systematic error” as conventionally used. Just as increasing the number of observations will not alleviate systematic effects, neither will it eliminate the bias.

3. *Minimum Variance*: Although unbiasedness involves the relationship between the expectation of the estimator and the parameter, the variances of the estimates from an estimator offer another criterion for estimator selection. In Figure 2.9, we can see that the variance of \hat{p}_1 is smaller than that for \hat{p}_2 , but the variance of \hat{p}_3 is smallest of the three. Hence, if we were to choose minimum variance as a criterion, \hat{p}_3 would be selected. However, realizing that \hat{p}_3 is biased, we must then weigh the benefits of unbiasedness against that of minimum variance.

ACCURACY AND PRECISION Figure 2.9 can be used to illustrate the related concepts of accuracy and precision. Precision may be defined as the degree of conformity among a set of observations of the same random variable. The spread (or dispersion) of the probability distribution is an indication of precision. Therefore in Figure 2.9 \hat{p}_2 is least precise and \hat{p}_3 is most precise.

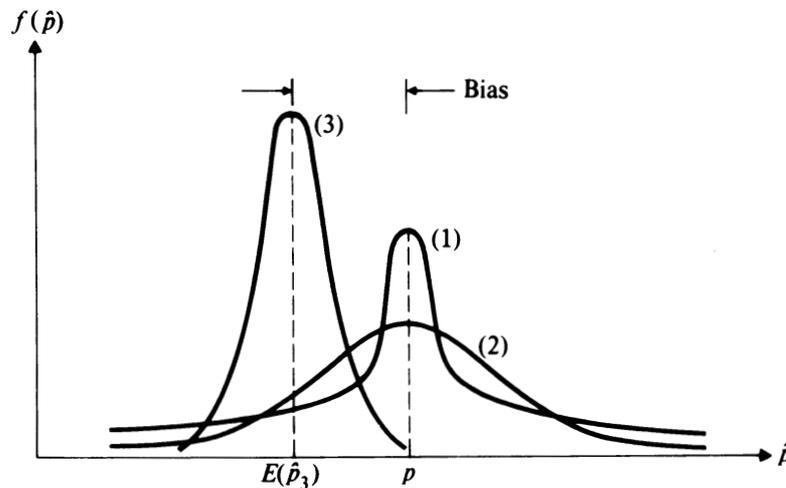


Figure 2.9. Biased Estimator

Accuracy, on the other hand, may be defined as the extent to which an estimate approaches its parameter. (In conventional terms, it was considered as the degree of closeness to the “true” value.)

In Figure 2.9, both \hat{p}_1 and \hat{p}_2 are equally accurate but neither is as precise as \hat{p}_3 . By contrast, \hat{p}_3 is least accurate although it is the most precise.

The difference between precision and accuracy lies in the possible presence of bias or “systematic error.” Although precision includes only random effects, accuracy encompasses both random and systematic effects. A measure of accuracy (proposed by Gauss) is the “mean square error” (MSE) given by

$$\text{MSE} = m^2 = E[(\hat{p} - E(\hat{p}))^2] \quad (2.100)$$

which can be shown to reduce to

$$\text{MSE} = m^2 = \sigma_{\hat{p}}^2 + (\text{bias})^2 \quad (2.101)$$

Although mean square error may be used for comparing relative accuracies of estimates that differ substantially in bias and precision, as do those in Figure 2.9, it is really not the most appropriate measure. Instead, we should assess accuracy on the basis of both the measure of precision and the bias.

4. *Efficiency and Sufficiency*: For unbiased estimators, that with minimum variance is called “efficient estimator.” Thus if $\text{var } \hat{p}_1 < \text{var } \hat{p}_2$ and \hat{p}_1 and \hat{p}_2 are estimators of the same parameter p , then \hat{p}_1 is efficient. The efficiency of \hat{p}_2 is equal to the ratio $\text{var } \hat{p}_1 / \text{var } \hat{p}_2$.

To demonstrate the above concepts, let us consider the random variable \bar{x} , whose expectation is

$$E(\bar{x}) = \mu = \text{the mean of the distribution}$$

Obviously any observation x in a set of, say, n independent observations would be an *unbiased* estimator of μ . The arithmetic mean \bar{x} is also an estimator of μ which is also *unbiased* as proved in equation (2.90). Thus on the basis of unbiasedness, both estimators are equally acceptable. However, the variance of x is obviously that of the population or σ^2 . The variance of the mean $\sigma_{\bar{x}}^2$, on the other hand, may be evaluated from the basic definition as

$$\sigma_{\bar{x}}^2 = E[(\bar{x} - E(\bar{x}))^2] = E(\bar{x} - \mu)^2$$

which can be readily deduced from equation (2.94) as

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad (2.102)$$

Since $\sigma_{\bar{x}}^2 < \sigma^2$, the mean \bar{x} satisfies the criterion of minimum variance better than an individual observation x . As a matter of fact \bar{x} is also an efficient

estimator since the efficiency of x as an estimator relative to \bar{x} is $\sigma_x^2/\sigma^2 = 1/n$, which can be a very small number. It can also be shown that \bar{x} is a consistent estimator.

An estimator is called sufficient if it uses all the information about the parameter that is contained in the sample. The formulation of this property is too involved and will not therefore be considered here.

METHODS OF ESTIMATION There are many ways of estimating a parameter. Estimators that have the four properties mentioned above are called “best” estimators. Some of these properties may be valid only asymptotically, however. In addition, these estimators may be called best asymptotically normally distributed with $n \rightarrow \infty$.

In practical applications of statistical estimation from samples it is not necessary to evaluate all criteria to determine which one to apply in a particular estimation. Instead, methods for estimation are applied for which the criteria, or some of them, hold in general. A few methods are briefly mentioned here.

The *moment method* takes the k th sample moment,

$$m_k = \frac{1}{n} \sum x_i^k \quad (2.103)$$

as an estimate for the k th moment of the probability distribution.

The sample mean is one example. However, when equation (2.103) is applied to the second central moment, it yields only asymptotically unbiased estimators for the variance and covariance.

The *maximum likelihood method* is an estimation method that is widely used in statistics. It is used for estimating parameters such that the sample values take the highest probability. If the random variables x_i are independent and have the same distribution (density function $f(x_i)$), then the sample vector (x_1, x_2, \dots, x_n) has the joint density function

$$L(x_1, \dots, x_n; p_1, \dots, p_m) = \prod_{i=1}^n f(x_i; p_1, \dots, p_m) \quad (2.104)$$

according to equation (2.22). Here, p_1, \dots, p_m are the parameters to be estimated. Equation (2.104) is called the likelihood function. The density $f(x_i)$ are functions of the unknown parameters p_1, \dots, p_m , which in turn are related to the sample values $p_i = g_i(x_i, \dots, x_n)$. It seems plausible to determine estimators \hat{p}_i of p_i such that the joint probability density L [equation (2.104)] becomes a maximum. This leads to the system

$$\frac{\partial L}{\partial p} = 0 \quad \text{or} \quad \frac{\partial \ln L}{\partial p} = 0 \quad (2.105)$$

The solution from equation (2.105) yields estimators of the parameters p_i and these are called maximum likelihood estimators. Note that this method

of estimation requires the knowledge of the distribution function of the random variable in question.

As an example, suppose that we would like to estimate the mean of a normal distribution using the data from the sample x_1, x_2, \dots, x_n . Since the density function for a normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}$$

the likelihood function is given by

$$L = \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\frac{(x_1-\mu)^2}{\sigma^2}} \right] \cdot \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\frac{(x_2-\mu)^2}{\sigma^2}} \right] \\ \dots \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\frac{(x_n-\mu)^2}{\sigma^2}} \right]$$

or

$$L = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-1/2\sum_i \frac{(x_i-\mu)^2}{\sigma^2}}$$

To get the first derivative of L it would be easier if we work with the natural logarithm. Thus

$$\ln L = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2$$

which when differentiated with respect to μ , gives

$$\frac{d(\ln L)}{d\mu} = \frac{-2}{2\sigma^2} \sum (x_i - \mu)(-1) = 0$$

for maximum

$$\sum_i (x_i - \mu) = \sum_i x_i - n\mu = 0$$

or

$$\mu = \frac{1}{n} \sum_i x_i$$

This shows that the maximum likelihood estimator of the mean of the normal distribution is the sample mean \bar{x} as previously defined.

It is interesting to note that it was possible to estimate μ without encountering the second normal distribution parameter σ^2 (variance). Perhaps it is worthwhile to seek an estimator for σ^2 using the maximum likelihood function. Differentiating $\ln L$ with respect to σ^2 yields

$$\frac{\partial(\ln L)}{\partial(\sigma^2)} = -\frac{n\pi}{2\pi\sigma^2} + \frac{1}{2\sigma^4} \sum_i (x_i - \mu)^2 = 0$$

for maximum

$$\frac{1}{\sigma^2} \sum (x_i - \mu)^2 = n$$

or

$$\sigma^2 = \frac{1}{n} \sum_i (x_i - \mu)^2$$

which is not the same as s_x^2 . In fact σ^2 is a biased estimate, which happens to be true about many maximum likelihood estimates. It is, however, consistent and asymptotically unbiased.

The *least squares* method is a very widely used procedure for estimating parameters of distributions from samples. Since it is the major topic of this book, its different techniques will be extensively treated in Parts II and III.

Least squares has been used in different branches of science and engineering for over a century and a half. Legendre published his own version of least squares in 1805 only to discover that Gauss had been using the same method since 1795. In addition to these two famed men, Laplace also holds a place of prominence with regard to the history of least squares and the theory of errors. Although both Gauss and Legendre were mainly interested in such sciences as astronomy and geodesy and the treatment of observations therein, Laplace had a purely theoretical interest. The latter's contributions were therefore important events in the history of calculus of probability and the theory of random errors.

Unlike the maximum likelihood method of estimation, the method of least squares does not require the knowledge of the distribution from which the observations are drawn for the purpose of parameter estimation. However, the construction of confidence intervals or the testing of hypotheses regarding the estimated parameters (which are given in the following sections) will require the knowledge of the distribution.

If the random variables to which the observations refer are normally distributed, the least squares method will give identical results to those from the maximum likelihood method. For instance, if \mathbf{v} is the vector of observational residuals (which is equal to a posteriori observational estimate $\hat{\mathbf{l}}$, minus the a priori given estimate \mathbf{l}), for which $E(\mathbf{v}) = \mathbf{0}$, and which is assumed to be normally distributed, and Σ is the covariance matrix of the observations, then

$$\begin{aligned} f(\mathbf{v}) &= C \exp \left[-\frac{1}{2}(\mathbf{v} - E(\mathbf{v}))' \Sigma^{-1} (\mathbf{v} - E(\mathbf{v})) \right] \\ &= C \exp \left[-\frac{1}{2} \mathbf{v}' \Sigma^{-1} \mathbf{v} \right] \end{aligned}$$

The least squares criterion is $\mathbf{v}' \Sigma^{-1} \mathbf{v} \rightarrow \text{minimum}$, as will be given in detail in Part II. It is clear, then, that minimizing $\mathbf{v}' \Sigma^{-1} \mathbf{v}$ would maximize $f(\mathbf{v})$, which is equivalent to yielding a maximum likelihood estimation.

With linear functions the estimated parameters (in particular the estimated expectations) are consistent, unbiased, efficient, sufficient, and have the minimum variance property, especially when there are no systematic effects in the observations.

2.7.3. Confidence Intervals The estimation of means, variances, and covariances of random variables from sample data is referred to as *point estimation*, because it results in one value for each parameter in question. By contrast to point estimation, establishing confidence intervals from sampling is referred to as *interval estimation*.

After having performed a point estimation—for instance, having estimated the coordinates of one intersected point—the question remains as to how much the deviation of the estimate is likely to be from the still unknown parameter. In other words, we would like to have an indication of how good the estimation is and how much it can be relied on. Here an absolute answer is not possible because sampling never leads to the true, theoretical distribution or its parameters. It is only possible to estimate probabilities with which the true value of the parameter in question is likely to be within a certain interval around the estimate. Such probabilities can be determined if the cumulative distribution function $F(x)$ of the random variable is given.

The probability that a random variable \tilde{x} takes values within the boundaries x_1 and x_2 is given, according to equation (2.5), by

$$P(x_1 < \tilde{x} < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx$$

By analogy to this, the probability statement for a *confidence interval* of a parameter p , the estimate of which is \hat{p} , is

$$P(p_1 < \tilde{p} < p_2) = 1 - \alpha \quad (2.106)$$

Here $(1 - \alpha)$ is called the *confidence level* or degree of confidence which is conventionally taken to be 90%, 95%, or 99%. The values p_1 and p_2 are the lower and upper confidence limits for the parameter p . Equation (2.106) defines the confidence interval for the parameter p as the interval around the estimate \hat{p} , such that the probability that this interval includes the (unknown) value of the parameter is $(1 - \alpha)$. The probability that the parameter does not fall in a given interval is the value α . The width of a confidence interval decreases as the degrees of freedom increase and as the level of probability associated with it decreases.

In constructing confidence intervals, it is essential to use suitable random variables whose values are determined by the sample data as well as by the parameters, but whose distributions do not involve the parameters in question. Examples are given in the following paragraphs.

CONFIDENCE INTERVALS FOR MEANS We shall consider a normally distributed random variable \tilde{x} , with unknown expectation μ and known variance σ^2 . We will let the sample mean be \bar{x} , computed from a sample of size n . Then the estimator function $(\bar{x} - \mu)/(\sigma/\sqrt{n})$ will be normally distributed with mean 0 and variance 1. We choose from the normal distribution function a value $z_{\alpha/2}$ such that the probability of the standard normal random variable to be outside the interval $(\pm z_{\alpha/2})$ is α (see Figure 2.10). The probability statement for the confidence interval, which is symmetrical here, is then

$$P\left\{-z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right\} = 1 - \alpha \quad (2.107a)$$

or

$$P\left\{\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha \quad (2.107b)$$

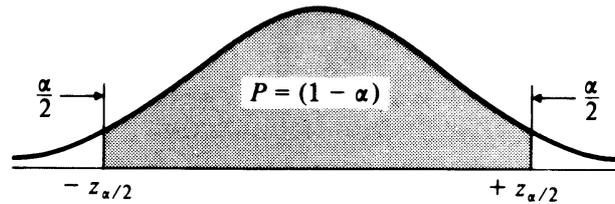


Figure 2.10

Since for $\alpha = 0.05$, $z_{\alpha/2} = 1.96$, we may write

$$P[\bar{x} - 1.96\sigma/\sqrt{n} < \mu < \bar{x} + 1.96\sigma/\sqrt{n}] = 0.95$$

This was an example of a two-sided confidence interval. On the other hand, for a one-sided confidence interval we write

$$P\left\{\mu < \bar{x} + z_{\alpha}\left(\frac{\sigma}{\sqrt{n}}\right)\right\} = 1 - \alpha \quad (2.108)$$

Example 2.3. Suppose a distance is measured $n = 8$ times with the mean $\bar{x} = 10.1$ cm. We assume that the variance of the normal population is known to be $\sigma^2 = 0.10$ cm². Then the 95% confidence interval on μ (which is unknown) according to equation (2.107b) is

$$P\left\{10.1 - 1.96\sqrt{\frac{0.10}{8}} < \mu < 10.1 + 1.96\sqrt{\frac{0.10}{8}}\right\} = 0.95$$

or

$$P[9.88 \text{ cm} < \mu < 10.32 \text{ cm}] = 0.95$$

for a two-sided confidence interval. For a one-sided interval, equation (2.108) may be applied,

$$P\left\{\mu < 10.1 + 1.645 \sqrt{\frac{0.10}{8}}\right\} = 0.95$$

or

$$P[\mu < 10.28 \text{ cm}] = 0.95$$

Let us now consider the case in which the standard deviation of the distribution σ is not known and has to be replaced by the standard deviation s , as estimated from the sample [according to equation (2.92)].

Then the estimator $(\bar{x} - \mu)/(s/\sqrt{n})$ has a t distribution (see Section 2.5.2) with $(n - 1)$ degrees of freedom. The probability statement for the confidence interval of the random variable of the estimator will be

$$P\left\{-t_{\alpha/2, n-1} < \frac{\bar{x} - \mu}{s/\sqrt{n}} < t_{\alpha/2, n-1}\right\} = 1 - \alpha \quad (2.109)$$

or

$$P\left\{\bar{x} - t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}\right\} = 1 - \alpha \quad (2.110)$$

For $\alpha = 0.05$ ($1 - \alpha = 0.95$) we obtain from Appendix Table D.2 of t distribution for $(n - 1) = 7$, for instance, $t_{\alpha/2} = 2.365$, and $z_{\alpha/2} = 1.96$. The confidence interval is wider in this case compared to the case of known σ , as illustrated in the following example. As the degrees of freedom increase (that is, the larger the sample size), the difference between the values of t and z decreases, all other factors remaining the same. In practice, for a sample size of 30 or larger, the t distribution is often approximated by the standard normal distribution.

Example 2.4. Considering the preceding example of the measured distance, $n = 8$ times, $\bar{x} = 10.1$ cm, let $s^2 = 0.10$ cm² be the *sample* variance. The 95% confidence interval on μ is, according to equation (2.110),

$$P\left\{10.1 - 2.365 \sqrt{\frac{0.10}{8}} < \mu < 10.1 + 2.365 \sqrt{\frac{0.10}{8}}\right\} = 0.95$$

or

$$P[9.84 \text{ cm} < \mu < 10.36 \text{ cm}] = 0.95$$

CONFIDENCE INTERVALS FOR VARIANCES AND COVARIANCES Suppose we are given a random sample of variance s^2 from a normal population with variance σ^2 . The random variable ms^2/σ^2 , where m is the number of degrees

of freedom used in computing s^2 , has a χ^2 distribution with m degrees of freedom. For the confidence level $(1 - \alpha)$ this random variable will take a value between $\chi_{1-\alpha/2, m}^2$ and $\chi_{\alpha/2, m}^2$ with the probability $(1 - \alpha)$. Thus

$$P\left\{\chi_{1-\alpha/2, m}^2 < \frac{ms^2}{\sigma^2} < \chi_{\alpha/2, m}^2\right\} = 1 - \alpha \quad (2.111)$$

or

$$P\left\{\frac{ms^2}{\chi_{\alpha/2, m}^2} < \sigma^2 < \frac{ms^2}{\chi_{1-\alpha/2, m}^2}\right\} = 1 - \alpha \quad (2.112)$$

For instance, with $m = 10$ and $\alpha = 0.05$ we get from Appendix Table D.3 $\chi_{0.025, 10}^2 = 20.48$ and $\chi_{0.975, 10}^2 = 3.25$. This interval is two sided but is asymmetric. For the one-sided interval we have $\chi_{0.05, 10}^2 = 18.31$.

Example 2.5. Suppose, as in preceding example, $n = 8$, $\bar{x} = 10.1$ cm, and $s_x^2 = 0.10$ cm². The 99% confidence interval on the population variance σ^2 may be established using equation (2.112). Here $m = n - 1 = 7$ degrees of freedom. Thus

$$P\left\{\frac{(7)(0.10)}{20.28} < \sigma^2 < \frac{(7)(0.10)}{0.99}\right\} = 0.99$$

or

$$P[0.035 \text{ cm}^2 < \sigma^2 < 0.707 \text{ cm}^2] = 0.99$$

which states that the probability is 0.99 that the distribution variance lies between 0.035 cm² and 0.707 cm².

For a confidence interval on the standard deviation σ , we have merely to take the square root of each side in the interval on the variance. Thus in this example

$$P[0.186 \text{ cm} < \sigma < 0.841 \text{ cm}] = 0.99.$$

CONFIDENCE INTERVALS FOR RATIOS OF VARIANCES If we are given two independent random samples of sizes n_1 and n_2 from normal populations with variances σ_1^2 and σ_2^2 , respectively, then each of the random variables $m_1 s_1^2/\sigma_1^2$ and $m_2 s_2^2/\sigma_2^2$ with $m_1 = n_1 - 1$ and $m_2 = n_2 - 1$ has a χ^2 distribution with m_1 and m_2 degrees of freedom, respectively. The ratio

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \quad (2.113)$$

has an F distribution with m_1, m_2 degrees of freedom. For the confidence level $(1 - \alpha)$ the random variable F in equation (2.113) will take a value between $F_{1-\alpha/2, m_1, m_2}$ and $F_{\alpha/2, m_1, m_2}$ with the probability $(1 - \alpha)$. Thus

$$P\left\{F_{1-\alpha/2, m_1, m_2} < \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} < F_{\alpha/2, m_1, m_2}\right\} = 1 - \alpha \quad (2.114)$$

OR

$$P\left\{\frac{s_2^2}{s_1^2} F_{1-\alpha/2, m_1, m_2} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{s_2^2}{s_1^2} F_{\alpha/2, m_1, m_2}\right\} = 1 - \alpha \quad (2.114a)$$

Because of the way the F tables are typically presented (see Appendix D) the following equality is often applied:

$$F_{1-\alpha/2, m_1, m_2} = \frac{1}{F_{\alpha/2, m_2, m_1}} \quad (2.115)$$

For instance, with $m_1 = 10$, $m_2 = 60$, and $\alpha = 0.05$,

$$F_{0.975, 10, 60} = \frac{1}{F_{0.025, 60, 10}} = \frac{1}{3.20} = 0.312$$

$$F_{0.025, 10, 60} = 2.27$$

which gives the following confidence interval for σ_2^2/σ_1^2 :

$$0.312 \frac{s_2^2}{s_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < 2.27 \frac{s_2^2}{s_1^2}$$

2.8. STATISTICAL TESTS

Statistical tests are increasingly applied in engineering and science and in combination with the least squares method. They are often used to compare results with previous ones or with given standards. In testing, one seeks a judgment as to whether some estimator function is consistent with the assumption (hypothesis) that the sample was drawn from a population with specified parameter values, such as normal distribution with a given standard deviation. Questions of this kind are the subject of hypothesis testing.

A hypothesis is a statement (explicit or implicit) about the probability distribution of a random variable. If the hypothesis covers the complete set of parameters for a distribution, we speak of a simple hypothesis. Conversely, a hypothesis is said to be composite if it covers only a number of distribution parameters, leaving others unspecified.

The general procedure of statistical testing always refers to a null hypothesis H_0 —that is, the set of distribution parameters with respect to which the sample estimates at hand are to be compared. The intention is to formulate a statement as to whether the sample parameters are in sufficient agreement with H_0 . In other words the result of the test is a statement whether, according to the available evidence, the null hypothesis can be considered acceptable or not.

A more specific task of a test is to make a decision between a null hypothesis H_0 and one or several alternative hypotheses. The alternative hypothesis can be simple, that is, specific or composite.

When accepting or rejecting a null hypothesis, there are four possible cases: (a) H_0 is true and is accepted; (b) H_0 is true but is rejected; (c) H_0 is false and is rejected; (d) H_0 is false and is accepted. Cases (a) and (c) are obviously correct decisions, and cases (b) and (d) are wrong decisions. Wrongly rejecting a true hypothesis is referred to as committing a type I error. The probability for type I error is designated by α , and is called the "significance level of the test." It is conventionally taken to be 5%, 2%, or 1%, as α has to be small in order that a test be useful. Case (d), which is the wrong acceptance of a false hypothesis, leads to committing a so-called type II error. The probability for type II error is designated β , and $(1 - \beta)$ is called the "power of a test."

Figure 2.11 demonstrates the relation between α and β . Let $D_1 = f(\hat{p} | H_0)$ be the conditional probability density function for the estimate of a parameter p when the null hypothesis H_0 is true; and let $D_2 = f(\hat{p} | H_1)$ be the

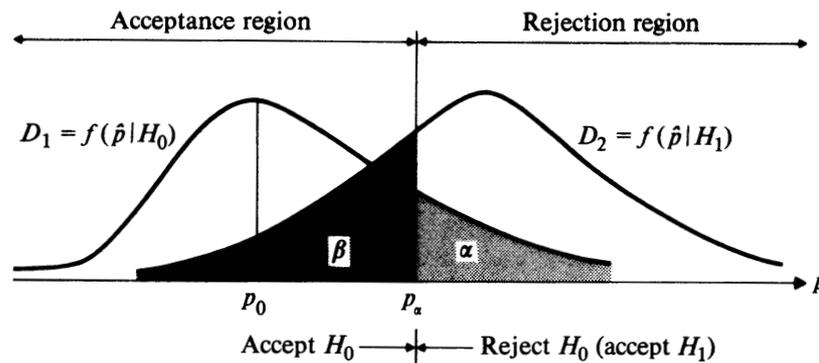


Figure 2.11

conditional probability density function for p in case the alternative hypothesis H_1 is true. We now consider the value p_α which marks the limit of sample values up to which H_0 is accepted and beyond which H_0 is rejected or H_1 is accepted. Thus H_0 is accepted if $\hat{p} < p_\alpha$, and H_0 is rejected or H_1 is accepted if $\hat{p} > p_\alpha$. The area α indicates the significance level of the test. It shows the probability that the sample parameter ($\hat{p} | H_0$) will exceed p_α in case H_0 is true. It shows, in other words, the probability of rejecting H_0 when it is true (type I error). Conversely, the area β shows the probability of accepting H_0 when H_1 is true (type II error). Figure 2.11 illustrates that for a certain alternative hypothesis H_1 it is not possible to make both α and β arbitrarily small. Decreasing the probability for a type I error increases the probability for a type II error and vice versa. Balancing type I versus type II errors depends on the purpose of the test.

Figure 2.11 represents the so-called one-tail or one-sided test. In this case H_0 is $p = p_0$ whereas the alternative hypothesis is $H_1: p > p_0$. Another one-sided test is depicted in Figure 2.12 for which $H_0: p = p_0$ versus

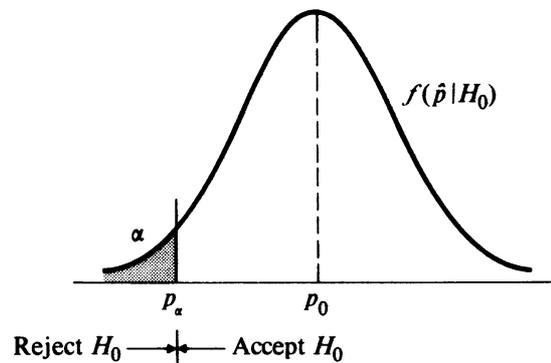


Figure 2.12

$H_1: p < p_0$. The two-tail or two-sided test is demonstrated in Figure 2.13. In this type of test the null hypothesis is $H_0: p = p_0$ whereas the alternative hypothesis is simply $H_1: p \neq p_0$. In the following sections hypothesis tests for different parameter estimates are given together with numerical examples.

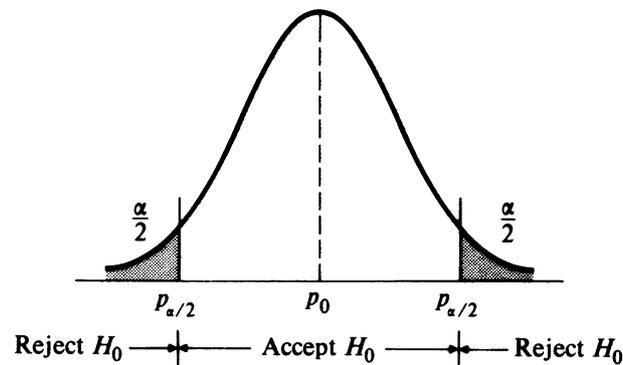


Figure 2.13

2.8.1. Test on Sample Means for Known σ In this case we are given a sample of size n with either the values x_i (from which the sample mean \bar{x} may be computed) or the sample mean \bar{x} , as well as the normal population standard deviation σ . The null hypothesis is $H_0: \mu = \mu_0$, which seeks to test that the population mean is equal to some a priori known value μ_0 . There are three possible alternative hypotheses: $\mu < \mu_0$, or $\mu > \mu_0$, or the two-tail case $\mu \neq \mu_0$. In order to perform any of the three possible hypothesis tests, the standardized normal random variable z is computed from

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad (2.116)$$

If z_α and $z_{\alpha/2}$ are values from the normal table (Appendix Table D.1) with the level of significance being α , then the three possibilities are as follows:

1. $H_0: \mu = \mu_0; H_1: \mu < \mu_0$; reject H_0 when $z < -z_\alpha$ because

$$P\left\{z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha\right\} = \alpha$$

2. $H_0: \mu = \mu_0; H_1: \mu > \mu_0$; reject H_0 when $z > z_\alpha$ because

$$P\left\{z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha\right\} = \alpha$$

3. $H_0: \mu = \mu_0; H_1: \mu \neq \mu_0$; reject H_0 when $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$ because

$$P\left\{-z_{\alpha/2} < \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < z_{\alpha/2}\right\} = 1 - \alpha$$

As examples, different values from Table D.1 are $z_{0.005} = 2.58$, $z_{0.01} = 2.33$, $z_{0.025} = 1.96$, $z_{0.05} = 1.64$.

2.8.2. Test on Sample Mean for Unknown σ When σ is unknown, the random variable

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (2.117)$$

is used in place of z of equation (2.116). All three tests can be carried out as in the preceding section except that z_α and $z_{\alpha/2}$ are replaced by $t_{\alpha, m}$ and $t_{\alpha/2, m}$, respectively, where $m = n - 1$ degrees of freedom.

We shall consider the problem of comparing two sample means where one is concerned with hypotheses regarding differences of means, and shall assume that one is dealing with independent random samples of sizes n_1 and n_2 from two normal populations having the means μ_1 and μ_2 and the known variances σ_1^2 and σ_2^2 . It is desired to test the null hypothesis $\mu_1 - \mu_2 = \delta$, where δ is a given constant, against one of the three alternatives $\mu_1 - \mu_2 \neq \delta$, $\mu_1 - \mu_2 > \delta$, or $\mu_1 - \mu_2 < \delta$. The test can be applied, based on the difference of the sample means $\bar{x}_1 - \bar{x}_2$, and using the standardized normal random variable

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \quad (2.118)$$

The rejection regions for the three tests are the same as those given in detail in Section (2.8.1.).

When the variances σ_1^2 and σ_2^2 are unknown but assumed equal, they are replaced by sample variances. In this case the random variable

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2.119)$$

is used for a test of the null hypothesis $\mu_1 - \mu_2 = \delta$ against the same alternative hypotheses given above. Here t is a random variable having a t distribution with $m = (n_1 + n_2 - 2)$ degrees of freedom. The same three possible hypothesis tests can be performed using $t_{\alpha, m}$ and $t_{\alpha/2, m}$ in place of z_α and $z_{\alpha/2}$, respectively.

2.8.3. Tests Concerning Variances Let us consider a random sample of size n from a normal population. We may test the null hypothesis $\sigma^2 = \sigma_0^2$ against the alternatives, $\sigma^2 \neq \sigma_0^2$, $\sigma^2 > \sigma_0^2$, and $\sigma^2 < \sigma_0^2$. In these tests, σ_0^2 is a specified value for the variance. Here the χ^2 test is used, based on the sample variance s^2 , with the variable

$$\chi_m^2 = \frac{ms^2}{\sigma_0^2} \quad (2.120)$$

computed from the sample data, where m is the degrees of freedom. If α is the level of significance, the three tests are as follows:

1. $H_0: \sigma^2 = \sigma_0^2$; $H_1: \sigma^2 \neq \sigma_0^2$; reject H_0 when $\chi_m^2 < \chi_{1-\alpha/2, m}^2$ or $\chi_m^2 > \chi_{\alpha/2, m}^2$ because

$$P\left\{\chi_{1-\alpha/2, m}^2 < \frac{ms^2}{\sigma_0^2} < \chi_{\alpha/2, m}^2\right\} = 1 - \alpha$$

where $m =$ degrees of freedom.

2. $H_0: \sigma^2 = \sigma_0^2$; $H_1: \sigma^2 < \sigma_0^2$; reject H_0 when $\chi_m^2 < \chi_{1-\alpha, m}^2$ because

$$P\left\{\frac{ms^2}{\sigma_0^2} < \chi_{1-\alpha, m}^2\right\} = \alpha$$

3. $H_0: \sigma^2 = \sigma_0^2$; $H_1: \sigma^2 > \sigma_0^2$; reject H_0 when $\chi_m^2 > \chi_{\alpha, m}^2$ because

$$P\left\{\frac{ms^2}{\sigma_0^2} > \chi_{\alpha, m}^2\right\} = \alpha$$

Example 2.6. Given a sample of size $n = 12$ with sample mean $\bar{x} = 10.0$ and sample variance $s^2 = 0.07$, test the hypothesis that the population variance $\sigma^2 = 0.10$ against the alternative that $\sigma^2 < 0.10$ for a level of significance of 0.05.

Solution:

$$\chi_m^2 = \frac{ms^2}{\sigma_0^2} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(11)(0.07)}{0.10} = 7.70$$

We would reject the null hypothesis of $\sigma^2 = 0.10$ if χ_m^2 computed above is less than $\chi_{1-\alpha, m}^2$. Since $\chi_{0.95, 11}^2 = 4.57$ is not larger than $\chi_m^2 = 7.70$, the null hypothesis is not rejected.

Another problem concerns the test on equality of the variances of two populations on the basis of the ratio between sample variances. Given two independent random samples of sizes n_1 and n_2 from two normal populations with the variances σ_1^2 and σ_2^2 , the null hypothesis is $H_0: \sigma_1^2 = \sigma_2^2$.

If the sample variances are s_1^2 and s_2^2 , with degrees of freedom $m_1 = n_1 - 1$ and $m_2 = n_2 - 1$, respectively, the random variable used in the test is

$$F_{m_1, m_2} = \frac{s_1^2}{s_2^2} \quad (2.121)$$

The three hypothesis tests possible are, in this case, as follows:

1. $H_0: \sigma_1^2 = \sigma_2^2$; $H_1: \sigma_1^2 < \sigma_2^2$; reject the null hypothesis if $F_{m_1, m_2} < F_{1-\alpha, m_1, m_2}$, because

$$P\left\{\frac{s_1^2}{s_2^2} < F_{1-\alpha, m_1, m_2}\right\} = \alpha$$

where α is the level of significance.

2. $H_0: \sigma_1^2 = \sigma_2^2$; $H_1: \sigma_1^2 > \sigma_2^2$; reject H_0 if $F_{m_1, m_2} > F_{\alpha, m_1, m_2}$ because

$$P\left\{\frac{s_1^2}{s_2^2} > F_{\alpha, m_1, m_2}\right\} = \alpha$$

3. $H_0: \sigma_1^2 = \sigma_2^2$; $H_1: \sigma_1^2 \neq \sigma_2^2$; reject H_0 if $F_{m_1, m_2} > F_{\alpha/2, m_1, m_2}$ when $s_1^2 > s_2^2$, which refers to the right boundary only. The rejection at the lower boundary is when $F_{m_1, m_2} < F_{1-\alpha/2, m_1, m_2}$.

Since $s_1^2 > s_2^2$, the value $F_{m_1, m_2} > 1$. Furthermore, $F_{1-\alpha/2, m_1, m_2} = 1/F_{\alpha/2, m_2, m_1}$ and $F_{\alpha/2, m_2, m_1} > 1$ (all entries in Table D.4 for the area under the F density function are larger than 1), then $F_{1-\alpha/2, m_1, m_2} < 1$. Henceforth F_{m_1, m_2} will always be larger than $F_{1-\alpha/2, m_1, m_2}$ and the test at the lower boundary is unnecessary.

Example 2.7. Given the following two sample data sets

$$\begin{array}{lll} n_1 = 8 & \bar{x}_1 = 10.1 & s_1^2 = 0.10 \\ n_2 = 12 & \bar{x}_2 = 10.0 & s_2^2 = 0.07 \end{array}$$

Test the hypothesis that $\sigma_1^2 = \sigma_2^2$ against the alternative that $\sigma_1^2 \neq \sigma_2^2$ using the level of significance of 0.05.

Solution:

$$m_1 = 8 - 1 = 7 \quad m_2 = 12 - 1 = 11$$

$$F_{7, 11} = \frac{s_1^2}{s_2^2} = \frac{0.10}{0.07} = 1.429 \quad (\text{note } s_1^2 > s_2^2)$$

We would reject the null hypothesis that $\sigma_1^2 = \sigma_2^2$ if F_{m_1, m_2} computed above is larger than $F_{\alpha/2, m_1, m_2}$. Since $\alpha/2 = 0.025$, and $F_{0.025, 7, 11} = 3.76$ is larger than $F_{7, 11} = 1.429$, then we cannot reject the null hypothesis.

There are of course a number of other tests that may be used in connection with adjustment problems such as (a) test for goodness of fit of frequency distribution on probability distribution, (b) test for correlation coefficients (see Section 11.5.3), and (c) test for regression coefficients. For such tests the reader should consult statistical literature. (See the Bibliography at the end of this book.)

3

Error Properties of Observations

3.1. INTRODUCTION

Modern information theory regards observations as signals the statistical properties of which are classified as having deterministic and stochastic components.

Until recently it was usual to speak of either the theory of observations or the theory of errors. Most of the foundation of present knowledge on the subject was laid down based on the idea of errors. The term “error” is actually somewhat misleading and seems to enjoy less and less usage. Instead, one often speaks in general of the statistical properties of observations.

Although the conventional concept of error should be avoided, the term still continues to be used to some extent. In order to avoid difficulties for the many who are still familiar with it, it will be occasionally used. In addition, an attempt will be made, whenever possible, to relate it to the modern concepts.

Classical ideas usually present errors as being of three types: random errors, systematic errors, and blunders.

3.2. RANDOM ERRORS

From the statistical standpoint, observations are considered to be samples from probability distributions of random variables. In the classical theory the variability due to sampling from probability distribution was what led to the idea of *observational errors*.

The term “observational error” or “random error,” as it is sometimes called, is very often restricted to a normal probability distribution. This is due to the fact that repeated observations usually display a normal frequency distribution. In principle, however, the term “random” is not necessarily restricted to normal distribution. Furthermore, it is not limited to one-dimensional distributions but holds equally for multidimensional distributions.

The error properties of observations are equivalent to statistical properties of sampling. Although in statistical sampling emphasis is placed on the independence of sampling, this feature is normally not considered a serious problem with observations, because the population from which the observations are drawn is hypothetically considered continuous and infinite. Independent sampling essentially means that observations should be set up such that the results of previous observations should not influence the results of the following observations. This is to be clearly distinguished from possible correlation of subsequent observations, perhaps because of common influence of a certain physical parameter (such as refraction). The two more essential parameters of a distribution are

μ_x = expectation or mean that is a location parameter

σ_x = standard deviation

or

σ_x^2 = variance, or the second moment about μ_x

(the parameter of dispersion)

In case of a normal distribution the two parameters μ_x and σ_x determine it completely.

When observations $x_i (i = 1, \dots, n)$ are made (or drawn), estimates for μ_x and σ_x are computed from them. Of direct importance is the sample mean \bar{x} [see equation (2.89)] which represents an estimate for μ_x . In case of one observation only, $\bar{x} = x$ is considered an estimate for μ_x . The standard deviation σ_x of the random variable \tilde{x} is often supposed to be known from previous experiments with similar equipment under similar conditions, and so on. If sufficient repetitions are available, then from the observations an estimate for the variance σ_x^2 can be obtained by computing the sample variance s_x^2 using equation (2.92).

The more general case refers to a multidimensional set of observations. Let us assume that there are several random variables $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_m$ with a joint probability distribution. The parameters of interest are therefore

$\mu_1, \mu_2, \dots, \mu_m$	the m expectations or means for the m random variables \tilde{x}_i
$\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2$	the variances of the marginal distributions for $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_m$
$\sigma_{12}, \sigma_{13}, \dots, \sigma_{1m}, \sigma_{23}, \dots, \sigma_{m-1, m}$	covariances between \tilde{x}_i and \tilde{x}_j in all combinations

In case of a multidimensional normal distribution, the above parameters determine it completely.

Considering the general case of multiple observations for m -dimensional distribution, we get the following observational set:

$$\begin{array}{l} x_{11}, x_{12}, \dots, x_{1n} \\ x_{21}, x_{22}, \dots, x_{2n} \\ \dots \dots \\ x_{m1}, x_{m2}, \dots, x_{mn} \end{array}$$

From such observations the sample means $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ may be computed as estimates for the population means $\mu_1, \mu_2, \dots, \mu_m$

$$\begin{array}{l} \bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_{1i} \\ \bar{x}_2 = \frac{1}{n} \sum_{i=1}^n x_{2i} \\ \dots \dots \\ \bar{x}_m = \frac{1}{n} \sum_{i=1}^n x_{mi} \end{array} \quad (3.1)$$

In practice we usually assume that the other parameters of the distribution ($\sigma_1^2, \dots, \sigma_m^2, \sigma_{12}, \dots, \sigma_{1m}, \dots, \sigma_{m-1, m}$) are known. In particular it is often assumed in practice that random variables \tilde{x}_i, \tilde{x}_j are uncorrelated (for all i and j). In case enough repeated observations are available, the variances

and covariances of the multidimensional distribution may also be estimated by the sample variances and sample covariances:

$$\begin{aligned}
 s_1^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 \\
 s_2^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2 \\
 &\dots \dots \\
 s_m^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_{mi} - \bar{x}_m)^2
 \end{aligned} \tag{3.2}$$

and

$$\begin{aligned}
 s_{12} &= \frac{1}{n-1} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \\
 &\vdots \\
 s_{kj} &= \frac{1}{n-1} \sum_{i=1}^n (x_{ki} - \bar{x}_k)(x_{ji} - \bar{x}_j) \\
 &\vdots \\
 s_{m-1, m} &= \frac{1}{n-1} \sum_{i=1}^n (x_{m-1, i} - \bar{x}_{m-1})(x_{mi} - \bar{x}_m)
 \end{aligned} \tag{3.3}$$

Thus the primary purpose of observations is to get estimates for the expectations of random variables. When we talk of *observational errors* or *random errors of observations*, we refer to the basic and inherent property that the estimates of a random variable \tilde{x} do not agree, in general, with its expectation. Thus an observational error may in this context be defined as

$$\varepsilon_{ji} = \mu_j - x_{ji} \tag{3.4}$$

It can also be extended to the sample mean,

$$\varepsilon_j = \mu_j - \bar{x}_j \tag{3.5}$$

From equation (3.4) we can readily show that the distribution of ε about zero, is identical to the distribution of \tilde{x} about its expectation μ_x . This is actually a result of the simple fact, directly obtainable from equation (3.4), that $E(\varepsilon) = 0$.

Error properties of observations are in general considered established once the complete probability distribution of the random variables to which they refer is specified. In practice, however, this is not always possible and often restricts consideration only to variances and covariances.

In this conception the mean of a sample of size n can be regarded as a *sample of size one* on a new random variable with mean μ and variance of σ^2/n . This case sometimes causes differing opinions. In the context of our discussion it may be considered as tantamount to having either n repeated observations, that is, a sample of size n on one random variable; or having n samples, each of size one on n different independent variables whose means μ and variance σ^2 are equal.

Closely related to the topic of random errors are the concepts of precision, accuracy, and weights.

3.3. PRECISION, ACCURACY, COFACTORS, AND WEIGHTS

Observations, as well as results of adjustments, are often assessed in practice by such terms as “accuracy” and “precision.” These terms are used to summarize statistical properties of random variables as introduced in Section 2.7.2.

Briefly, *accuracy* refers to the degree of closeness of an estimate to its parameter, whereas *precision* expresses the degree of closeness of observations to their mean. Therefore accuracy reflects the closeness of a location statistic to the value of the parameter for which it is an estimate, and precision is directly related to the dispersion of a distribution. In the one-dimensional case precision is taken to be represented by the value of the standard deviation σ .

The term “precision” is also applied to multidimensional distribution. The description of the precision properties of multidimensional random variables (random vector) is the complete set of second central moments, or the covariance matrix $\Sigma = \{\sigma_{ij}\}$.

A term related to the trace of this covariance matrix is often taken as an indication of the average precision:

$$\text{average precision} = \sqrt{\frac{\text{tr } \Sigma}{n}} = \frac{1}{\sqrt{n}} \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2} \quad (3.6)$$

The covariance matrix can be used to determine the error ellipsoids of an n -dimensional random variable as explained in Section 2.6.3.

In practical application of adjustment the variances and covariances are often replaced by what should be called relative variances and covariances. For these the terms “weight coefficient” or “cofactors” are in common use. The term “cofactor” is selected and the letters q for one element and Q for a matrix, are used as symbols for it.

A cofactor is related to a covariance by

$$q_{ij} = \frac{\sigma_{ij}}{\sigma_0^2} \quad \text{or} \quad \sigma_{ij} = q_{ij} \sigma_0^2 \quad (3.7)$$

The definition in equation (3.7) includes as a special case the relation between a cofactor and the variance:

$$q_i = \frac{\sigma_i^2}{\sigma_0^2} \quad \text{or} \quad \sigma_i^2 = q_i \sigma_0^2 \quad (3.8)$$

In equations (3.7) and (3.8) the term σ_0^2 is an arbitrary constant with arbitrary dimension, called the *reference variance*. (Its square root, σ_0 , will be referred to as the “reference standard deviation.”) It has also been called “variance factor” and the “reference variance associated with the weight unity” (see below).

Applying equations (3.7) and (3.8) to the covariance matrix, results in the cofactor matrix,

$$\mathbf{Q} = \{q_{ij}\} = \frac{1}{\sigma_0^2} \{\sigma_{ij}\} = \frac{1}{\sigma_0^2} \mathbf{\Sigma} \quad (3.9)$$

Because of the symmetry of the covariance matrix $\mathbf{\Sigma}$ (when referring to the elements of the same random vector), the cofactor matrix \mathbf{Q} is also symmetric for this case. It must also have only positive diagonal elements.

With equation (3.7) we can always switch from cofactors to covariances and vice versa. The inverse of the cofactor matrix \mathbf{Q} (when it is square and nonsingular) is called the *weight matrix* \mathbf{W} , thus

$$\mathbf{W} = \{w_{ij}\} = \mathbf{Q}^{-1} \quad (3.10)$$

Again, when referring to the elements of the same random vector, the weight matrix must also be symmetric.

The concept of weights has had extensive use in the classical theory of error and of adjustment. In fact, the term “weight” was used to express precision by way of inverse relationship. Thus high weight meant high precision, which in turn meant small standard deviation. However, this type of relationship is not true in general. In fact the use of the term “weight” relating to individual (diagonal) elements of \mathbf{W} should be totally restricted to the noncorrelation case. Only then is its use safe.

For the case of *no correlation* the covariance and cofactor matrices will both be diagonal, with all off-diagonal elements being equal to zero. In this case the diagonal elements are then

$$w_{ii} = \frac{1}{q_{ii}}$$

or, using equation (3.8),

$$w_{ii} = \frac{\sigma_0^2}{\sigma_{ii}^2} = \frac{c}{\sigma_{ii}^2}$$

These relationships are used often in textbooks on surveying. It must be stressed that they hold *only in the uncorrelated case*. In the general case, the diagonal elements w_{ii} of a weight matrix are *not* just the reciprocals of the associated cofactors,

$$w_{ii} \neq \frac{1}{q_{ii}}$$

It should be noted that in the case of perfect correlation, that is, $\rho = \pm 1$, between variables it is still possible to establish a covariance matrix and a cofactor matrix. *It is not possible, however, to establish a weight matrix in this case*. In addition, since covariances can be derived from observations directly, according to Equation (3.3), whereas weights cannot, the concept of weight is considered as a secondary concept. Caution must therefore be exercised when using weights.

3.4. BLUNDERS

Under the broad concept of error properties of observations, the conventional theory of errors also includes, in addition to random errors, blunders, or gross errors, as well as systematic errors. As regards blunders, the observer may take the wrong reading of a scale or a dial, or if he reads the proper value, he may record the wrong one by, for example, transposing numbers. If the operation of collecting the observations is performed through an automatic recording technique, mistakes may still occur due to failure of equipment, although they may be less frequent in this case. Another way of causing the occurrence of blunders is by failure in technique as in the case of reading the fraction on a tape on the wrong side of the zero mark, or by selecting the wrong whole degree in the measurement of angles that are very close to an integer of degrees. Finally, a mistake may also occur due to misinterpretation, such as sighting to the wrong target or selecting the wrong image for a control point in an aerial photograph. In fact, this last group of possibilities contains the more common causes of mistakes.

From a statistical point of view, blunders are observations that cannot be considered as belonging to the same samples from the distribution in question. They should not therefore be used together with other observations. Consequently, measurements should be planned and observational procedures designed in such a way as to allow for detecting blunders so they may be rejected. In practice, there are a variety of ways that can be employed: taking multiple readings and checking for reasonable consistency; careful checking of both pointing and recording; using simple and quick techniques for verification, applying logic and common sense; checking and verifying the performance of equipment, particularly those with automatic readouts; repeating the experiment with perhaps slightly different techniques; increasing the redundancy of the observations used in a model;

in case of relatively complex models, applying simplified geometric or algebraic checks to detect the mistakes; and, finally, simply noting that mistakes are deviations of large magnitude, which may lead directly to their detection. (However, blunders of moderate magnitude are hard to detect even with modern statistical techniques). For example,

1. Measuring an angle or a distance several times and computing the average. Any single measurement deviating from that average by an amount that is larger than a preset value can be assumed to contain a blunder.
2. Taking two readings by the transit differing by 180 degrees.
3. Realizing simple facts such as a tape is usually 100 feet long; vertical angles are normally between ± 90 degrees; when checking photocoordinates, the format is mostly 230×230 mm (for aerial photography).
4. Making a quick check on a spherical triangle using plane trigonometry.
5. Using solar time instead of sidereal time to check astro observations.
6. Using small portable calculators or even slide rules to check traverses and other nets to a low number of significant digits.

Despite design precautions, some blunders may still remain. Their detection and rejection should be carried out according to principles of statistical testing (data editing). It is worth mentioning here, that in practice too many observations are often discarded, although careful testing may prove the rejection to be unjustified.

3.5 SYSTEMATIC EFFECTS (ERRORS)

The term "systematic error," though used extensively in the past, is gradually being reconsidered. The practitioner is well aware of the existence of systematic effects and usually strives to minimize their presence through instrument calibration, observational procedures, and data preprocessing for atmospheric and other effects. However, from the statistical point of view it is essential to note that systematic errors will affect repeated observations in much the same way, hence they cannot be detected by having repetitive measurements. The concept of considering observations as samples from random probability distributions shows us where to place the effects of systematic errors. Although the variability of sampling relates to the deviations of the means of a distribution, the systematic effects of observations (or biases) concern its location parameter (μ). Thus perhaps with some exaggeration we can say that when systematic errors are present, there is nothing wrong with the observations; it is the interpretation that is wrong. In other words it is the functional model that is not appropriate. For example, a triangle on the earth's surface may be treated by one of three functional models: plane, spherical, or ellipsoidal. The choice of one over the others may result in systematic errors. Taking the concept of systematic error away

from the observation and attaching it to the functional model may seem to be of little practical importance, as this does not solve the problem of how to consider such effects. But it is of basic importance to understand that systematic errors indicate an inconsistency between the observations (or better, between the random variables to which the observations refer) and the functional model. The linking of the observations to the model, that is, assigning sample means (which are estimates for the expectations) to certain variables of the functional model, will be inconsistent when systematic errors are present.

Systematic effects take on different forms depending on the value and sign of each of the effects. If the value and sign remain the same all through the measuring process, we would have the so-called constant error. An example of this is making distance measurements with a tape that is either too short (or too long) by a constant value. All lengths measured by that tape will undergo the same systematic effect due to the tape alone. If the sign of the systematic effect changes, perhaps due to personal bias of an observer, the resulting systematic errors are often called "counteracting." For example, on an aerial photograph earth curvature and atmospheric refraction cause opposite displacements of image points. Thus the systematic effect due to the first counteracts that due to the second.

In photogrammetry, geodesy, and surveying, systematic errors occur due to physical causes, due to instrumental factors, and due to the observer's human limitations. Physical sources are numerous of which we mention temperature, humidity, and pressure changes. These will affect angle measurements and distance measurements either by tapes or electro-optical equipment, and will cause the bending of photogrammetric light rays due to atmospheric refraction, to name a few. Instrumental factors are caused by either imperfections in construction or lack of adequate adjustment of equipment before their use in data acquisition. Examples include unequal graduations on linear and circular scales, lack of centering of different components of the instrument, compromise in optical design thus leaving certain amounts of distortions and aberration (for example, photogrammetric camera lenses), and physical limitations in machining parts such as straight ways and pitch of screws in precision photogrammetric equipment.

Although automation has been considered and in some cases introduced (with its own sources of systematic effects) to several tasks, the human observer remains an important element in the activities of photogrammetry, geodesy, and surveying. Of his natural senses, he relies most on his vision and hearing abilities, both of which have limitations, and vary due to circumstances and from one individual to another. Although some of the personal systematic errors are constant and some are counteracting, a lot more may be erratic.

As a matter of principle, the functional model should be set up such that systematic effects are accounted for. This can be done in several ways:

1. The effects of systematic errors can be determined separately and considered directly a priori to the adjustment.
2. The effects can be considered as unknown random variables to be assessed by the observations during the adjustment.
3. The effects can be treated as observations themselves, perhaps with a priori given numerical values, for instance, refraction coefficients.

It is interesting to note that systematic effects are often not taken into account by the functional model but rather by “correcting” the observations for them. In other words the observations are replaced by another set of values, which are then treated as observations and which are more consistent with the functional model. It is said that the functional model is a “computational model” onto which the observations are “reduced.” Working with such a simplified computational model is preferred over the “best” functional model in the sense of a best description of the physical system.

As an example we consider the case in which image coordinates in an aerial photograph are corrected for lens distortion and refraction in order to maintain the fictitious functional model of a perspective image. Sometimes the image coordinates are even corrected for earth curvature in order to have a simpler computational system.

In order to eliminate the effects of systematic errors, they must first be detected. One should always strive to analyze the measuring process and the functional model in order to determine as well as possible the systematic effects. In some instances even the best modeling of the systematic errors leaves some effects in the observations. This is often detected by performing tests on the residuals from the adjustment (for example, using regression on the residuals from least squares computation). In such cases the functional model may be extended in such a way as to include some added parameters to absorb the effect of the remaining systematic errors. Alternatively, some interpolation treatment a posteriori to the adjustment may be performed on the results (see Chapter 14). Before closing this chapter, some examples may help in focusing the attention of the reader on how to detect and treat systematic errors.

Example 3.1. Let us consider the taking of aerial photography for the purpose of photogrammetric aerotriangulation. Object points on the surface of the earth emanate light rays that will eventually impinge on the photographic emulsion in the aerial camera. As these rays travel through the atmosphere, the air density changes (decreases) thus causing their refraction and bending before they reach the camera lens assembly. While passing through the lens, the rays undergo further deviation from straight line paths due to the radial and tangential distortions of the lens. Therefore the latent images registered on the negative material will be at positions different from those that would result from perfectly straight line rays (perspective image).

The negative material is brought back and then subjected to a variety of changeable conditions (stresses, temperature and humidity changes, and so on) during development and printing to make the photographic positive. Image positions undoubtedly change further during this process. What are now measured on a comparator are the images on the positive plate, which are obviously shifted from the perspective position. The magnitudes of these shifts depend on the amounts and types of all systematic effects, some of which have already been discussed.

Once the existence of systematic errors is ascertained, their effects are modeled functionally. When formulation is made to fit the systematic portion of the problem at hand, the correction procedure for those systematic effects is not difficult. For example, to compensate for radial lens distortion the image coordinates need to be shifted a certain amount which is a function of the radial distance from the principal point. Knowing the function and the radial distance for any one image point, the x and y shifts can be readily computed and (algebraically) added to the measured coordinates. Other systematic effects may be treated in the same manner and the total correction computed and applied to compensate for all effects, as far as possible.

Example 3.2. Example 3.1, from photogrammetry, exemplifies the procedure followed to trace in an orderly manner the sources of systematic effects. Examples of a variety of systematic errors also abound in geodetic and surveying operations. Let us consider the operation of *taping* for the determination of distances between points on the earth's surface. The length of a given tape may be physically different from the values indicated by the numbers written on its graduations due to some or all of the following factors:

1. The temperature changes between that used for tape standardization (calibration) and the temperature actually recorded in the field during observation.
2. The tension or pull applied to the tape during measurement is different from that used during calibration.
3. The method of tape support is different during measurement from that used during calibration.
4. The end points of the distance to be measured are at different elevation. In this case a correction is needed due to the fact that the slope distance would be measured instead of the horizontal distance.

Example 3.3. Electronic (and electro-optical) distance measuring techniques are also subject to a number of systematic effects whose sources and characteristics must be determined and alleviated. We mention a few of these sources here: There may be a change in the density of air through which the signal travels as it causes a change in the signal frequency (due to variations in wave propagation velocity); the instrument (and sometimes the reflector or remote unit) may not be properly centered on the end of the line to be measured; and the path of propagation of the signal may not conform to the straight-line assumption and may be bent due to environmental and other factors.

Example 3.4. Another instrument that is used extensively in surveying and geodesy to measure horizontal and vertical angles is the transit. The following are

some of the sources of systematic effects that may be associated with such instruments:

1. Horizontal circle may be off center.
2. Graduations on either or both circles may not be uniform.
3. The horizontal axis of the telescope (about which it rotates) may not be perpendicular to the vertical axis of the instrument.
4. The longitudinal (or optical) axis of the telescope may not be normal to the horizontal axis of rotation. (In this case, the axis of the telescope would describe a cone instead of a plane as it rotates through a complete revolution.)
5. When the optical axis of the telescope is horizontal, the reading on the vertical circle is different from zero.
6. The telescope axis and the axis of the leveling bubble may not be parallel.
7. This source does not pertain to the transit but to the target that is used to sight on. If the natural illuminating conditions are such that part of the target is in the shadow, the observer will tend to center the transit's cross hair so as to bisect the illuminated portion of the target. This type of error is often referred to as the phase error.

We have by no means exhausted all systematic sources in the general field of geometronics. There are numerous others in photogrammetry (particularly instrumental errors in stereoplotters, comparators, and so on) and geodesy (spherical and spheroidal excess, gravimeter and other instrument errors, and timing and other errors in astrogeodetic work). Systematic errors, once found, can often be accounted for as follows:

1. Actual formulation and computation of corrections that are then applied to the raw observations.
2. Careful calibration and adjustment of equipment and measuring under the same conditions specified by calibration results.
3. Devising observational procedures that will result in the elimination of systematic errors that would otherwise occur.
4. Extending the functional model to include the effects of the systematic errors.

Finally, systematic effects may arise due to highly correlated random errors that are not accommodated in the stochastic model. In fact, particularly after an adjustment, residual systematic effects are often treated as correlated random errors and filtering techniques are used for their reduction. It is important to note that in the phase on adjustment techniques in this book we do not deal with systematic effects or biases. The assumption will be made that both the stochastic and functional models are appropriate.

4

Principle and Techniques of Propagation

4.1. INTRODUCTION

Propagation involves obtaining the stochastic characteristics of (functionally) dependent variables given the characteristics of the independent variables and the functional relationships relating the two sets of variables.

Let \tilde{x}_i be a set of random variables with associated density function (n dimensional) $f(x_1, x_2, \dots)$. Let \tilde{y}_k be another set of random variables related to \tilde{x}_i by functional relationships

$$\tilde{y}_k = u_k(\tilde{x}_1, \tilde{x}_2, \dots)$$

Then the task of propagation is to determine the stochastic properties of \tilde{y}_k from those of the \tilde{x}_i , in other words, to determine basically the density function $f(y_1, y_2, \dots)$.

In practical application this general problem is often simplified. We can distinguish conventionally between three cases:

1. Propagation of means (expectations).
2. Propagation of random errors (variances and covariances).
3. Propagation of systematic errors.

Each of these cases will be treated separately, following the general case of distribution propagation. In all the cases, however, practical application is concerned with linear (or linearized) functional relationships only.

4.2. PROPAGATION OF DISTRIBUTIONS

Let $(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ be a set of stochastic variables, the joint probability distribution of which is given by the density function $f_x(x_1, x_2, \dots, x_n)$. From them a set of variables $(\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_m)$ is derived by given functional relationships. The joint probability distribution $f_y(y_1, y_2, \dots, y_m)$ is to be derived using f_x and the given functions.

4.2.1. One-Dimensional Case For simplicity, consider first a one-dimensional case. Let us take a continuous, differentiable function $\tilde{y} = g(\tilde{x})$ with $f(x)$ being the density function of \tilde{x} . The inverse function $\tilde{x} = h(\tilde{y})$ is assumed to exist and to be also unique, continuous, and differentiable.

Each event \tilde{x}_i can be transformed to an event $\tilde{y}_i = g(\tilde{x}_i)$ and, vice versa, $\tilde{x}_i = h(\tilde{y}_i)$. Therefore the probability of \tilde{x} falling into the interval $x_2 - x_1$ is the same as \tilde{y} falling into the associated interval $y_2 - y_1$.

$$P(x_1 \leq \tilde{x} \leq x_2) = \int_{x_1}^{x_2} f_x(x) dx = \int_{y_1}^{y_2} f_y(y) dy = P(y_1 \leq \tilde{y} \leq y_2)$$

Using the functional relationships, we can substitute

$$f_x(x) dx = \left[f_x(h(y)) \frac{\partial h(y)}{\partial y} \right] dy = [f_y(y)] dy$$

Thus the density function for the random variable $\tilde{y} = g(\tilde{x})$ is

$$f_y(y) = f_x(h(y)) \left| \frac{\partial h(y)}{\partial y} \right| \quad (4.1)$$

The absolute value used in equation (4.1) is taken in order to insure having the proper sign.

Example 4.1

$$y = x^2$$

$$g(x) = x^2; \quad f(x) = f_x(x)$$

$$h(y) = x = +\sqrt{y} \quad (\text{using only the positive value})$$

$$\frac{\partial h(y)}{\partial y} = \frac{1}{2\sqrt{y}}$$

density function for y : $f_y(y) = f_x(y) \cdot \frac{1}{2\sqrt{y}}$. Thus, if

$$f_x(x) = e^{-x^2/2}$$

then

$$f_y(y) = e^{-y/2} \left(\frac{1}{2\sqrt{y}} \right)$$

4.2.2. Multidimensional Case Let us consider the functions

$$\tilde{u}_i = g_i(\tilde{x}_1, \tilde{x}_2, \dots) = g_i(\tilde{\mathbf{x}}) \quad (4.2)$$

of the random vector $\tilde{\mathbf{x}}$ and the inverse functions

$$\tilde{x}_j = h_j(\tilde{u}_1, \tilde{u}_2, \dots) = h_j(\tilde{\mathbf{u}}) \quad (4.3)$$

If $f(x_1, x_2, \dots)$ is the joint density function of the random vector $\tilde{\mathbf{x}}$, then

$$\tilde{f}(u_1, u_2, \dots) = f[h_1(u_1, u_2, \dots), h_2(u_1, u_2, \dots), \dots] \cdot |\mathbf{J}| \quad (4.4)$$

is the joint density function of the derived random variables $\tilde{u}_1, \tilde{u}_2, \dots$. Here, $|\mathbf{J}|$ is the determinant of the Jacobian matrix of the inverse transformations

$$\tilde{x}_j = h_j(\tilde{\mathbf{u}}) \quad \text{or} \quad \mathbf{J} = \frac{\partial \mathbf{x}}{\partial \mathbf{u}}$$

The difficulty with propagation of distributions is that we have to assume that the inverse function exists. But in the case of linear transformation of normally distributed variables the result will also be normally distributed variables. Therefore propagation is generally limited to linear or linearized functions.

4.3. PROPAGATION OF MEANS

4.3.1. General Means or expectations of distributions are a special class of parameters by which probability distributions or density functions are defined. Therefore the general equation (4.4) must contain the special relationship for the propagation of means. Instead of using equation (4.4), however, it is simple to derive the propagation of means directly from the concept of statistical expectation introduced in Chapter 2. Let $\tilde{x}_1, \tilde{x}_2, \dots$ be a set of random variables with expectations (or means) $E(\tilde{x}_1), E(\tilde{x}_2), \dots$. We let $\tilde{y}_1, \tilde{y}_2, \dots$ be a set of random variables that are functions of $\tilde{x}_1, \tilde{x}_2, \dots$

$$\tilde{y}_i = g_i(\tilde{x}_1, \tilde{x}_2, \dots) \quad (4.5)$$

The expectation of \tilde{y}_i being the expectation of the function g_i is given by

$$E(\tilde{y}_i) = E(g_i(\tilde{x}_1, \tilde{x}_2, \dots)) \quad (4.6)$$

Equation (4.6) holds in general, but it does not provide a simple computational technique, although its general statistical definition makes possible the computation of $E(\tilde{y}_i)$ from

$$E(\tilde{y}_i) = \int \int_{-\infty}^{+\infty} \dots \int g_i(\tilde{x}_1, \tilde{x}_2, \dots) f(x_1, x_2, \dots) dx_1 dx_2, \dots \quad (4.7)$$

Here, $f(x_1, x_2, \dots)$ is the density function of $\tilde{x}_1, \tilde{x}_2, \dots$

4.3.2. Propagation of Means for Linear Functions For linear or linearized functions, however, the relationship for the propagation of expectations can be easily derived. In this case the general formula (4.7) simplifies considerably. For example, for $\tilde{y} = a\tilde{x}$ we get

$$E(\tilde{y}) = E(a\tilde{x}) = \int_{-\infty}^{\infty} axf(x) dx = a \int_{-\infty}^{\infty} xf(x) dx = aE(\tilde{x}) \quad (4.8)$$

Extending equation (4.8) to include several variables

$$\begin{aligned} \tilde{y}_i &= a_{0i} + a_{1i}\tilde{x}_1 + a_{2i}\tilde{x}_2 + a_{3i}\tilde{x}_3 + \cdots \\ E(\tilde{y}_i) &= E(a_{0i} + a_{1i}\tilde{x}_1 + a_{2i}\tilde{x}_2 + a_{3i}\tilde{x}_3 + \cdots) \\ &= E(a_{0i}) + E(a_{1i}\tilde{x}_1) + E(a_{2i}\tilde{x}_2) + E(a_{3i}\tilde{x}_3) + \cdots \\ &= a_{0i} + a_{1i}E(\tilde{x}_1) + a_{2i}E(\tilde{x}_2) + a_{3i}E(\tilde{x}_3) + \cdots \end{aligned} \quad (4.9)$$

or

$$\mu_{yi} = a_{0i} + a_{1i}\mu_{x1} + a_{2i}\mu_{x2} + a_{3i}\mu_{x3} + \cdots \quad (4.10)$$

Thus the linear relationship holds for the expectations or means as well. It is important to point out that this propagation of means for linear functions is independent of the probability distribution of the random variables $\tilde{x}_1, \tilde{x}_2, \dots$. Equation (4.10) combines the rules given in Section 2.4.1. When the relations are nonlinear, the probability distribution is needed.

Example 4.2. Given a normally distributed random variable \tilde{x} with

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right]$$

that is, with a mean μ and a variance of σ^2 . If $\tilde{y} = \tilde{x}^2$ is another random variable, evaluate its mean μ_y .

Solution

$$\begin{aligned} \mu_y &= E(\tilde{y}) = E(\tilde{x}^2) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] dx \end{aligned}$$

Let

$$\begin{aligned} z &= \frac{x - \mu}{\sigma\sqrt{2}} \\ z^2 &= \frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \\ x &= \mu + \sigma\sqrt{2}z \\ dx &= \sigma\sqrt{2} dz \end{aligned}$$

then

$$\begin{aligned}\mu_y &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma\sqrt{2}z)^2 e^{-z^2} (\sigma\sqrt{2}) dz \\ &= \frac{1}{\sqrt{\pi}} \left[\mu^2 \int_{-\infty}^{\infty} e^{-z^2} dz + 2\mu\sigma\sqrt{2} \int_{-\infty}^{\infty} ze^{-z^2} dz + 2\sigma^2 \int_{-\infty}^{\infty} z^2 e^{-z^2} dz \right] \\ &= \frac{1}{\sqrt{\pi}} \left[\sqrt{\pi}\mu^2 + 0 + 2\sigma^2 \frac{\sqrt{\pi}}{2} \right]\end{aligned}$$

or

$$\mu_y = \mu^2 + \sigma^2$$

This result could have been directly obtained from equation (2.38a) in Chapter 2, which, by the way, does not require knowledge of the density function.

In case \tilde{x} has a standardized distribution $N(0, 1)$, or $\mu = 0$ and $\sigma^2 = 1$, then

$$\mu_y = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-(1/2)x^2} dx$$

Let

$$t = \frac{x}{\sqrt{2}}$$

$$t^2 = \frac{1}{2}x^2$$

$$dx = \sqrt{2} dt$$

then

$$\begin{aligned}\mu_y &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 2t^2 e^{-t^2} \sqrt{2} dt \\ &= \frac{4}{\sqrt{\pi}} \int_0^{\infty} t^2 e^{-t^2} dt\end{aligned}$$

or

$$\mu_y = \frac{4}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{4} = 1$$

This result shows that whereas $\mu = 0$ (for \tilde{x}), $\mu_y = 0 + \sigma^2 = 0 + 1$ for \tilde{y} .

4.4. PROPAGATION OF VARIANCES AND COVARIANCES

4.4.1. General The propagation of variances and covariances is commonly known in practice as propagation of errors. In its general form it may be extracted as the second central moments from equation (4.4). It can in general be formulated for nonlinear functions in the following manner.

We let \tilde{x}_1, \tilde{x}_2 be random variables, with expectation μ_1, μ_2 and density function $f(x_1, x_2)$. We consider the functions

$$\tilde{y}_1 = g_1(\tilde{x}_1, \tilde{x}_2) \quad \text{and} \quad \tilde{y}_2 = g_2(\tilde{x}_1, \tilde{x}_2) \quad (4.11)$$

Variances and covariances are defined as the expectations

$$\sigma_{x_1}^2 = E[(\tilde{x}_1 - \mu_1)^2] = \int_{-\infty}^{\infty} (x_1 - \mu_1)^2 f(x_1) dx_1 \quad (4.12)$$

$$\sigma_{x_2}^2 = E[(\tilde{x}_2 - \mu_2)^2] = \int_{-\infty}^{\infty} (x_2 - \mu_2)^2 f(x_2) dx_2 \quad (4.13)$$

$$\begin{aligned} \sigma_{x_1 x_2} &= E[(\tilde{x}_1 - \mu_1)(\tilde{x}_2 - \mu_2)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_1)(x_2 - \mu_2) f(x_1, x_2) dx_1 dx_2 \end{aligned} \quad (4.14)$$

Accordingly we get for the variances and covariances of \tilde{y}_1, \tilde{y}_2

$$\sigma_{y_1}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_1(x_1, x_2) - \mu_{y_1})^2 f(x_1, x_2) dx_1 dx_2 \quad (4.15)$$

$$\sigma_{y_2}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_2(x_1, x_2) - \mu_{y_2})^2 f(x_1, x_2) dx_1 dx_2 \quad (4.16)$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_1(x_1, x_2) - \mu_{y_1})(g_2(x_1, x_2) - \mu_{y_2}) \\ &\quad \times f(x_1, x_2) dx_1 dx_2 \end{aligned} \quad (4.17)$$

Formulas (4.15) to (4.17) are general for any two functions of two random variables. They can readily be extended to the case of n random variables.

Such general formulas are not used in practice, however. The integrals involved in these equations often present problems. Therefore the propagation of variances and covariances is simplified to linear (or linearized) functions.

4.4.2. Variance and Covariance Propagation for Linear Functions Let \tilde{x}_1, \tilde{x}_2 be the random variables, with expectations μ_1, μ_2 and joint density function $f(x_1, x_2)$. Consider the linear functions

$$\begin{aligned} \tilde{y}_1 &= a_0 + a_1 \tilde{x}_1 + a_2 \tilde{x}_2 \\ \tilde{y}_2 &= b_0 + b_1 \tilde{x}_1 + b_2 \tilde{x}_2 \end{aligned} \quad (4.18)$$

The variances and covariances of \tilde{y}_1 , \tilde{y}_2 are derived by applying the definition,

$$\begin{aligned}
\sigma_{y_1}^2 &= E(\tilde{y}_1 - \mu_{y_1})^2 = E(a_0 + a_1 \tilde{x}_1 + a_2 \tilde{x}_2 - a_0 - a_1 \mu_1 - a_2 \mu_2)^2 \\
&= E(a_1(\tilde{x}_1 - \mu_1) + a_2(\tilde{x}_2 - \mu_2))^2 \\
&= E(a_1^2(\tilde{x}_1 - \mu_1)^2 + a_2^2(\tilde{x}_2 - \mu_2)^2 + 2a_1 a_2(\tilde{x}_1 - \mu_1)(\tilde{x}_2 - \mu_2)) \\
&= a_1^2 E(\tilde{x}_1 - \mu_1)^2 + a_2^2 E(\tilde{x}_2 - \mu_2)^2 + 2a_1 a_2 E(\tilde{x}_1 - \mu_1)(\tilde{x}_2 - \mu_2) \\
&= a_1^2 \sigma_{x_1}^2 + a_2^2 \sigma_{x_2}^2 + 2a_1 a_2 \sigma_{x_1 x_2}
\end{aligned} \tag{4.19}$$

Similarly,

$$\sigma_{y_2}^2 = b_1^2 \sigma_{x_1}^2 + b_2^2 \sigma_{x_2}^2 + 2b_1 b_2 \sigma_{x_1 x_2} \tag{4.20}$$

For the covariance, we have

$$\begin{aligned}
\sigma_{y_1 y_2} &= E(\tilde{y}_1 - \mu_{y_1})(\tilde{y}_2 - \mu_{y_2}) \\
&= E[(a_0 + a_1 \tilde{x}_1 + a_2 \tilde{x}_2 - a_0 - a_1 \mu_1 - a_2 \mu_2) \\
&\quad \times (b_0 + b_1 \tilde{x}_1 + b_2 \tilde{x}_2 - b_0 - b_1 \mu_1 - b_2 \mu_2)] \\
&= E[(a_1(\tilde{x}_1 - \mu_1) + a_2(\tilde{x}_2 - \mu_2)) \\
&\quad \times ((b_1(\tilde{x}_1 - \mu_1) + b_2(\tilde{x}_2 - \mu_2)))] \\
&= E(a_1 b_1(\tilde{x}_1 - \mu_1)^2 + 2(a_1 b_2 + a_2 b_1) \\
&\quad \times (\tilde{x}_1 - \mu_1)(\tilde{x}_2 - \mu_2) + a_2 b_2(\tilde{x}_2 - \mu_2)^2) \\
&= a_1 b_1 E(\tilde{x}_1 - \mu_1)^2 + a_2 b_2 E(\tilde{x}_2 - \mu_2)^2 + 2(a_1 b_2 + a_2 b_1) \\
&\quad \times E(\tilde{x}_1 - \mu_1)(\tilde{x}_2 - \mu_2) \\
&= a_1 b_1 \sigma_{x_1}^2 + a_2 b_2 \sigma_{x_2}^2 + 2(a_1 b_2 + a_2 b_1) \sigma_{x_1 x_2}
\end{aligned} \tag{4.21}$$

Again it is noticeable that the propagation of variances and covariances with linear functions is independent of the density functions. Equations (4.19) to (4.21) are valid for any probability distribution.

These equations were derived in a lengthy manner due to the application of the basic definitions of expectation. They can, however, be more concisely presented using matrices. Let

$$\tilde{\mathbf{y}} = [\tilde{y}_1 \tilde{y}_2]^t \quad \text{and} \quad \tilde{\mathbf{x}} = [\tilde{x}_1 \tilde{x}_2]^t$$

be the two random vectors involved. Equation (4.18) may be written in the form

$$\tilde{\mathbf{y}} = \mathbf{c} + \mathbf{C}\tilde{\mathbf{x}} \tag{4.22}$$

where

$$\mathbf{c} = [a_0 \quad b_0]^t$$

$$\mathbf{C} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}$$

Further, let

$$\Sigma_{xx} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 \end{bmatrix} \quad \Sigma_{yy} = \begin{bmatrix} \sigma_{y_1}^2 & \sigma_{y_1y_2} \\ \sigma_{y_1y_2} & \sigma_{y_2}^2 \end{bmatrix}$$

be the covariance matrices for the two random vectors $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$. The three equations (4.19), (4.20), and (4.21) may all be combined in one

$$\Sigma_{yy} = \mathbf{C} \Sigma_{xx} \mathbf{C}' \quad (4.23)$$

The (1, 1) terms in equation (4.23) give equation (4.19), the (1, 2) or (2, 1) terms give equation (4.20), and the (2, 2) terms give equation (4.21). The matrix \mathbf{C} represents the Jacobian of $\tilde{\mathbf{y}}$ with respect to $\tilde{\mathbf{x}}$; thus equation (4.23) may be further generalized symbolically to read

$$\Sigma_{yy} = \mathbf{J}_{yx} \Sigma_{xx} \mathbf{J}'_{yx} \quad (4.24)$$

with

$$\mathbf{J}_{yx} = \frac{\partial \tilde{\mathbf{y}}}{\partial \tilde{\mathbf{x}}} \quad (4.25)$$

If Σ_{xx} is a diagonal matrix, Σ_{yy} may still in general be a full matrix. Therefore even if the original variables are uncorrelated, the new ones are usually correlated.

4.4.3. Variance and Covariance Propagation for Nonlinear Functions Linear equations are not often encountered in practice, and therefore equations (4.18) are usually the result of linearizing functions that were originally nonlinear. For instance, if the original equations are those given by equation (4.11), their linearized form at the approximate (or initial) values x_1^0, x_2^0 (see Appendix B) will be identical to equation (4.18).

Of course, the linearized form is stopped after the zero and first-order terms neglecting the second and higher-order terms. In fact, the zero-order terms are not needed since $g_1(x_1^0, x_2^0)$ and $g_2(x_1^0, x_2^0)$, which correspond to a_0, b_0 of the directly linear form, would not appear in the forms for propagated $\sigma_{y_1}^2, \sigma_{y_1y_2}$, and $\sigma_{y_2}^2$. This leaves only the partial derivatives and thus makes equation (4.24) general, in as much as it can be applied to linear and

nonlinear cases alike. Here, the Jacobian matrix \mathbf{J} would include all four partial derivatives or

$$\mathbf{J} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{bmatrix} \quad \text{evaluated at } x_1^0, x_2^0 \quad (4.26)$$

Although in practical applications linearized functions are used regularly for the propagation of variances and covariances, it should be pointed out that this is permitted if the range of dispersion in \tilde{x}_1, \tilde{x}_2 is small when linear approximation is compared to the curvature of the function in the neighborhood of x_1^0, x_2^0 . In other words the function should be approximated well by its tangent within the region of interest—that is, the region of dispersion of the random variables.

From a formal point of view it should be noted that in linearization the properties of random variables change from the variables themselves (in the nonlinear form) to the increment,

$$x_i = x_{i0} + \Delta x_i \quad y_i = y_{i0} + \Delta y_i$$

Thus the error properties—that is, the probability distribution—is now associated with Δx_i and Δy_i instead of x_i, y_i , respectively.

Within the range of linearity, the geometrical consideration confirms the result, obtained previously, that linear transformation of normally distributed variables \tilde{x}_i also yields normally distributed variables \tilde{y}_i . It also allows us to assess the validity of the linearization (see Figure 4.1).

As a final comment, propagation rules for other moments could be derived in the same way as for the second moments in the case of linear

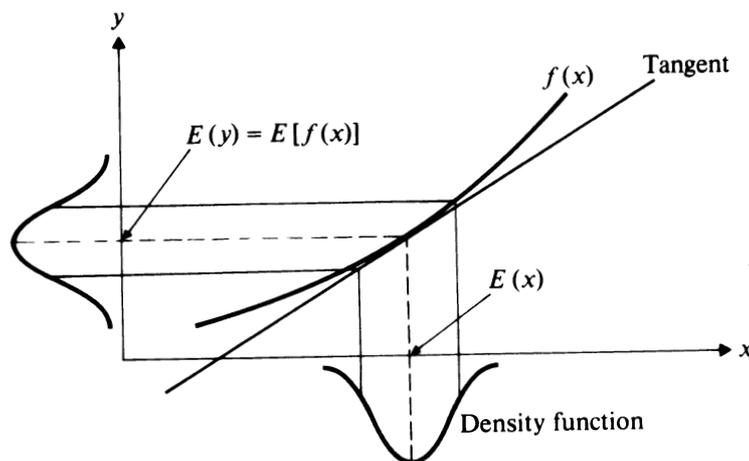


Figure 4.1

functions. In general, the propagation of moments is independent of the distribution with linear functions.

4.4.4. Alternate Derivation for Variance and Covariance Propagation A derivation, which is due to Brown,[†] is possible without the application of the concept of moments. It is based on the following alternate definitions of variance and covariance. If \tilde{x}_i is a random variable taking many values x_{ik} , each of which having a random error dx_{ik} (as a deviation from the mean), then the variance σ_{xi}^2 of x_i is given by

$$\sigma_{xi}^2 = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m (dx_{ik})^2 \quad (4.27)$$

In a similar manner the covariance of two elements \tilde{x}_i, \tilde{x}_j of a random vector $\tilde{\mathbf{x}}$ may be defined by

$$\sigma_{xixj} = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m (dx_{ik} dx_{jk}) \quad (4.28)$$

With these two basic definitions, the propagation of variances and covariances from one set of random variables $\tilde{\mathbf{x}}$ to another set $\tilde{\mathbf{y}}$ can be formulated. We let

$$\begin{aligned} \tilde{y}_1 &= y_1(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) \\ &\vdots \\ \tilde{y}_q &= y_q(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) \end{aligned} \quad (4.29)$$

be a set of q functions relating the variables $\tilde{\mathbf{x}}$ to $\tilde{\mathbf{y}}$. In order to propagate the "errors" $dx_{1k}, dx_{2k}, \dots, dx_{nk}$ into all \tilde{y}_i , let us define the auxiliaries (partial derivatives),

$$y'_{ij} = \frac{\partial y_i}{\partial x_j} \quad i = 1, 2, \dots, q, \quad j = 1, 2, \dots, n \quad (4.30)$$

Thus

$$\begin{aligned} dy_{1k} &= y'_{11} dx_{1k} + y'_{12} dx_{2k} + \dots + y'_{1n} dx_{nk} \\ &\vdots \\ dy_{qk} &= y'_{q1} dx_{1k} + y'_{q2} dx_{2k} + \dots + y'_{qn} dx_{nk} \end{aligned} \quad (4.31)$$

or in matrix form

$$dy_k = Y' dx_k \quad (4.32)$$

[†] See bibliography at end of book.

where dy_k is a $q \times 1$ vector, and dx_k is a $n \times 1$ vector; Y' is a $q \times n$ matrix of partial derivatives which is the same as the Jacobian J_{yx} . Hence, equation (4.32) may be written as

$$dy_k = J_{yx} dx_k \quad (4.33)$$

If we choose the i th and r th elements to represent any two typical elements, then

$$\begin{aligned} dy_{ik} &= \mathbf{j}_i dx_k \\ dy_{rk} &= \mathbf{j}_r dx_k \end{aligned} \quad (4.34)$$

where \mathbf{j}_i , \mathbf{j}_r are the i th and r th rows of \mathbf{J} . Next,

$$dy_{ik} dy_{rk} = \mathbf{j}_i dx_k \mathbf{j}_r dx_k$$

or

$$dy_{ik} dy_{rk} = \mathbf{j}_i dx_k (dx_k)^t \mathbf{j}_r^t \quad (4.35)$$

Let

$$\mathbf{M}_k = dx_k \cdot (dx_k)^t = \begin{bmatrix} (dx_1)^2 & (dx_1 dx_2) & \cdots & (dx_1 dx_n) \\ \vdots & \vdots & & \vdots \\ (dx_n dx_1) & (dx_n dx_2) & \cdots & (dx_n)^2 \end{bmatrix}_k \quad (4.36)$$

then

$$dy_{ik} dy_{rk} = \mathbf{j}_i \mathbf{M}_k \mathbf{j}_r^t \quad (4.37)$$

We let k run from 1 to m , take the sum, divide by m , and take the limit, or

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m dy_{ik} dy_{rk} = \mathbf{j}_i \left[\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m \mathbf{M}_k \right] \mathbf{j}_r^t \quad (4.38)$$

In view of the definitions of equations (4.27) and (4.28), it is readily seen that equation (4.38) leads to

$$\sigma_{y_i y_r} = \mathbf{j}_i \Sigma_{xx} \mathbf{j}_r^t \quad (4.39)$$

Equation (4.39) represents *one element* on the i th row and r th column of the total covariance matrix Σ_{yy} . It should be rather straightforward to extend equation (4.39) to

$$\Sigma_{yy} = \mathbf{J}_{yx} \Sigma_{xx} \mathbf{J}_{yx} \quad (4.40)$$

which is identical to equation (4.24). Equation (4.40) is sufficiently general in that it concerns a vector $\tilde{y}_{(q, 1)}$ as a general (nonlinear) function of another vector $\tilde{x}_{(n, 1)}$ which is of different dimension. Since cofactor matrices are related to covariance matrices by only a scale factor, that is,

$$\mathbf{Q}_{xx} = \frac{1}{\sigma_0^2} \boldsymbol{\Sigma}_{xx} \quad \text{and} \quad \mathbf{Q}_{yy} = \frac{1}{\sigma_0^2} \boldsymbol{\Sigma}_{yy}$$

Then equation (4.40) may be written in terms of cofactor matrices

$$\mathbf{Q}_{yy} = \mathbf{J}_{yx} \mathbf{Q}_{xx} \mathbf{J}'_{yx} \quad (4.41)$$

It is worth noting here that we may speak of covariance and cofactor propagation but *not of weight* propagation. This follows in view of the fact that earlier discussion showed that weight matrices may not even be defined. Furthermore, the Jacobian matrices are often rectangular and the inverse of equation (4.41) [or equation (4.40)] cannot be effected on individual matrices.

A further generalization of equation (4.41) is possible when multiple functions of different random vectors are involved. For example, let $\mathbf{x}_{(n, 1)}$ and $\mathbf{t}_{(m, 1)}$ be two correlated vectors with cofactor matrices \mathbf{Q}_{xx} , \mathbf{Q}_{xt} and \mathbf{Q}_{tt} . Two other vectors $\mathbf{y}_{(q, 1)}$ and $\mathbf{z}_{(p, 1)}$ are functions of \mathbf{x} and \mathbf{t} (\mathbf{y} is functionally independent of \mathbf{t} and \mathbf{z} of \mathbf{x})

$$\mathbf{y} = \mathbf{y}(\mathbf{x}) \quad (4.42)$$

$$\mathbf{z} = \mathbf{z}(\mathbf{t})$$

from which the following Jacobians are derived

$$\mathbf{J}_{yx} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \quad (4.43)$$

$$\mathbf{J}_{zt} = \frac{\partial \mathbf{z}}{\partial \mathbf{t}}$$

Two expanded vectors

$$\mathbf{r} = \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} \quad \text{and} \quad \mathbf{s} = \begin{bmatrix} \mathbf{x} \\ \mathbf{t} \end{bmatrix}$$

can be set up with the corresponding expanded Jacobian

$$\mathbf{J}_{rs} = \begin{bmatrix} \mathbf{J}_{yx} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{zt} \end{bmatrix} \quad (4.44)$$

and equation (4.41) may be applied

$$\mathbf{Q}_{rr} = \mathbf{J}_{rs} \mathbf{Q}_{ss} \mathbf{J}'_{rs}$$

or

$$\begin{aligned} \begin{bmatrix} Q_{yy} & Q_{yz} \\ Q_{zy} & Q_{zz} \end{bmatrix} &= \begin{bmatrix} J_{yx} & 0 \\ 0 & J_{zt} \end{bmatrix} \begin{bmatrix} Q_{xx} & Q_{xt} \\ Q_{tx} & Q_{tt} \end{bmatrix} \begin{bmatrix} J'_{yx} & 0 \\ 0 & J'_{zt} \end{bmatrix} \\ &= \begin{bmatrix} J_{yx} Q_{xx} J'_{yx} & J_{yx} Q_{xt} J'_{zt} \\ J_{zt} Q_{tx} J'_{yx} & J_{zt} Q_{tt} J'_{zt} \end{bmatrix} \end{aligned}$$

from which we can write the following four general relationships

$$Q_{yy} = J_{yx} Q_{xx} J'_{yx} \quad (4.45a)$$

$$Q_{yz} = J_{yx} Q_{xt} J'_{zt} \quad (4.45b)$$

$$Q_{zy} = J_{zt} Q_{tx} J'_{yx} = Q'_{yz} \quad (4.45c)$$

$$Q_{zz} = J_{zt} Q_{tt} J'_{zt} \quad (4.45d)$$

The crosscofactor matrices are easily constructed as, for example,

$$Q_{xt} = \begin{bmatrix} q_{x1t1} & q_{x1t2} & \cdots & q_{x1tm} \\ \vdots & \vdots & & \vdots \\ q_{xnt1} & q_{xnt2} & \cdots & q_{xntm} \end{bmatrix} \quad (4.46)$$

which directly leads to the fact that

$$Q_{tx} = Q'_{xt}$$

4.4.5. Symbolic Multiplication In an effort to assist the practitioner, Tienstra† devised a scheme for the application of propagation rules that he called “symbolic multiplication.” It is a strictly mnemotechnical rule that makes possible obtaining the elements of the covariance matrix separately. It is briefly given here for the sake of those readers consulting European literature.

In order to illustrate the scheme, we consider the following equations,

$$u = a_0 + a_1 x + a_2 y + a_3 z \quad (4.47)$$

$$v = b_0 + b_1 x + b_2 y + b_3 z$$

According to equation (4.41), the cofactors of u, v are

$$\begin{aligned} q_{uu} &= a_1^2 q_{xx} + a_2^2 q_{yy} + a_3^2 q_{zz} + 2a_1 a_2 q_{xy} + 2a_1 a_3 q_{xz} + 2a_2 a_3 q_{yz} \\ q_{vv} &= b_1^2 q_{xx} + b_2^2 q_{yy} + b_3^2 q_{zz} + 2b_1 b_2 q_{xy} + 2b_1 b_3 q_{xz} + 2b_2 b_3 q_{yz} \\ q_{uv} &= a_1 b_1 q_{xx} + a_2 b_2 q_{yy} + a_3 b_3 q_{zz} + (a_1 b_2 + a_2 b_1) q_{xy} \\ &\quad + (a_1 b_3 + a_3 b_1) q_{xz} + (a_2 b_3 + a_3 b_2) q_{yz} \end{aligned} \quad (4.48)$$

† See bibliography at end of book.

The coefficients of the cofactors on the right-hand side of equation (4.48) can be obtained directly from corresponding multiplication of the original equations themselves [equations (4.47)] disregarding a_0 , b_0 . For instance, the product (uu) would be

$$(uu) = a_1^2(xx) + a_2^2(yy) + a_3^2(zz) + 2a_1 a_2(xy) + 2a_1 a_3(xz) + 2a_2 a_3(yz) \quad (4.49)$$

The terms between parentheses refer to the indices of cofactors.

Comparing equation (4.49) to the first line of equation (4.48) shows that such a “multiplication” does in fact yield the coefficients of the cofactors. Thus with the understanding that products of random variables are shifted to indices of cofactors, any auto- or crosscofactor can be obtained by “symbolic multiplication” of the equations. It was devised before the extensive use of matrix algebra, and was helpful in particular when a single element of the total cofactor matrix is required. Recently the same scheme has also been applied to matrix equations. Of course, in this case the matrix sequence must be carefully observed, thus leading to transposition. For example, for

$$\begin{aligned} \mathbf{y} &= \mathbf{Ax} + \mathbf{a} \\ \mathbf{z} &= \mathbf{Bt} + \mathbf{b} \end{aligned}$$

The crosscofactor matrix \mathbf{Q}_{yz} may be obtained by “symbolic multiplication” of \mathbf{y} and \mathbf{z} (again disregarding the terms \mathbf{a} and \mathbf{b} which are irrelevant to propagation). Thus

$$\mathbf{yz}^{(t)} = (\mathbf{Ax})(\mathbf{Bt})^t = \mathbf{A}(\mathbf{xt}^{(t)})\mathbf{B}^t \quad (4.50)$$

The random vectors are transferred to indices for cofactor matrices and dropping the transpose symbol, or

$$\mathbf{Q}_{yz} = \mathbf{AQ}_{xt}\mathbf{B}' \quad (4.51)$$

Noting that $\mathbf{J}_{yx} = \mathbf{A}$ and $\mathbf{J}_{zt} = \mathbf{B}$, equation (4.51) becomes identical to equation (4.45b).

4.4.6. Further Elaboration on Propagation Using Matrices Matrix formulation expounded in the preceding sections allows for the propagation of variances through several transformations. As an example, we consider the following three relations:

$$\begin{aligned} \mathbf{y} &= \mathbf{Ax} + \mathbf{a} \\ \mathbf{z} &= \mathbf{By} + \mathbf{b} \\ \mathbf{r} &= \mathbf{Cz} + \mathbf{c} \end{aligned} \quad (4.52)$$

Let the random vector $\tilde{\mathbf{x}}$ be known, with its cofactor matrix \mathbf{Q}_{xx} . Then there are two ways of obtaining the cofactor matrices of \mathbf{z} and \mathbf{r} , known as “substitution” and “stepwise execution.”

PROPAGATION THROUGH SUBSTITUTION We express z and r in terms of x , by substitution:

$$\begin{aligned}
 y &= Ax + a \\
 z &= By + b = B(Ax + a) + b = (BA)x + (Ba + b) \\
 r &= Cz + c = C(By + b) + c = (CB)y + (Cb + c) \\
 &= (CB)(Ax + a) + (Cb + c) \\
 &= (CBA)x + (CBa + Cb + c)
 \end{aligned} \tag{4.53}$$

Applying propagation of cofactor relations in equation (4.45) (or using symbolic multiplication) we get

$$\begin{aligned}
 Q_{yy} &= AQ_{xx} A^t \\
 Q_{zz} &= (BA)Q_{xx}(BA)^t = BAQ_{xx} A^t B^t \\
 Q_{rr} &= (CBA)Q_{xx}(CBA)^t = CBAQ_{xx} A^t B^t C^t \\
 Q_{yz} &= AQ_{xx}(BA)^t = AQ_{xx} A^t B^t \\
 Q_{yr} &= AQ_{xx}(CBA)^t = AQ_{xx} A^t B^t C^t \\
 Q_{zr} &= BAQ_{xx}(CBA)^t = BAQ_{xx} A^t B^t C^t
 \end{aligned} \tag{4.54}$$

STEPWISE PROPAGATION The same result in equation (4.54) can be obtained by applying propagation in steps as follows:

$$\begin{aligned}
 y &= Ax + a & Q_{yy} &= AQ_{xx} A^t \\
 z &= By + b & Q_{zz} &= BQ_{yy} B^t = BAQ_{xx} A^t B^t \\
 r &= Cz + c & Q_{rr} &= CQ_{zz} C^t = CBAQ_{xx} A^t B^t C^t \\
 & & Q_{yz} &= AQ_{xy} B^t \\
 & & Q_{yr} &= AQ_{xz} C^t \\
 & & Q_{zr} &= BQ_{yz} C^t
 \end{aligned} \tag{4.55}$$

The last three (crosscofactor) relations of equation (4.55) do not correspond to those in equation (4.54), particularly because of the absence of the matrices Q_{xy} , Q_{xz} , and Q_{yz} . However, these matrices can be derived if equations (4.52) are supplemented by simple identities. Thus

$$\begin{aligned}
 x &= Ix \\
 y &= Ax + a \\
 x &= Ix \\
 z &= By + b \\
 y &= Iy \\
 z &= By + b
 \end{aligned}$$

and

$$\begin{aligned} \mathbf{Q}_{xy} &= \mathbf{I}\mathbf{Q}_{xx} \mathbf{A}' = \mathbf{Q}_{xx} \mathbf{A}' \\ \mathbf{Q}_{xz} &= \mathbf{Q}_{xy} \mathbf{B}' = \mathbf{Q}_{xx} \mathbf{A}'\mathbf{B}' \\ \mathbf{Q}_{yz} &= \mathbf{Q}_{yy} \mathbf{B}' \end{aligned} \quad (4.56)$$

Substituting equation (4.56) in equation (4.55) leads directly to the equations in (4.54), thus ascertaining the equivalence of both procedures.

Example 4.3. Given

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \quad \Sigma_{yy} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}; \quad z = x_1 + x_2$$

1. Compute σ_z^2 , Σ_{zy} , and Σ_{xy}
2. Compute the elements of the standard ellipse for \mathbf{x} and \mathbf{z} .

Solution: 1. First

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

and

$$\Sigma_{xx} = \mathbf{J}_{xy} \Sigma_{yy} \mathbf{J}'_{xy} = \frac{1}{9} \begin{bmatrix} 11 & -7 \\ -7 & 11 \end{bmatrix}$$

$$\sigma_z^2 = \mathbf{J}_{zx} \Sigma_{xx} \mathbf{J}'_{zx} = \frac{8}{9}$$

$$\Sigma_{zy} = \mathbf{J}_{zx} \Sigma_{xx} \mathbf{J}'_{yx} = \frac{1}{3} [4 \quad 4]$$

$$\Sigma_{xy} = \mathbf{J}_{xy} \Sigma_{yy} \mathbf{J}'_{yy} = \frac{1}{3} \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}$$

2. (a) Standard ellipse for \mathbf{x} : The characteristic equation is

$$\lambda^2 + \text{tr}(\Sigma_{xx})(-\lambda) + |\Sigma_{xx}| = 0$$

or

$$\lambda^2 - \frac{22}{9}\lambda + \frac{8}{9} = 0$$

thus

$$\lambda_1 = 2 \quad \text{and} \quad \lambda_2 = \frac{4}{9}$$

For $\lambda_1 = 2$ the eigenvector is $(x_1, -x_1)$, and for $\lambda_2 = \frac{4}{9}$ the eigenvector is (x_1, x_1) . Hence the semimajor axis of the standard ellipse is $a = \sqrt{\lambda_1} = \sqrt{2}$ and is oriented along a line making an angle -45° with the x_1 axis. The semiminor axis is normal to it and is equal to $b = \sqrt{\lambda_2} = \frac{2}{3}$.

(b) Standard ellipse for \mathbf{z} : Since this is a one-dimensional case, the ellipse degenerates into a line segment with $+\sigma_z = (2\sqrt{2})/3$ along the positive z axis from the estimate of the point, and $-\sigma_z$ in the opposite direction.

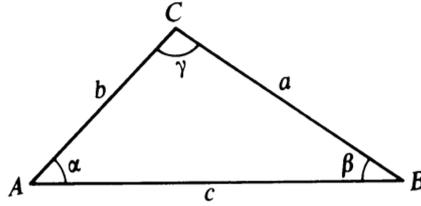


Figure 4.2

Example 4.4. Figure 4.2 shows a triangle for which

$$f = \alpha + \beta + \gamma - (\pi + \varepsilon)$$

represents the lack of closure, with ε being the spherical excess.

1. If $s_\alpha, s_\beta, s_\gamma$ are the standard deviations for the three angles, which are assumed to be uncorrelated, then

$$s_f^2 = [1 \quad 1 \quad 1] \begin{bmatrix} s_\alpha^2 & 0 & 0 \\ 0 & s_\beta^2 & 0 \\ 0 & 0 & s_\gamma^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = s_\alpha^2 + s_\beta^2 + s_\gamma^2$$

If, further, $s_\alpha = s_\beta = s_\gamma = s$, then

$$s_f^2 = 3s^2$$

2. Each angle may be considered as the difference between two directions that are uncorrelated and of equal precision, s_d^2 . For example: $\alpha = d_1 - d_2$ and

$$s^2 = s_\alpha^2 = [1 \quad -1] \begin{bmatrix} s_d^2 & 0 \\ 0 & s_d^2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2s_d^2$$

Thus

$$s_f^2 = 6s_d^2 \quad \text{or} \quad s_d^2 = \frac{1}{6}s_f^2$$

The latter relationship is often referred to as Ferrero's.

Example 4.5. In the plane triangle of Figure 4.2, the two angles α and β and the side c are measured and known to be uncorrelated. It is desired to derive the autocorrelation factor of side a .

Solution:

$$a = \frac{c \sin \alpha}{\sin (\pi - \alpha - \beta)}$$

$$\mathbf{J} = \left[c \frac{(\cos \alpha \sin \gamma \sin \alpha \cos \gamma)}{\sin^2 \gamma} \quad \frac{c \sin \alpha \cos \gamma}{\sin^2 \gamma} \quad \frac{\sin \alpha}{\sin \gamma} \right]$$

$$= \left[(a \cot \alpha + a \cot \gamma) \quad a \cot \gamma \quad \frac{a}{c} \right]$$

$$q_{aa} = a^2(\cot \alpha + \cot \gamma)^2 q_{\alpha\alpha} + a^2 \cot^2 \gamma q_{\beta\beta} + \frac{a^2}{c^2} q_{cc}$$

Another procedure is as follows:

$$a = \frac{\sin \alpha}{\sin \gamma} \quad \gamma = \pi - \alpha - \beta$$

$$\mathbf{J} = \begin{bmatrix} a \cot \alpha & -a \cot \gamma & \frac{a}{c} \end{bmatrix}$$

$$q_{aa} = a^2 \cot^2 \alpha q_{\alpha\alpha} + a^2 \cot^2 \gamma q_{\gamma\gamma} - 2a^2 \cot \alpha \cot \gamma q_{\alpha\gamma} + \frac{a^2}{c^2} q_{cc}$$

Since $q_{\gamma\gamma} = q_{\alpha\alpha} + q_{\beta\beta}$ and $q_{\alpha\gamma} = -q_{\alpha\alpha}$, then q_{aa} reduces to the same expression above.

4.5. PROPAGATION OF SYSTEMATIC AND “TRUE” ERRORS

In this section we shall discuss briefly the propagation of “known” terms through known functions. Let \tilde{x} , \tilde{y} be random variables, for which the following function is given

$$r = g(x, y) \quad (4.57)$$

Let Δx , Δy be increments, the nature of which will be discussed below. Then we have directly for the propagation of the increments

$$\begin{aligned} r + \Delta r &= g(x + \Delta x, y + \Delta y) \\ \Delta r &= g(x + \Delta x, y + \Delta y) - g(x, y) \end{aligned} \quad (4.58)$$

If the increments are small, the function can be expanded and only linear terms used. Thus

$$r + \Delta r = g(x, y) + \frac{\partial g}{\partial x} \Delta x + \frac{\partial g}{\partial y} \Delta y + \dots$$

or

$$\Delta r = \frac{\partial g}{\partial x} \Delta x + \frac{\partial g}{\partial y} \Delta y \quad (4.59)$$

To this point, no probability considerations enter into the picture. It is treated as a calculus problem as far as the formula (4.59) is concerned. In practical applications, however, the increments Δx , Δy are in a way related to the error concept. They can be (a) corrections due to computational errors; (b) “true” errors that are assumed to be known; or (c) systematic errors.

It is essential that in all cases Δx , Δy are supposed to be known and not associated with random variables. This is, of course, a simplification. From a statistical point of view the case can be considered as a special case of propagation of means. The random variables \tilde{x} , \tilde{y} are replaced by derived variables $u = x + \Delta x$, $v = y + \Delta y$, where Δx and Δy are also, in principle,

random variables. We ask here for the expectation of $r = g(u, v)$ as compared to that of $r = g(x, y)$.

The above rule in equation (4.59) can be considered a limit case when Δx , Δy are considered as constants, their stochastic properties being reduced to zero. This propagation rule is from a theoretical point of view of very limited importance. However, it is useful in practice for studying certain effects, in particular in connection with simplified mathematical models. Systematic errors are, in principle, deficiencies of the mathematical model. Consequently, such rules can be used in studying the propagation of residual systematic effects.

Problems for Part I

1. From the bivariate normal density function for two correlated random variables \tilde{x} , \tilde{y} [see equation (2.70)], derive the marginal distributions for each variable \tilde{x} and \tilde{y} .
2. Given the bivariate normal density function $f(x, y)$ [equation (2.70)] derive the conditional density function $f(x|y)$.
3. From equation (2.69) give the (noncentral) joint normal density function for the variables \tilde{x} , \tilde{y} , \tilde{z} such that there is no correlation between \tilde{x} and \tilde{z} ($\rho_{xz} = 0$) and between \tilde{y} and \tilde{z} ($\rho_{yz} = 0$). Show that the joint density can be written as the product of the bivariate density function $f(x, y)$ and the univariate density function $f(z)$.
4. Show that the mean is zero and the variance is equal to one for the standardized random variable

$$\tilde{z} = \frac{\tilde{x} - E(\tilde{x})}{\sigma_x}$$

5. Derive the mean and variance of the χ^2 distribution, given in equations (2.61) and (2.62).
6. Given the following 20 measurements of an angle in order of their observation:

(a) 9.3"	(b) 6.6"	(c) 7.9"	(d) 6.8"
(e) 6.1"	(f) 5.9"	(g) 6.1"	(h) 7.8"
(i) 7.2"	(j) 8.0"	(k) 7.1"	(l) 4.9"
(m) 5.7"	(n) 5.2"	(o) 6.9"	(p) 6.2"
(q) 10.1"	(r) 9.7"	(s) 4.1"	(t) 6.2"

Sketch a histogram of the measurements using an interval of 0.5" and a lower bound of 4.0" and compute the sample mean value, the sample standard deviation of a single measurement, and of the mean value.

7. Given a random variable \tilde{y} with standard normal distribution, what is the probability that (a) \tilde{y} assumes a value between 0.35 and 1.65; (b) \tilde{y} assumes a value between -1.96 and 2.42 .

8. The distribution of the diameters of the ball bearings in a certain shipment is approximately normal with mean 0.500 cm and standard deviation 0.010 cm. If a ball bearing is effective when its diameter lies between 0.490 and 0.515 cm, find the probability of obtaining an effective ball bearing from the shipment. (Use the standard normal tables.)

9. The following observations were recorded for the measurement of two angles, $\tilde{\alpha}$ and $\tilde{\beta}$, each measured 20 times.

OBSERVATION	$\tilde{\alpha}$	$\tilde{\beta}$
1	20° 05' 17.9"	16° 42' 29.8"
2	17.0	29.9
3	19.1	30.1
4	17.1	31.3
5	17.7	31.3
6	16.8	30.9
7	17.6	30.7
8	18.5	29.5
9	17.6	30.0
10	16.4	31.2
11	17.5	30.2
12	18.4	30.2
13	17.6	30.6
14	16.6	29.1
15	18.2	30.3
16	17.8	29.9
17	17.0	29.8
18	18.3	29.5
19	17.7	29.6
20	18.0	30.6

(a) Compute the sample variance and sample standard deviation for a single observation of each of the angles $\tilde{\alpha}$ and $\tilde{\beta}$. (b) Compute the sample covariance and estimate of the correlation coefficient between $\tilde{\alpha}$ and $\tilde{\beta}$. (Assume $\tilde{\alpha}$ and $\tilde{\beta}$ are the random variables of a bivariate distribution.) (c) Write out the variance-covariance matrix for the observations $\tilde{\alpha}$, $\tilde{\beta}$.

10. A distance x is measured nine times with sample mean of 190.09 meters and sample variance of the mean of 4 cm^2 . Present the confidence intervals for μ_x and σ_x at the 95% and 99% levels.

11. A distance x is measured 30 times with sample mean of 190.09 meters and sample standard deviation of a single value of 6 cm. Present the confidence intervals for μ_x and σ_x at the 95% and 99% levels.

12. A sample of nine measurements of a distance yields a mean value of 8.4 cm and a standard deviation in the mean value of 0.4 mm. Test the hypothesis that the true length of the distance is 8.0 cm against the alternative that the true length is greater than 8.0 cm. Use a 5% level of significance.

13. The average of 100 measurements is 21.0 cm with a standard deviation in a single measurement of 1.5 cm. Can it be claimed that the true length of the measurement exceeds 20.0 cm? Use a 5% level of significance.

14. The following 15 measurements were observed and assumed to be from a normal distribution:

132.2	128.0	125.0	130.2	126.6
128.6	127.8	128.4	128.0	128.4
128.0	128.5	128.4	128.7	127.2

Test the following hypotheses at a 5% level of significance:

- a. $H_0: \mu = 130.0$ against $H_1: \mu \neq 130.0$
- b. $H_0: \mu = 130.0$ against $H_1: \mu < 130.0$
- c. $H_0: \sigma^2 = 9.0$ against $H_1: \sigma^2 \neq 9.0$
- d. $H_0: \sigma^2 = 5.0$ against $H_1: \sigma^2 < 5.0$

15. The following two sets of measurements were supposedly taken of the same physical quantity by identical techniques:

Sample I:	0.806	0.827	0.816	0.809	0.814
Sample II:	0.813	0.803	0.810	0.807	0.804

Test the following hypotheses at a 5% level of significance:

- a. $H_0: \mu_I - \mu_{II} = 0$ against $H_1: \mu_I - \mu_{II} \neq 0$
- b. $H_0: \sigma_I^2 = \sigma_{II}^2$ against $H_1: \sigma_I^2 \neq \sigma_{II}^2$

16. The random variable \tilde{X} has mean μ_x . The random variable \tilde{Y} has mean μ_y . If a and b are constants, determine the expected value of \tilde{Z} in terms of a, b, μ_x , and μ_y if $\tilde{Z} = \tilde{X} - a\tilde{X} + \tilde{Y} - b\tilde{Y}$.

17. A sample of 12 measurements of an angle θ yields a mean value of 60 degrees. The standard deviation in the mean value of the angle is 30 seconds. Present a 95% confidence interval for μ_θ and σ_θ^2 .

18. From two sets of measurements supposedly from the same population, the following data were recorded:

SET	n	\bar{X}	σ_x
I	8	10.1	0.10
II	12	10.0	0.07

At the 10% level of significance, test the hypothesis that the variances are equal against the alternative that they are not equal.

19. If $x_1 = [1 \ 2] \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $x_2 = [2 \ 3] \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, and $\Sigma_{uu} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ is the covariance matrix of u_1 and u_2 , compute the covariance matrix Σ_{zz} for

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

20. Given the cartesian coordinates $x = 300$ mm with $\sigma_x = 0.04$ mm, and $y = 400$ mm with $\sigma_y = 0.05$ mm, where x and y are assumed uncorrelated, compute the polar coordinates r and θ and their covariance matrix.

21. $x_1 = [1 \ 3] \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $x_2 = [1 \ 4] \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $x_1^2 + x_2^2 = y^2$, and $\Sigma_{uu} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix}$.

Compute σ_y if $u_1 = 0$ and $u_2 = 1$.

22. $x = [2 \ 3]^t$, $\Sigma_{xx} = \begin{bmatrix} 1 & -0.1 \\ -0.1 & 2 \end{bmatrix}$, $y = x^t x$. Compute σ_y^2 .

23. $y = x_1 \cos \theta - x_2 \sin \theta$, $\theta = 45^\circ$, $\mathbf{W}_{xx} = \begin{bmatrix} 5 & 1 \\ 1 & 4 \end{bmatrix}$. Compute the weight of y if $\sigma_0^2 = 1$.

24. $z = [1 \ 2]^t$, $\Sigma_{zz} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$, $u = 2z^t A z$ for $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$. Compute σ_u^2 .

25. The figure below shows a small triangulation chain, in which L is the measured baseline and $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2$, and β_3 are known angles. The exit base M may be computed from

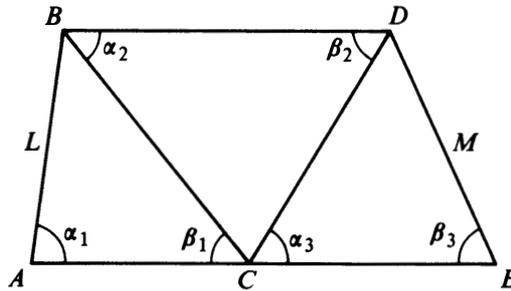
$$M = L \frac{\sin \alpha_1 \sin \alpha_2 \sin \alpha_3}{\sin \beta_1 \sin \beta_2 \sin \beta_3}$$

Compute the value of M and its standard deviation if the measured quantities and their corresponding estimated accuracies are

$$L = 5286.278 \text{ meters; } \sigma_L = 12.7 \text{ mm}$$

ANGLE	DEGREES	MINUTES	SECONDS	σ IN SECONDS
α_1	53	56	41.3	7.6
β_1	74	47	40.2	10.5
α_2	50	42	26.3	6.8
β_2	81	32	56.3	2.3
α_3	63	50	46.6	3.7
β_3	69	22	06.5	5.3

Note: All observations are assumed to be uncorrelated.



26. The following two functions x_1 and x_2 relate three observed correlated variates u_1 , u_2 , and u_3 , where $u_1 = 3.0$, $u_2 = 4.0$, and $u_3 = 5.0$,

$$x_1 = u_1^2 + 3u_2^3 + u_3$$

$$x_2 = \frac{1}{u_1^2} + 4u_2 + \frac{1}{u_3}$$

Compute the covariance matrix of the functions x_1 and x_2 given that the weight matrix of the observations is

$$\mathbf{W}_{uu} = \begin{bmatrix} 2 & 0.5 & 0.6 \\ 0.5 & 1 & -0.7 \\ 0.6 & -0.7 & 3 \end{bmatrix}$$

and the reference variance is 0.8.

27. It is estimated that in 1 km of differential leveling, the standard deviation in elevation difference determination is 2 mm. The elevation of bench mark A is considered known at a standard deviation of 3 mm. How long a level line can be run from BM A to establish the elevation of point Q at a standard deviation of 1 cm?

28. It is estimated that each interior angle of a nine-sided closed traverse is measured with a standard deviation of 10 seconds. Within what range would you expect the sum of the measured interior angles of the traverse to fall, knowing that the sum of the interior angles of an n -sided polygon is $(n - 2) 180$ degrees?

29. The upper base (UB), lower base (LB), and height (H) of a trapezoid are each measured n times and the respective standard deviation of a single measurement for each is determined (σ_{UB} , σ_{LB} , and σ_H). The mean value of the area of the trapezoid is determined from $\bar{A} = [(UB + LB)/2]H$. What is the standard deviation of \bar{A} in terms of the variables identified? (Assume the measurements on UB , LB , and H are independent.)

30. σ_A , σ_B , σ_C , and σ_D are the standard deviations of A , B , C , and D , respectively. The measurements on A , B , C , and D are not independent, hence $\sigma_{AB} \neq \sigma_{AC} \neq \sigma_{AD} \neq \sigma_{BC} \neq \dots \neq 0$. If $S = A + 2B - 3C + D$, what is the standard deviation of S ?

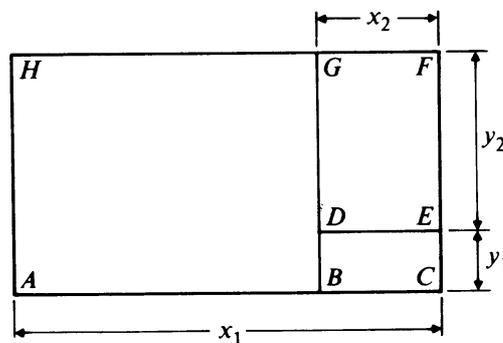
31. In a triangle, the misclosure $w = a + b + c - (180^\circ - e)$, where a , b , and c are the three measured angles with standard deviations σ_a , σ_b , and σ_c , respectively, and e is the spherical excess. (a) Determine the standard deviation of w , σ_w . (In terms of variables defined above.) (b) Determine σ_w if $\sigma_a = \sigma_b = \sigma_c = \sigma_A$. (c) If each angle is

determined as the difference of two directions as read from a horizontal circle of a theodolite, and if the standard deviation for a direction σ_D is assumed equal for all directions, represent σ_a , σ_b , σ_c , and σ_w in terms of σ_D .

32. In stadia distance measurement, the horizontal distance $D = KR \cos^2 \alpha$, where K is the stadia constant (normally 100), R is the level rod intercept (upper wire reading – lower wire reading), and α is the vertical inclination of the telescope above horizontal. At a sight distance of 300 feet, the wire readings on the level rod can be made with a standard deviation of approximately ± 0.01 feet. (a) Determine the standard deviation in D for a horizontal sight distance of 300 feet. (b) State or show why the effects of error in K and α are much less than those in the reading of the level rod.

33. In a subtense bar distance measurement process, the horizontal distance $D = (b/2) \cot(\alpha/2)$, where b is the length of the subtense bar, α is the horizontal angle measured between the targets designating the length of the subtense bar, and the line of sight is perfectly perpendicular to the subtense bar at its midpoint. If the length of the subtense bar is 2 meters ± 0.1 mm, determine the standard deviation that must be attained in the measurement of the angle α in order to measure a 100-foot line with a relative error of 1/5000. Repeat for a 200-foot line and a 300-foot line. (Note: For the purposes of this problem consider the relative error to be the ratio of the estimated *standard deviation* in the line length over the *line length*.)

34. The figure shows a scaled layout of a land area where all the angles in the figure are perfectly known to be right angles. You are given the following information:



$$\begin{aligned} x_1 &= 7 \text{ cm} \\ x_2 &= 2 \text{ cm} \end{aligned} \quad \Sigma_{xx} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \text{ mm}$$

$$\begin{aligned} y_1 &= 1 \text{ cm} \\ y_2 &= 3 \text{ cm} \end{aligned} \quad \Sigma_{yy} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \text{ mm}$$

Assuming no correlation between the x_i and y_i measurements, compute the area of the rectangle $ABGH$ and its variance.

35. An observable quantity l has been measured three times with assumed uncorrelated values l_1 , l_2 , and l_3 having standard deviation σ_1 , σ_2 , and σ_3 , respectively. Using the *propagation* principle derive the *standard deviation* of the weighted mean of

these three observations. (Assume that the reference variance is unity.) The weighted mean \bar{x} , of any variable x , is given by

$$\bar{x} = \frac{(w_1 x_1 + w_2 x_2 + \cdots w_n x_n)}{w_1 + w_2 + \cdots w_n}$$

where x_i is the i th observation and w_i is the i th weight, *assuming no correlation*.

36. a. Sketch the error ellipse for

$$\Sigma_{xx} = \begin{bmatrix} \frac{5}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{5}{2} \end{bmatrix}$$

b. Describe in your own words what the above ellipse means.

