INTRODUCTION TO ADJUSTMENT CALCULUS

P. VANICEK

September 1973

LECTURE NOTES
35

INTRODUCTION TO ADJUSTMENT CALCULUS

(Third Corrected Edition)

Petr Vaníček

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

February, 1980
Latest Reprinting October 1995

PREFACE

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

FOREWORD

It has long been the author's conviction that most of the existing courses tend to slide over the fundamentals and treat the adjustment purely as a technique without giving the student a deeper insight without answering a good many questions beginning with "why". This course is a result of a humble attempt to present the adjustment as a discipline with its own rights, with a firm basis and internal structure; simply as an adjustment calculus. Evidently, when one tries to take an unconventional approach, one is only too liable to make mistakes. It is hoped that the student will hence display some patience and understanding.

These notes have evolved from the first rushed edition - termed as preliminary - of the Introduction to Adjustment Calculus, written for course SE 3101 in 1971. Many people have kindly communicated their comments and remarks to the author. To all these, the author is heavily indebted. In particular, Dr. L. Hradílek, Professor at the Charles University in Prague, and Dr. B. Lund, Assistant Professor at the Mathematics Dept. UNB, made very extensive reviews that helped in clarifying many points. Mr. M. Nassar, a Ph.D. student in this department, carried most of the burden connected with rewriting the notes on his shoulders. Many of the improvements in formulations as well as most of the examples and exercise problems contained herein originated from him.

None of the contributors should however, be held responsible for any errors and misconception still present. Any comment or critism communicated to the author will be highly appreciated.

P. Vaníček
October 7, 1974

CONTENTS

4. Fundamentals of the Theory of Errors

5. Least-Squares Principle

6. Fundamentals of Adjustment Calculus

INTRODUCTION

In technical practice, as well as in all experimental sciences, one is faced with the following problem: evaluate quantitatively parameters describing properties, features, relations or behaviour of various objects around us. The parameters can be usually evaluated only on the basis of the results of some measurements or observations. We may, for example, be faced with the problem of evaluating the length of a string. This can be measured directly. Here the only parameter we are trying to determine is the observed quantity itself and the problem is fairly simple. More complicated proposal would be, for instance, to determine the coefficient of expansion of a rod. Then the parameter--the coefficient of expansion--cannot be measured directly, as in the previous case, and we have to deduce its value from the results of observations of length, by performing some computations using the mathematical relationship connecting the observed quantities and the wanted parameters. The more complicated the problems get, of course, the more complex is the system whose parameters we are trying to determine. Obviously, the determination of the orbital parameters of a satellite from various angles observed on the surface of the earth would be an example of one such still more sophisticated task.

The adjustment is a discipline that tries to categorise those problems and attempts to deal with them symmetrically. In order to be able to deal with such problems systematically the adjustment has to use a language suitable for this purpose, the obvious choice being mathematics.

Hence, the problem to be treated has to be first "translated" into the language of mathematics, i.e., the problem has to be first mathematically formulated. The mathematical formulation of the problem would really be the mathematical formulation of the relation between the observed quantities (observables) and the wanted quantities (parameters). This relationship is called the mathematical model. Denoting the observables by L (L stands for one, two, or n quantities) and the parameters by X (X stands for one, two or m quantities) the most general form of the mathematical model cna be written as

$$F(X, L) = 0 .$$

The above equation merely states that there is a (implicit) relation between the observables and the parameters. The formulation of an actual mathematical model has to be done taking into account all the physical and geometrical laws--simply using the accumulated experience. The complexity of the mathematical model reflects the complexity of the problem itself. Thus the mathematical model of our first problem is practically trivial:

$$X = L$$

where X is the wanted length and L is the observed length.

The mathematical model for the coefficient of expansion α of the rod is more complicated, namely, for instance

$$\ell = \ell_o (1 + \alpha t)$$

where α = X, the observed length ℓ and the observed difference in temperature t create L and ℓ_o is another parameter (length of the rod at

a fixed temperature) which we happen to know. The mathematical model for the satellite orbital elements would be more complicated still.

Once the mathematical model has been formulated it can become a subject of rigorous mathematical treatment, a subject of adjustment calculus. Hence, the formulation of the mathematical model itself is to be considered as being beyond the scope of adjustment calculus and only the various kinds of mathematical models alone constitute the subject of interest.

There is one particular class of models, that are very often encountered in practice, and that can be termed as overdetermined. By an overdetermined model we understand a model which does not have a unique solution for X because there are "unnecessarily many" observations supplied. This can be the case, say, with our first example, if the length is measured several times. The model in this case would be formulated as

$$\begin{aligned} X &= \ell_1 \\ X &= \ell_2 \\ &\vdots \\ X &= \ell_n \ , \end{aligned}$$

where $\ell_1, \ell_2, \ldots, \ell_n$ are all encompassed by the symbol L. Or, in the second example, we may have

$$\begin{aligned} \ell_1 &= \ell_o \ (1 + \alpha t_1) \\ \ell_2 &= \ell_o \ (1 + \alpha t_2) \\ &\vdots \\ \ell_n &= \ell_o \ (1 + \alpha t_n) \ , \end{aligned}$$

where $(\ell_1, \ell_2, \ldots, \ell_n, t_1, t_2, \ldots, t_n) = L$.

As we can easily see, these overdetermined models may or may not have a unique solution. They usually have not. Therefore, in order to produce a unique solution of some kine, we have to assume that the observations were not quite correct, that there were errors in their determinations.

This leads us into the area of the theory of errors with its prerequisites--the theory of probability and statistics. With the help of these disciplines we are aboe to define the most probable unique solution (if it exists) for the parameters of the mathematical models. We also are usually able to establish the degree of reliability of the solution.

The notes are divided into six sections: Fundamentals of the intuitive Theory of Sets, Fundamentals of the Mathematical Theory of Probability, Fundamentals of Statistics, Fundamentals of the Theory of Errors, Least-Squares Principle, Fundamentals of the Adjustment Calculus. The first four sections describe the relevant parts of the individual fields that are necessary to understand what adjustment is all about. They, by no means, claim any completeness and it is envisaged that an interested student will supplement his reading from other sources, such as those listed at the end of these notes.

A separate section (5) is devoted to the philosophical basis of the adjustment calculus. Although not very extensive it should be regarded important, giving the reasons why the least-squares technique is used in adjustment.

Finally, the last section deals with the basics of the adjustment proper. Here again, only the introductory parts of the adjustment calculus could be treated with the understanding that only the subsequent courses will develop the full picture.

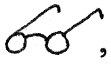
Throughout the course emphasis is placed on the parallel development of concepts of "discrete statistics", i.e. statistics of random samples, and "continuous statistics", i.e. statistics of random variables. While random samples are the quantities we deal with in every-day practice, the mathematical tools used are predominently from the continuous domain. Good understanding of the interplay of the two concepts is indispensable for anyone who wants to be able to use the adjustment calculus properly.

The bibliography given at the end of these notes lists some of the useful books dealing with statistics and adjustments. Interested reader is recommended to complement the reading of these notes by turning to at least some of the listed sources.

1. FUNDAMENTALS OF THE INTUITIVE THEORY OF SETS

1.1. Sets, Elements and Subsets

A set is an ensemble of objects (elements) that can be distinguished one from another. The set is defined when all its elements are defined.

Example 1.1: $A_1 \equiv \{$  $\}$,

$A_2 \equiv \{ 1,\ 8,\ 15,$  $,\ 0,$  $,\ 4\}$,

$A_3 \equiv \{ 0,\ 1\}$,

$A_4 \equiv$ {all the left feet} ,

$A_5 \equiv$ {all the cities with more than one million inhabitants in New Brunswick} ,

$R \equiv$ {all the real numbers} , and

$I \equiv$ {all the positive integers} , are all sets.

The text within the brackets {...} is known as the list of the set. If an element a exists in the list of a set A, we say that the element a belongs to the set A, and this is denoted by

$$a \in A \ ,$$

which is read as "a belongs to A". On the other hand, if an element a does not belong to a set A, we write

$$a \notin A \ ,$$

which is read as "a does not belong to A".

Example 1.2: Referring to Example 1.1, we see that:

 $\in A_1$, $8 \in A_2$, $2 \notin A_1$, and a right foot $\notin A_4$.

A part of a set G is called a subset of G whether it contains one or several elements. The fact that a set H is contained in G is hence written as

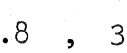
$$H \subset G \quad .$$

If H is not contained in G, i.e. if not all the elements of H are at the same time elements of G, we write

$$H \not\subset G \quad .$$

Example 1.3: Referring to Example 1.1, we see that:

$$A_3 \subset A_2 \quad , \quad \{2 \ , \ 35 \ , \ 118\} \subset I \ ,$$

$\{1.8 \ , \ 3 \ , \ 6.2\} \subset R$, and $\{$  $\}$ $\not\subset A_1$.

A set which does not contain any element is known as the empty (void or null) set, and is denoted by $\emptyset$.

Example 1.4: The set $A_6 \equiv$ {all people taller than 10 feet} contains no elements, i.e. $A_6 \equiv \emptyset$. Also from Example 1.1, we find that $A_5 \equiv \emptyset$.

The sets are called equal if they contain the same and none but the same elements.

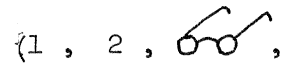
Example 1.5: All the following sets are equal

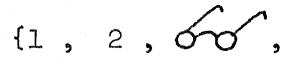
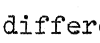
$$\{1 \ , \ 2 \ , \ 3\} \ , \{3 \ , \ 1 \ , \ 2\} \ , \{2 \ , \ 3 \ , \ 1\} \ , \ \ldots$$

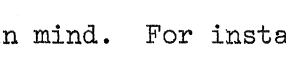
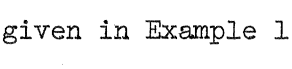
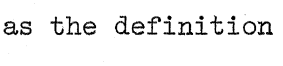
1.2. Progression and Definition Set

A progression ξ is an ordered (by ordered we mean that ξ is arranged such that each of its elements has a specific position) ensemble of objects (elements) that may not all be distinguishable one from another. The definition set D of a progression ξ is the set composed from all the distinguishable elements of ξ. In such a case, we shall also say that D

belongs to ξ.

Example 1.6: $\xi \equiv$ (1 , 2 ,  , 2 , 1 , 8 , ) , is a progression, and its definition set D is given by

$D \equiv$ {1 , 2 ,  , 8 ,  } .

At this point, the difference between a progression and a set should be clear in mind. For instance, the progression ( , 8 , 2 , 1 , 1 , 2 , ) represents a different progression than the one given in Example 1.6. However, the sets {  , 8 ,  , 2 , 1} , {2 , 1 , 8 ,  , }, ... are all the same as the definition set D in Example 1.6.

1.3. Cartesian Product of Sets

The Cartesian product of two sets A and B is a set, called the product set and denoted by A x B (reads A cross B), whose elements are all the ordered two-tuples of elements of the component sets A and B. Hence, if $a \in A$ and $b \in B$, then the two-tuple $(a,b) \in A \times B$. However, if $b \notin A$ or $a \notin B$, the two tuple $(b,a) \notin A \times B$.

The above definition can be extended to more than two sets, say n sets. In such a case, the elements of the product set will be all the ordered n-tuples of elements of the component sets. Accordingly, we can define the Cartesian n-power A^n (or A_n if no danger of confusion with indexed set exists) of a set A as the Cartesian product of the same set A with itself n-times.

Example 1.7: If A ≡ {3 , 1 , 5} and B ≡ {2 , 4} , then the product set A x B is A x B ≡ {(3 , 2), (3 , 4), (1 , 2), (1 , 4), (5 , 2), (5 , 4)} . Referring to Example 1.1, we can easily see that:

( , ) $\in A_1 \times A_2$, (4.18 , ) $\in A_1 \times A_2$,

$(1\ ,\ \heartsuit\ ) \notin A_1 \times A_2$ but $(\ 1\ ,\ \heartsuit\ ) \in A_2 \times A_1$,

$(1\ ,\ 2\ ,\ 15\ ,\ 1\ ,\ 8) \in I^5$, and $(5.16,\ 3.26,\ 1,\ 0,\ 1) \in R^5$.

1.4. Intersection of Sets

The intersection of two sets A and B, denoted by $A \cap B$, is a set which is a subset of both A and B and does not contain any elements other than the common elements to A and B. The intersection of two sets can be represented by the shaded area in Figure 1.1. Diagrams of this kind are called "Venn diagrams".

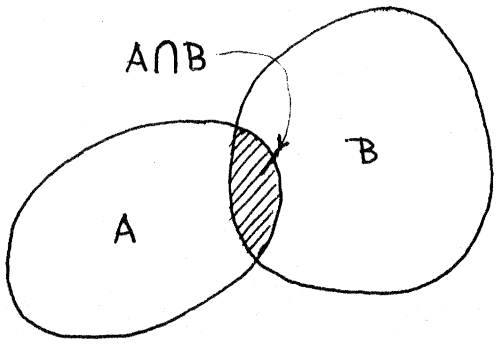


Figure 1.1

From the above diagram we can easily see that

$$A \cap B \equiv B \cap A \ .$$

Example 1.8: Referring to Example 1.1, we find that:

$A_1 \cap A_2 \equiv \{$ ☼ $\}$, $R \cap I \equiv I$,

$A_2 \cap I \equiv \{1, 4, 8, 15\}$, and $A_3 \cap A_2 \equiv \{0, 1\} \equiv A_3$.

Note that we can define a subset A of B as such a set whose intersection with B is A itself. In other words, if $A \subset B$ then $A \cap B \equiv A$, or vice versa (see Figure 1.2).

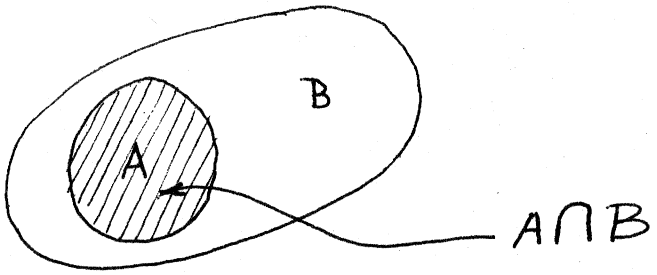


Figure 1.2

If $A \cap B \equiv \emptyset$, then the sets A and B are said to be disjoint sets.

The intersection of n sets $A_1, A_2, \ldots, A_n$ is usually denoted by

$$\bigcap_{i=1}^{n} A_i \text{ , where}$$

$$\bigcap_{i=1}^{n} A_i \equiv A_1 \cap A_2 \cap A_3 \ldots \cap A_n \ .$$

This is illustrated in Figure 1.3 by the common area to all sets.

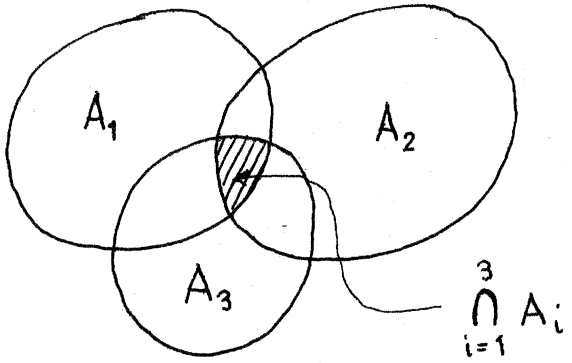


Figure 1.3

1.5. Union of Sets

The union of two sets A and B, denoted by $A \cup B$, is a set that contains all the elements of A and B and none else. Similar to the intersection, the union of the two sets is represented by the shaded area in

Figure 1.4

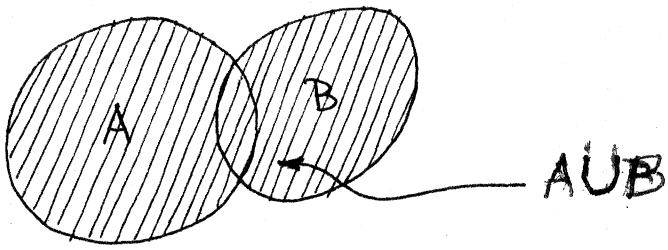


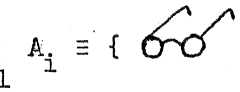
Figure 1.4

The union of n sets $A_1, A_2, \ldots, A_n$ is denoted by

$$\bigcup_{i=1}^{n} A_i \text{ , where}$$

$$\bigcup_{i=1}^{n} A_i \equiv A_1 \cup A_2 \cup A_3 \ldots \cup A_n \quad .$$

Example 1.9: Referring to Example 1.1, we obtain

$$\bigcup_{i=1}^{3} A_i \equiv \{ \text{  } , 4.18, 1, 8, 15, \text{  } ,$$

$$0, 4 \} ,$$

and $I \cup R \equiv R$.

Thinking of the union as the addition of sets, the subtraction of two sets is known as the complement of one into the other. Referring to Figure 1.5, and considering the two sets $A \subset B$, the set of all the elements contained in B and not contained in A is called the complement of A in B, and is denoted by B - A.

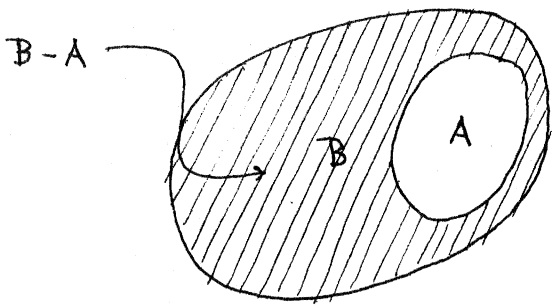


Figure 1.5

Example 1.10: Referring to Example 1.1, we get:

The complement of A_3 in A_2 is

$A_2 - A_3 \equiv \{8, 15,$ ☾ , ☼ , $4\}$, and

R-I ≡ { all real numbers that are not integers }.

1.6. Mapping of Sets

f is called a mapping of A into B if it relates one and only one element from B to each element of A. This means that for each element a ε A there will be only one corresponding image b ε B (see Figure 1.6).

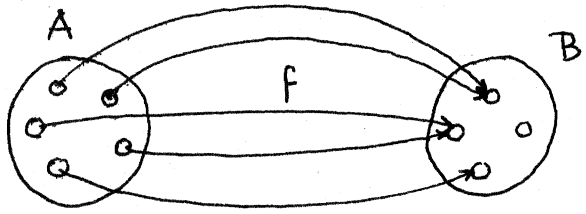


Figure 1.6

Note here that the one-to-one relationship (i.e each b ε B has got one and only one argument a ε A) is not required. We shall denote any such mapping by

$$f \varepsilon \{A \rightarrow B\} ,$$

and read it as "f is an element of the set of all the mappings of A into B",

or simply "f is a mapping of A into B", or "f maps A into B".

If the elements of B are all images of the elements of A, then f is called an onto mapping, or simply we say that "f maps A onto B".

If A and B are numerical sets, then f is called a function (which gives the mathematical relationship between each $a \in A$ and its corresponding image $b \in B$). In this case, the image b of a will be nothing else but the functional value $f(a)$.

Example 1.11: Given the set $A \equiv \{a_1, a_2, a_3\} \equiv \{2, -1, 3\}$ and the mapping $f \in \{A \to B\}$, where $f(a_i) = a_i^3$ for each $a_i \in A$, i=1,2,3, then the images $b_i \in B$ are computed as the functional values of the corresponding elements $a_i \in A$, i.e. $b_i = f(a_i) = a_i^3$, which give $b_1 = (2)^3 = 8$, $b_2 = (-1)^3 = -1$, and $b_3 = (3)^3 = 27$.

Generally, f is an into function, hence we write $(8, -1, 27) \in B$. However, if f is an onto function, then the image set B of this example is given as $B \equiv \{8, -1, 27\}$.

1.7. Exercise 1

1. Which of the following sets are equal?

   $\{t, r, s\}$, $\{s, r, t\}$, $\{r, s, t\}$, $\{t, s, r\}$.

2. Let $A \equiv \{d\}$, $B \equiv \{c, d\}$, $C \equiv \{a, b, c\}$, $D \equiv \{a, b\}$ and $H \equiv \{a, b, d\}$;

   (i) is $B \subset D$? (ii) is $C \equiv B$?

   (iii) is $D \subset C$? (iv) is $B \neq H$?

   (v) is $A \subset H$? (vi) is $(A \cup D) \subset H$?

   (vii) is $(A \cap B) \not\subset C$? (viii) is $(H \cap C) \equiv D$?

3. Let $U \equiv \{1, 2, 3, \ldots, 8, 9\}$, $A \equiv \{1, 2, 3, 4\}$, $B \equiv \{2, 4, 6, 8\}$, $C \equiv \{3, 4, 5, 6\}$ and $D \equiv \{1, 3, 5, 7, 9\}$; then find the following:

   (i) $B \cup D$; (ii) $A \cap C$;

   (iii) $A \cup B$; (iv) $U - A$;

   (v) a set H, which is a subset of all the sets U, A and D.

4. Considering the following Venn diagram with the sets A, B, C, D and H, indicate by shading the suitable areas on separate diagrams, the following sets:

   (i) $D \cup H$;

   (ii) $H \cap C$;

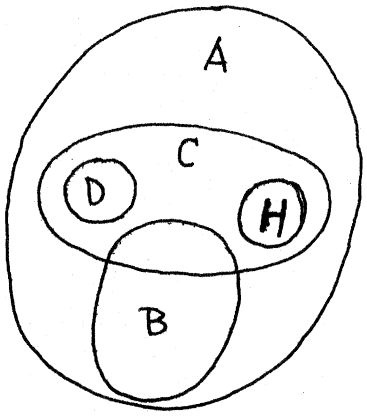
   (iii) $C \cap B$;

   (iv) $A - C$;

   (v) $B \cup C$;

   (vi) $(A - B) \cup (B \cap C)$;

   (viii) $A - (C \cup B)$.



5. Considering the two sets:

   $A \equiv \{3, 4, 0, -1\}$ and $B \equiv \{-2, 5\}$, find the Cartesian products A x B and B x A. Also find the second power B^2 of the set B.

6. Given the set $X \equiv \{-2, -1, 0, 1, 2\}$, with $f \in \{X \to Y\}$. If for each $x \in X$, $f(x) = x^2 + 1$, find the image set Y considering that f is an onto function.

2. FUNDAMENTALS OF THE MATHEMATICAL THEORY OF PROBABILITY

2.1 Probability Space, Probability Function and Probability

Let us have a set $D \neq \emptyset$ and let us assume that it can be partitioned into mutually disjoint subsets $D_j \subset D$ such that $D \equiv \bigcup_j D_j$ (by mutually disjoint subsets we mean such subsets that $D_i \cap D_j \equiv \emptyset$ for any pair D_i, D_j, $i \neq j$). Such a set D we shall call probability space.

Any mapping P of D onto [0, 1] (that is the set of all positive real numbers "b" satisfying the inequalities ($0 \leq b \leq 1$)) that has the following two properties:

(1) If $D' \subset D$, then $P(D') = 1 - P(D - D')$, (note that $D - D'$) is the complement of D' in D; see section 1.5), and

(2) If $D_1, D_2, \ldots, D_n \subset D$ are mutually disjoint, then $P(\bigcup_{i=1}^{n} D_i) = \sum_{i=1}^{n} P(D_i)$, is called a probability function. The value (P(D')) of the probability function P (takes any value from [0, 1]) is called the probability. Note that the difference between the function and the functional value has been mentioned in section 1.6.

The above two properties of the probability function have the following consequences:

(1) $P(D) = 1$,

(2) $P(\emptyset) = 0$,

(3) If $D' \subset D$; then $P(D') \leq 1$,

(4) If $D'' \subset D'$; then $P(D'') \leq P(D')$, and

(5) If $A, B \subset D$, and $A \cap B \equiv \emptyset$; then $P(A \cup B) = P(A) + P(B)$.

If D is a point set, i.e. its elements can be represented by points, it is always decomposable.

The value $\sum_i P(D_i) \in [0, 1]$ is sometimes called the total or accumulative probability of $\bigcup_i D_i$.

2.2. Conditional Probability

If A, $B \subset D$; then the ratio $P(A \cap B)/P(B) = P(A/B)$ is called the conditional probability. The right hand side, that is P(A/B), is read as "probability of A given B". In other words, the conditional probability P(A/B) can be interpreted as the probability of occurrence of A under the condition that B occurred.

From the above definition of the conditional probability, we notice that:

(1) If P(B) = 0; then P(A/B) is not defined,

(2) If $B \subset A$; then $A \cap B = B$ (see section 1.4), and then P(A/B) = 1,

(3) If $A \cap B \equiv \emptyset$, i.e. A and B are disjoint sets ; then P(A/B) = 0.

2.3. Combined Probability

If the conditional probability P(A/B) equals to P(A), then it is clear that the occurrence of A does not depend on the occurrence of B. In such a case we say that A and B are independent. Using the definition of the conditional probability from the previous section, we can write:

$$P(A \cap B) = P(A) \cdot P(B) \quad .$$

This can be understood as the probability of simultaneous occurrence of A and B, which is usually denoted by P(A, B) and read as probability of A

and B, and known as the combined (compound) probability of A and B, that is

$$P(A, B) = P(A) \cdot P(B).$$

Similarly, we define the combined probability of occurrence of the independent $D_1, D_2, \ldots, D_n \subset D$ as the product of their individual probabilities, i.e.

$$P(D_i, D_j) = P(D_i)\ P(D_j) \quad i \neq j$$

$$P(D_i, D_j, D_k) = P(D_i)\ P(D_j)\ P(D_k) \quad i \neq j,\ j \neq k,\ i \neq k,$$

$$\vdots$$

$$P(D_1, D_2, \ldots, D_n) = \prod_{i=1}^{n} P(D_i) .$$

Example 2.1: Suppose we have decomposed the probability space D into seven mutually disjoint subsets $D_1, D_2, \ldots, D_7$ as shown in Figure 2.1 such that:

$$D = \bigcup_{i=1}^{7} D_i .$$

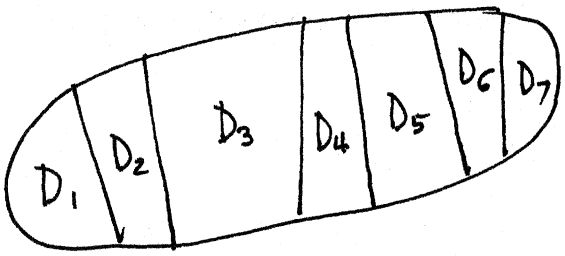


Figure 2.1.

Assuming that the probabilities $P(D_i)$ of the individual subsets D_i are found to be:

$P(D_1) = 1/28$, $P(D_2) = 2/28$, $P(D_3) = 3/28$, $P(D_4) = 4/28$, $P(D_5) = 5/28$, $P(D_6) = 6/28$, and $P(D_7) = 7/28$; then we get:

Total probability of D_i, $i = 1, 2, \ldots, 7$ is

$$P(D) = P\left(\bigcup_i D_i\right) = \sum_{i=1}^{7} P(D_i) = (1+2+3+4+5+6+7)/28 = 28/28 = \underline{1.0.}$$

Combined probability of all $D_i = \prod_{i=1}^{7} P(D_i) = 0$

Example 2.2: In this example we assume that our probability space D is decomposed into five elements $d_j \in D$, $j = 1, 2, \ldots, 5$. If the probabilities $P(D_j)$, as represented by the ordinates in Figure 2.2, are given as:

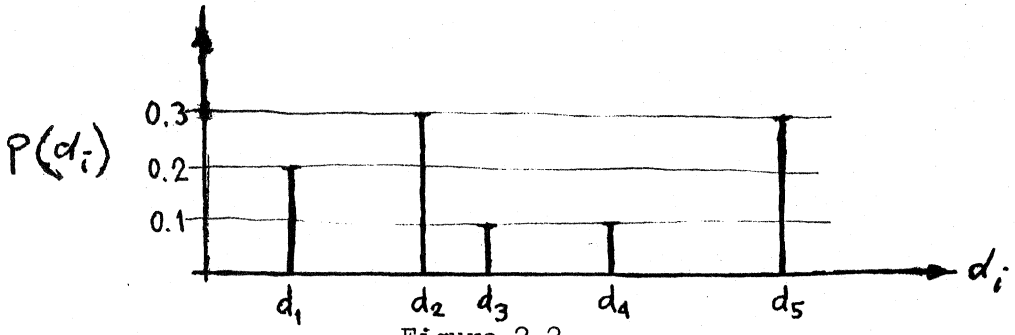


Figure 2.2

$P(d_1) = 0.2$, $P(d_2) = 0.3$, $P(d_3) = 0.1$, $P(d_4) = 0.1$, and $P(d_5) = 0.3$; then we get:

$$\text{Total probability } P(D) = P(\bigcup_j d_j) = \sum_{j=1}^{5} P(d_j) = 0.2+0.3+0.1+0.1+0.3$$

$$= \underline{1.0} .$$

Combined probability of d_1 and d_2 (for example) $= P(d_1, d_2)$

$$= \prod_{j=1}^{2} P(d_j) = 0.2.0.3 = \underline{0.06}.$$

This combined probability has to be understood as the probability of simultaneous occurrence of d_1 and d_2 under the assumption of their independence.

2.4. Exercise 2.

We have determined that every number of a die have the probability of appearing when the die is tossed, proportional to the number itself.

Let us denote: A = {even numbers}, B = {prime numbers}, and C = {odd numbers}; all subsets of the set of numbers appearing on the die.

Required: (1) Construct the probability space D.

(2) Find the probability of each individual element $d_i \in D$.

(3) Find P(A), P(B) and P(C).

(4) Find the probability that:

(i) an even or prime number occurs,

(ii) an odd prime number occurs,

(iii) A but not B occurs.

III. FUNDAMENTALS OF STATISTICS

3.1 Statistics of an Actual Sample

3.1.1 Definition of a Random Sample

Any finite (i.e. containing only a finite number n of elements) ordered progression of elements (see section 1.2) $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ such that:

(i) its definition set D (see section 1.2) can be declared a probability space (see section 2.1); and

(ii) it has the probability function P defined for every $d_i \in D$ in such a way that $P(d_i) = c_i/n$, where c_i is the count (frequency), of the element d_i in ξ.

may be called a random sample. The ratio c_i/n is known as the relative frequency.

Example 3.1: Consider the following progression ξ

$$\xi_1\ \xi_2\ \ \xi_3\ \ \xi_4\ \xi_5\ \xi_6\ \ \xi_7$$
$$\xi = \{\ 1,\ \heartsuit\ ,\ \text{}\ ,\ 1,\ 1,\ \text{}\ ,\ \heartsuit\ \}\ ,$$

which has seven elements, (i.e. $n = 7$).

The definition set D of ξ will be

$$d_1\ d_2\ \ d_3\ \ d_4$$
$$D = \{1,\ \heartsuit\ ,\ \text{  }\ ,\ \text{}\ \}\ ,$$ which

consists of four elements (i.e. $m = 4$),

the counts of which are:

$c_1 = 3$, $c_2 = 2$, $c_3 = 1$ and $c_4 = 1$, and their

corresponding probabilities (relative frequencies) are:

$P(d_1) = P(1) = 3/7$, $P(d_2) = P(\heartsuit) = 2/7$,

$P(d_3) = P(\clubsuit\clubsuit) = 1/7$, and

$P(d_4) = P(\spadesuit) = 1/7$.

Note here that really both properties required from P to be a probability function (section 2.1) are satisfied. In particular we have (from the above example): the total probability

$$P(D) = P\left(\bigcup_{i=1}^{m} d_i\right)$$

$$= \sum_{i=1}^{4} P(d_i) = \frac{3}{7} + \frac{2}{7} + \frac{1}{7} + \frac{1}{7} = \frac{7}{7} = 1 \ .$$

Accordingly, any finite ordered progression of elements may be declared a random sample. This is a very important discovery and has to be born in mind throughout the following development. As a result, it is always possible to construct the probability space and the associated probabilities "belonging" to the sample (i.e. the probability associated with each element in the definition set of the sample).

From now on we shall deal with $D \subset R$ (recall that R is the set of all real numbers), i.e. with numerical sets and progressions only. Also, D will be considered ordered in either ascending or descending sense; usually the former is used.

It has to be noted here that our definition of a random sample is not standard in the sense that it admits much larger family of objects to be called random samples than the standard definition. More will be said about it in 3.2.4.

Example 3.2: A die is tossed 100 times. The following table lists the six numbers and the frequency (count) with which each number appeared:

number d_i	1	2	3	4	5	6
count c_i	14	17	20	18	15	16

.

Find the probability that:

(i) a 3 appears ;

(ii) a 5 appears ;

(iii) an even number appears ;

(iv) a prime number appears.

Solution:

(i) $P(3) = \frac{20}{100} = 0.20$,

(ii) $P(5) = \frac{15}{100} = 0.15$,

(iii) $P(2,4,6) = P(2) + P(4) + P(6)$

$= \frac{17}{100} + \frac{18}{100} + \frac{16}{100} = \frac{51}{100} = 0.51$,

(iv) $P(2,3,5) = P(2) + P(3) + P(5)$

$= \frac{17}{100} + \frac{20}{100} + \frac{15}{100} = \frac{52}{100} = 0.52$.

3.1.2 Actual (Experimental) Probability Distribution Function (PDF) and Cumulative Distribution Function (CDF)

If the random sample ξ is a progression of numbers only (and, of course, its definition set D is a numerical set), which we shall from now on always assume, then P is a discrete function mapping D into [0,1].

This function is usually called experimental (actual) probability distribution function (or experimental frequency function, etc.) of the sample ξ, and abbreviated by PDF. The values $P(d_i)$, $d_i \in D$, are known as experimental probabilities of d_i, which are equal to the corresponding relative frequencies.

Example 3.3: Assume that a certain experiment gave us the following random sample:

$\xi \equiv (1, 2, 4, 1, 1, 2, 1, 1, 2)$, $n = 9$.

Then its definition set is:

$D \equiv \{1, 2, 4\} = \{d_i, i=1,2,3\}$, $m = 3$.

Therefore, the frequencies c_i of d_i are:

$c_1 = 5$, $c_2 = 3$ and $c_3 = 1$.

The corresponding experimental probabilities are: $P(1) = 5/9$, $P(2) = 3/9$ and $P(4) = 1/9$.

As a check $\sum_{i=1}^{3} P(d_i) = \frac{1}{9}(5+3+1) = 1$.

The discrete PDF of the given ξ in this example is depicted in Figure 3.1 (which is sometimes called a bar diagram), in which the abscissas represent d_i and the ordinates represent the corresponding $P(d_i)$.

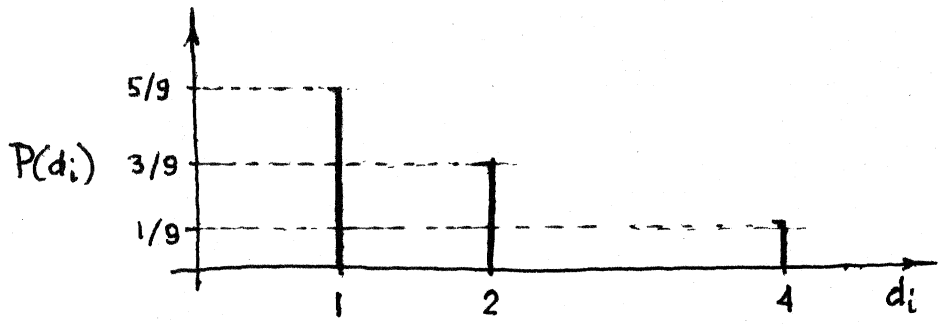


Figure 3.1

Since we are using numerical sets only (and therefore ordered), it makes sense to ask, for instance, what is the actual probability of d being within an interval $D' \subseteq D$, where $D' \equiv [d_k, d_j]$. Such probability is denoted by $P(D')$ or $P(d_k \leq d \leq d_j)$. To answer this question, we use the actual PDF and get

$$P(d_k \leq d \leq d_j) = \sum_{i=k}^{j} P(d_i). \qquad (3.1)$$

The above expression (equation 3.1) must be understood as giving the actual probability of $d \varepsilon D' \equiv \{d_k, \ldots, d_j\} \subseteq D$ rather than $d \varepsilon [d_k, d_j]$ (i.e. the probability that d will acquire a specific discrete value equal to $d_k, d_{k+1}, \ldots, d_{j-1}, d_j$ rather than the probability that d will be anywhere in the continuous interval $[d_k, d_j]$). This is not always properly understood in practice.

The function C of $d_i \in D$ given by

$$\boxed{C(d_i) = \sum_{j \leq i} P(d_j) \; \varepsilon[0,1]} \qquad (3.2)$$

is called <u>experimental (actual) cumulative distribution function</u> (or summation density function, ... etc.) of the sample ξ, and usually abbreviated by <u>CDF</u>.

<u>Example 3.4</u>: Using the data and results from example 3.3, we can compute the CDF of the given sample ξ by computing each $C(d_i)$ as follows:

$C(d_1) = P(d_1) = 5/9,$

$C(d_2) = P(d_1) + P(d_2) = 5/9 + 3/9 = 8/9,$

and $C(d_3) = (P(d_1) + P(d_2)) + P(d_3)$

$= C(d_2) + P(d_3) = 8/9 + 1/9 = 9/9 = 1.$

Figure 3.2 illustrates the discrete CDF of the given sample ξ.

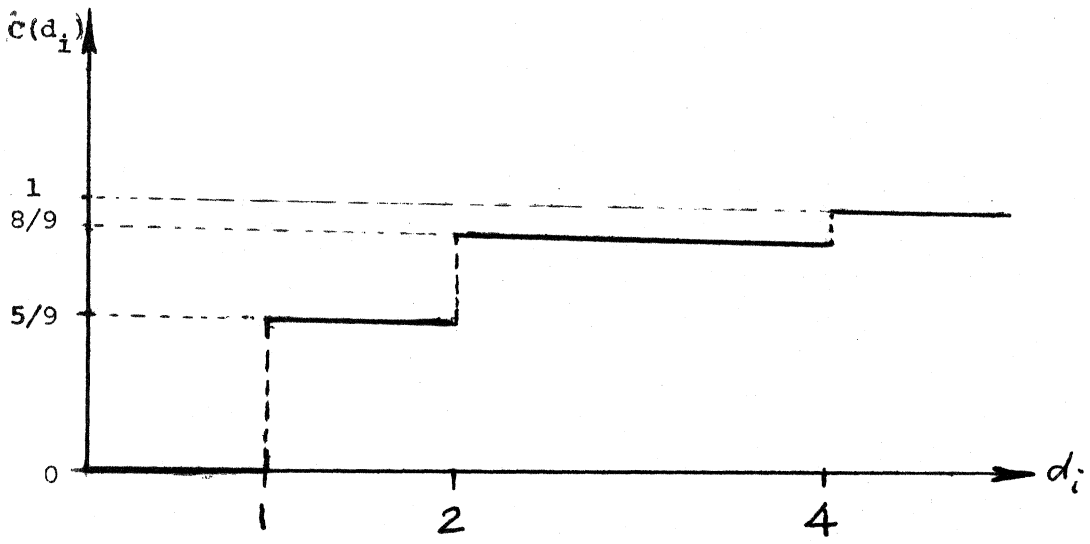


Figure 3.2

From Figure 3.2, we notice the following properties of the CDF:

(i) the value (ordinate) of the CDF is always positive,

(ii) the CDF is a never decreasing function,

(iii) the cumulative probability $C(d_m)$, where d_m is the largest $d_i \in D$, is always equal to 1.

Example 3.5: Using the data from example 3.2, we can construct the CDF of the die tossing experiment as follows:

$C(1) = P(1) = 0.14$,

$C(2) = C(1) + P(2) = 0.14 + 0.17 = 0.31$,

$C(3) = C(2) + P(3) = 0.31 + 0.20 = 0.51$,

$C(4) = C(3) + P(4) = 0.51 + 0.18 = 0.69$,

$C(5) = C(4) + P(5) = 0.69 + 0.15 = 0.84$, and

$C(6) = C(5) + P(6) = 0.84 + 0.16 = \underline{1.00}$.

Note again that the maximum value of the CDF is one. The graphical representation of the above CDF can be constructed similar to Figure 3.2.

3.1.3 Mean of a Sample

Consider the sample $\xi \equiv (\xi_1, \xi_2, \ldots, \xi_n)$ with its definition set $D \equiv \{ d_1, d_2, \ldots, d_m \}$. The real number M defined as:

$$\boxed{M = \frac{1}{n} \sum_{i=1}^{n} \xi_i \in [d_1, d_m],} \tag{3.3}$$

is called the mean (average) of the actual sample.

We can show that M equals also to:

$$\boxed{M = \sum_{i=1}^{m} d_i \, P(d_i) .} \tag{3.4}$$

The proof of (3.4) reads as follows:

$$\text{R.H.S.} = \sum_{i=1}^{m} d_i(c_i/n) = \frac{1}{n}\sum_{i=1}^{m} d_i c_i = \frac{1}{n}(\xi_1+\xi_2+ \ldots + \xi_n) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} \xi_i = M .$$

The mean M of a sample can be interpreted as the outcome of applying the summation operator Σ divided by n on ξ, and is often written as:

$$M = E(\xi) = \text{mean } (\xi) = \text{ave } (\xi) = \bar{\xi} , \qquad (3.5)$$

where the symbol E (an abbreviation for the "mathematical Expectation") must be understood as another name for the summation operator Σ operating on $P(d_i)d_i$ (and not on ξi!).

Note that E is a linear operator, and hence it has the following properties (where k is a constant and ξ is a random sample):

(i) $E(k) = k$,

(ii) $E(k\xi) = kE(\xi)$,

(iii) $E(\xi+k) = E(\xi) + k$,

(iv) $E(\sum_j \xi^j) = \sum_j E(\xi^j)$, where ξ^j, $j = 1, 2, \ldots, s$, are s random samples with the same number of elements m in their corresponding definition sets D^j (Do not confuse ξ_j with ξ^j; the former is an element in the latter. In other words, ξ_j is a single element in a sample, but ξ^j is one sample in a class of samples),

(v) If $\xi = (\xi_1)$, then $E(\xi) = \xi_1$,

(vi) $E(E(\xi)) = E(\xi)$.

Example 3.6: Using the random sample ξ given in example 3.3, let us compute its mean from equation (3.3) as follows:

$$M=E(\xi)=\frac{1}{n}\sum_{i=1}^{n}\xi_i=\frac{1}{9}(1+2+4+1+1+2+1+1+2)=\frac{15}{9}=1\frac{2}{3}\quad .$$

Also, we can use equation (3-4), from which we get:

$$M=E(\xi)=\sum_{j=1}^{m}d_jP(d_j)=1.\frac{5}{9}+2\ .\frac{3}{9}+4\ .\frac{1}{9}$$
$$=\frac{1}{9}(5+6+4)=\frac{15}{9}=1\frac{2}{3}\ .$$

Obviously, both formulae (3.3) and (3.4) give identical answers.

It is interesting to note that computing the mean of a sample using equation (3.4) is analogous to computing the centre of balance in mechanics. This can be simply seen by considering the probabilities $P(d_i)$ or the counts c_i as weights, and then taking the Σ moments = 0 about any point, e.g. the origin 0 (see Figure 3.3 which uses the data from example 3.3).

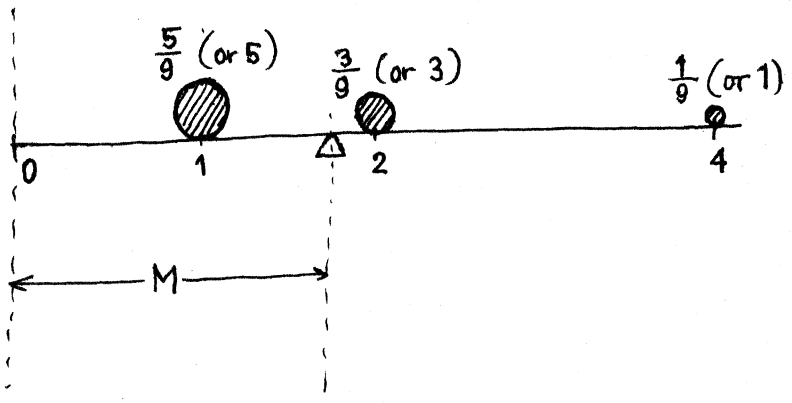


Figure 3.3

The resulting distance of the centre of balance from the point is nothing else but the sample mean M.

It is worthwhile mentioning here that, based on the above analogy with mechanics, the mean M computed from equation (3.4) is also called the weighted mean, in which each element $d_i \in D$ is weighted (the concept of weights is to be discussed later in details) by its probability $P(d_i)$.

3.1.4 Variance of a Sample

Let us have again an actual sample $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ with a mean M. Then, the real number S^2 defined as

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (\xi_i - M)^2, \tag{3.6}$$

is called the variance (dispersion) the actual sample. The square root of the variance S^2, i.e. S, is known as the standard deviation of the sample.

Keeping in mind the relationship between the random sample ξ and its definition set D, we can write:

$$\frac{1}{n} \sum_{i=1}^{n} \xi_i^2 = \sum_{j=1}^{m} d_j^2 \, P(d_j) \ ,$$

which will provide another expression for S^2, namely:

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (\xi_i - M)^2$$

$$= \sum_{j=1}^{m} P(d_j) \, (d_j - M)^2 \ . \tag{3.7}$$

S^2 can be also interpreted as the outcome of the application of the operator E on $(\xi-E(\xi))^2$ meaning really $P(d_j)\ (d_j-M)^2$ and is often written as

$$S^2 = E\ ((\xi-E(\xi))^2 = var\ \ (\xi). \qquad (3.8a)$$

Carrying out the prescribed operation, we get

$$S^2 = E(\xi^2-2\xi E(\xi) + (E(\xi))^2)\ .$$

Applying the calculus with E operator (as summarized in section 3.1.3), we obtain:

$$S^2 = E(\xi^2)-2E(\xi)E(\xi) + E^2(\xi)$$

$$= E(\xi^2) - E^2(\xi)\ .$$

From equation (3.5) we have $E(\xi) = M$, then by substituting for $E(\xi)$ we get

$$S^2 = E(\xi^2) - M^2\ . \qquad (3.8b)$$

Consequently, the corresponding expression to equation (3.7b) will be:

$$S^2 = \sum_{j=1}^{m} d_j^2\ P(d_j)-M^2\ . \qquad (3.9)$$

It is worth mentioning that giving the analogy with mechanics (as discussed in the previous section) we can regard the variance of the sample (equation 3.7)) as the moment of inertia of the system of corresponding mass points with respect to M.

Example 3.7: Let us compute the variance S^2 of the sample ξ given in example 3.3, by using equation (3.8b). First, we compute the first term $E(\xi^2)$ as follows:

$$E(\xi^2)= \frac{1}{n} \sum_{i=1}^{n} \xi_i^2 = \frac{1}{9}(1+4+16+1+1+4+1+1+4) = \frac{33}{9} .$$

Substituting in equation (3.8b), and knowing that $M = \frac{15}{9}$ from example 3.6, we get:

$$S^2 = \text{var}(\xi) = \frac{33}{9} - (\frac{15}{9})^2 = \frac{9 \cdot 33 - (15)^2}{81}$$

$$= \frac{297-225}{81} = \frac{72}{81} = \frac{8}{9} \doteq \underline{0.89} .$$

Taking the square root of the computed variance, we obtain the standard deviation of the sample as:

$$S = \sqrt{\frac{8}{9}} = \frac{2\sqrt{2}}{3} = \frac{2.828}{3} \doteq \underline{0.943} .$$

The same result is obtained if we use equation (3.9), firstly we have

$$\sum_{j=1}^{m} d_j^2 P(d_j) = 1.\frac{5}{9} + 4.\frac{3}{9} + 16.\frac{1}{9}$$

$$= \frac{1}{9}(5+12+16) = \frac{33}{9} ,$$

and since $M = \frac{15}{9}$, we obtain

$$S^2 = \frac{33}{9} - (\frac{15}{9})^2 = \frac{8}{9} \doteq \underline{0.89}.$$

It should be noted here that the same value for the sample variance can be obtained from equations (3.6) and (3.7). The verifica-

tion is left to the student (e.g. using the data from the above example). However, equation (3.9) is advantageous from the computational point of view, especially for large samples. A similar statement holds for computing the sample mean M using equation (3.4).

3.1.5 Other "Characteristics" of a Sample: Median and Range

The median, Med (ξ) of the sample $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ is defined differently for n odd and for n even. For n odd, Med (ξ) equals the ξ that is in the middle of the ordered progression ξ, that is

$$\boxed{\text{Med}\ (\xi) = \xi_{(\frac{n+1}{2})}} \quad . \tag{3.10}$$

For n even, Med (ξ) is the mean of $\xi_{(\frac{n}{2})}$ and $\xi_{(\frac{n}{2}+1)}$ that is:

$$\boxed{\text{Med}(\xi) = \frac{1}{2}\left(\xi_{(\frac{n}{2})} + \xi_{(\frac{n}{2}+1)}\right)} \quad . \tag{3.11}$$

Example 3.8: Consider the sample $\xi \equiv (5,3,6,4,1,2)$.
To obtain Med (ξ), we first arrange the sample in either ascending or descending order, for instance: $\xi \equiv (1,2,3,4,5,6)$, n=6. Since we have n even, we get:

$$\text{Med}(\xi) = \frac{1}{2}\left(\xi_{(\frac{n}{2})} + \xi_{(\frac{n}{2}+1)}\right) = \frac{1}{2}(\xi_3 + \xi_4)$$

$$= \frac{1}{2}(3+4) = \frac{7}{2} = 3.5 \quad .$$

Similarly, the ascending progression of the sample ξ given in example 3.3 is:
$\xi \equiv (1,1,1,1,1,2,2,2,4)$, n = 9.
In this case n is odd, and we get:

$$\text{Med}(\xi) = \xi_{(\frac{n+1}{2})} = \xi_5 = 1 \ .$$

The <u>range</u>, $\text{Ra}(\xi)$ of the sample $\xi \equiv (\xi_i \ , \ i=1,2,\ldots n)$ is defined as the difference between the largest (ξ_ℓ) and the smallest (ξ_s) elements of ξ that is:

$$\boxed{\text{Ra}(\xi) = \xi_\ell - \xi_s \ .} \qquad (3.12)$$

Consequently, for an ascendingly ordered sample ξ, we get

$$\text{Ra}(\xi) = \xi_n - \xi_1 \ . \qquad (3.12a)$$

Note that the range of the sample can be also determined from its definition set $D \equiv \{d_j, \ j=1,2,\ldots m\}$. The corresponding expressions to (3.12) and (3.12a), respectively are:

$$\text{Ra}(\xi) = \text{Ra}(D) = d_\ell - d_s \ , \qquad (3.13)$$

and

$$\text{Ra}(\xi) = \text{Ra}(D) = d_m - d_1 . \qquad (3.13a)$$

<u>Example 3.9</u>: From example 3.8, we have the ascendingly ordered sample $\xi \equiv (1,1,1,1,1,2,2,2,4)$, <u>n=9</u> , whose definition set is $D = \{1,2,4\}$, $m=3$. To obtain the range, we use either equation (3.12a), i.e. $\text{Ra}(\xi) = \xi_n - \xi_1 = 4-1 = \underline{3}$, or we use equation (3.13a), i.e.

$$R_a(\xi) = d_m - d_1 = 4 - 1 = \underline{3}.$$

At this point, we can summarize the different characteristics of the sample ξ originated in example 3.3; as computed in the last three

sections, namely: $M = 1.\overline{6}$, $S^2 = 0.\overline{8}$, $S = 0.94$,

$$Med(\xi) = 1 \quad \text{and} \quad Ra(\xi) = 3$$

(Note that the "bar" above the last digit means that it is a periodic number).

3.1.6 Histograms and Polygons

From now on, the number of elements n of a sample ξ will be called the size of the sample. A sample with large size n, is often divided into classes (categories). Each class is a group of n_i individual elements ($n_i < n$). To achieve this, we usually determine the range of ξ (see section (3.1.5)),and then divide the range into k intervals* by (k+1) class-boundaries (class-limits). It is usual to make the intervals equidistant. The difference between the upper and lower boundaries of a class is called the class-width. The number c of elements in each class is called the class-count (class-frequency). This process in statistics is called classification of the sample. The "box" (or rectangular) graphical representation of the classified sample is called the histogram of the sample.

Example 3.10: Let us have the following random sample ξ:

$$\xi \equiv (17,3,2,8,1,5,2,4,6,15,8,9,2,3,10,9,$$
$$11,12,4,5,8,6,7,4,5),\ n = 25$$

---

* The interval from a to b is either:

- open , denoted by $(a,b) = (x : a<x<b)$
- closed , " " $[a,b] = (x : a\leq x\leq b)$
- Open-closed, " " $(a,b] = (x : a<x\leq b)$
- closed-open, " " $[a,b) = (x : a\leq x<b)$

To reconcile this known notation with the terminology of the theory of sets, it has to be understood that any such interval can be regarded as a set. To distinguish such a set from a point set, we shall call it a compact set.

First, we compute the range of ξ using equation (3.12), i.e.

$Ra(\xi) = \xi_{\ell} - \xi_{s} = 17-1 = 16.$

Let us use four intervals:

[1,5], (5,9], (9,13] and (13,17].

Hence, the class-counts will be:

$c_1([1,5]) = 12$, $c_2((5,9]) = 8$,

$c_3((9,13]) = 3$ and $c_4((13,17]) = 2$.

The histogram of the given sample in this example is shown in Figure 3.4, in which the horizontal axis represents the class boundaries and the ordinates represent the class-frequencies c_i (see the left-hand scale).

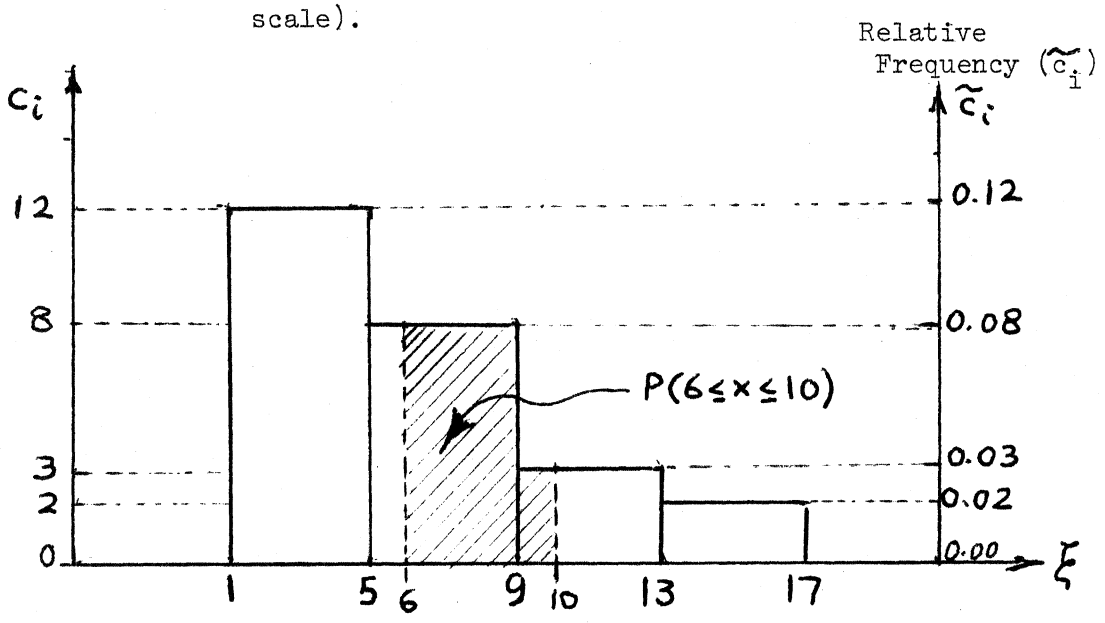


Figure 3.4

Note in the above figure that a rectangle is drawn over each interval with constant height equal to the corresponding class-count.

It is usually required that the area of, or under, the histogram has to be equal to one. Assume that we have k classes with corresponding class-counts c_i such that: $\sum_{i=1}^{k} c_i = n$. Let us denote the class-width, assumed to be constant, by Δ. Hence, the area a of the histogram is given by:

$$a = \Delta c_1 + \Delta c_2 + \ldots + \Delta c_k$$

$$= \Delta(c_1+c_2+\ldots+c_k) = \Delta \sum_{i=1}^{k} c_i = \Delta n.$$

This means that the area under the histogram equals the class-width multiplied by the size of the sample.

Therefore, to make the area of the histogram equal to one, we simply have to divide each ordinate c_i by the quantity $n\Delta$. The new (transformed) ordinate $\tilde{c}_i$ is also called the relative count (compare this to the relative count mentioned in section 3.1.2, which represents the experimental probability of an individual element; however, here we are dealing with counts in an interval).

Example 3.11: Using the data from example 3.10, we have:

$n = 25$ and $\Delta = 4$. The quantity $n\Delta = 25.4 = 100$.

Hence, to compute the relative counts $\tilde{c}_i$ of the classified sample ξ, we divide each ordinate c_i (obtained in example 3.10) by 100. This gives us:

$$\tilde{c}_1 = \frac{12}{100} = 0.12, \quad \tilde{c}_2 = \frac{8}{100} = 0.08,$$

$$\tilde{c}_3 = \frac{3}{100} = 0.03 \text{ and } \tilde{c}_4 = \frac{2}{100} = 0.02.$$

The histogram of the sample in this case will be the same as in example 3.10, with the only difference that the ordinate scale is going to be changed (see Figure 3.4, the right-hand scale).
Using the relative counts $\tilde{c}_i$, the area "a" under the histogram equals to one, as we can see from the following computation (using Figure 3.4):

$a = 4.0.12 + 4.0.08 + 4.0.03 + 4.0.02$

$= 0.48 + 0.32 + 0.12 + 0.08 = \underline{1.0}$,

which may be used as a check on the correctness of computing $\tilde{c}_i$.

Let us denote the largest and the smallest abscissas of a histogram by ℓ and s, respectively (e.g. in Figure 3.4, $\ell = 17$ and $s=1$). Notice that for any subinterval $D' = [a,b]$ of the interval $\Delta = [s, \ell]$, we can compute the area $\alpha(D')$ under the histogram. This $\alpha(D')$ will be given as a real number from $[0,1]$. Hence, α can be regarded as a function mapping any subinterval of $[s, \ell]$ onto $[0,1]$. Therefore, it is easy to see that α can be considered as a probability function (see section 2.1), more specifically one of the possible probability distribution functions (PDF's) of the sample. Obviously, such PDF (i.e. α) depends on the particular accepted classification of the sample.

From the above discussion, we find that the probability of any subinterval of $[s,\ell]$ is represented by the corresponding area under the histogram. On the other hand, the ordinates of the histogram do not represent probabilities (again compare the histogram with the bar diagram given in section 3.1.2).

Example 3.12: Referring to Figure 3.4, we may ask: what is the probability of $D' = [6, 10]$; or, what is the probability that the sample element, say x, lies between 6 and 10. This can be written as:

$$P(6 \leq x \leq 10) = \ ?$$

The answer will be given by the area under the histogram between 6 and 10 (which is shaded in Figure 3.4), i.e.

$$P(6 \leq x \leq 10) = P(6 \leq x < 9) + P(9 < x \leq 10) =$$

$$= (9-6).0.08 + (10-9).0.03$$

$$= 3.0.08 + 1.0.03 = 0.24 + 0.03$$

$$= \underline{0.27} \ .$$

On the other hand, by inspecting the actual sample ξ originated in example 3.10 we find out that the actual number of elements in the interval [6,10] is nine. This number represents $(\frac{9}{25}).100\% = \underline{36\%}$ of the sample. Or, we say that the actual probability $P(6 \leq x \leq 10) = \underline{0.36}$, which does not agree precisely with the result obtained when using the corresponding histogram (i.e. 0.27).

The difference between the actual probability and the computed probability using the histogram, as experienced in example 3.12, is largely dependent on the chosen classification of the sample (selection of the class-intervals). Usually, one gets a smaller difference (better agreement) by selecting the class-boundaries so as not to coincide with any of the elements of the given sample. The construction of histograms can be considered a subject of its own right. We are not going to venture into this subject any deeper.

Example 3.13: If we, for instance, use the following classification (for the sample ξ given in example 3.10): [0.7, 4.8], [4.8, 8.9], [8.9, 13] and [13, 17.1], i.e. we have again four equal intervals, for which $\Delta = 4.1$. Then we get the class-counts as $c_1 = 9$, $c_2 = 9$, $c_3 = 5$ and $c_4 = 2$. The quantity $n\Delta = 25.4.1 = 102.5$. Hence, the relative counts are:

$$\tilde{c}_1 = \tilde{c}_2 = \frac{9}{102.5} \doteq 0.0878,$$

$$\tilde{c}_3 = \frac{5}{102.5} \doteq 0.0488, \text{ and}$$

$$\tilde{c}_4 = \frac{2}{102.5} \doteq 0.0195.$$

In this case, the new histogram of the sample ξ is shown in Figure 3.5.

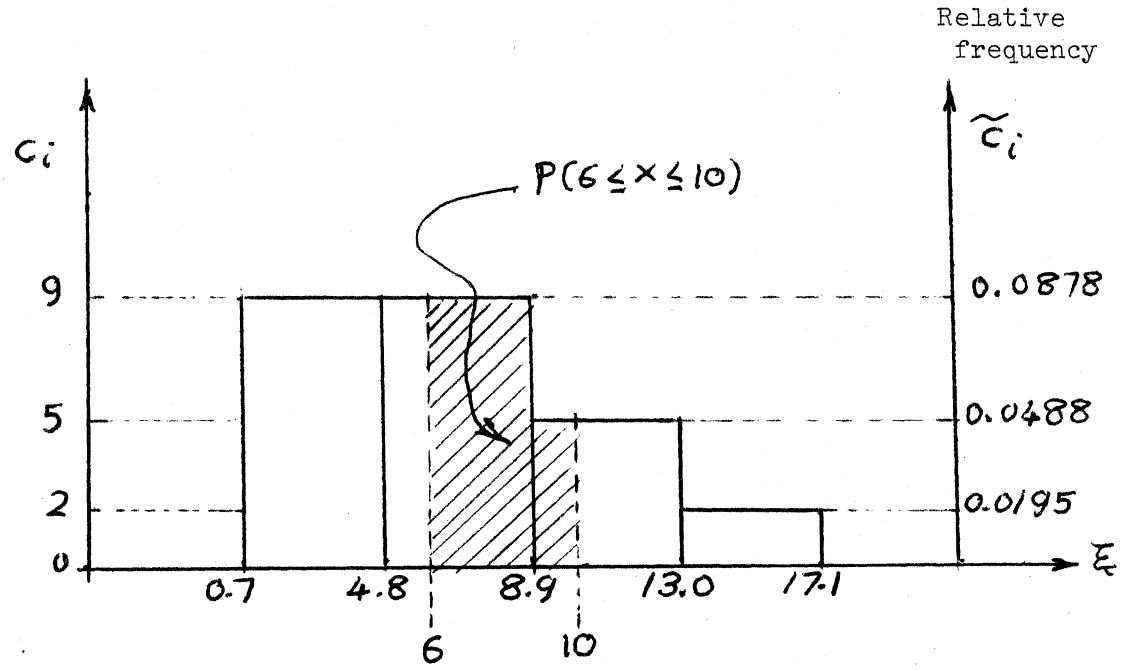


Figure 3.5

The probability $P(6 \leq x \leq 10)$ is computed as follows (shaded area in Figure 3.5):

$$P(6 \leq x \leq 10) = P(6 \leq x \leq 8.9) + P(8.9 < x \leq 10)$$

$$= 2.9.0.0878 + 1.1.0.0488 \doteq$$

$$\doteq 0.2546 + 0.0537 \doteq$$

$$\doteq 0.3083 \doteq \underline{0.31},$$

which gives a better agreement with the actual probability than the classification used in example 3.11.

The graphical representation of a histogram, which uses the central point of each box (class-midpoint) and its ordinate (the corresponding relative class-count), is called a polygon.

In order to make the total area under the polygon equal to one we have to add one more class interval on each side (tail) of the corresponding histogram. The midpoints s' and ℓ' of these two, lower and upper tail intervals, are used to close the polygon.

Therefore, it can be easily seen that the area α' under the polygon has again the properties of probability. This means that α' is one of the possible PDF's of the sample. Hence α' can be used for determining the probability of any $D' = [a, b] \subset [s', \ell']$. Note also here that the ordinates of the polygon do not represent probabilities.

Example 3.14: The polygon corresponding to the histogram of Figure 3.4 is illustrated in Figure 3.6.

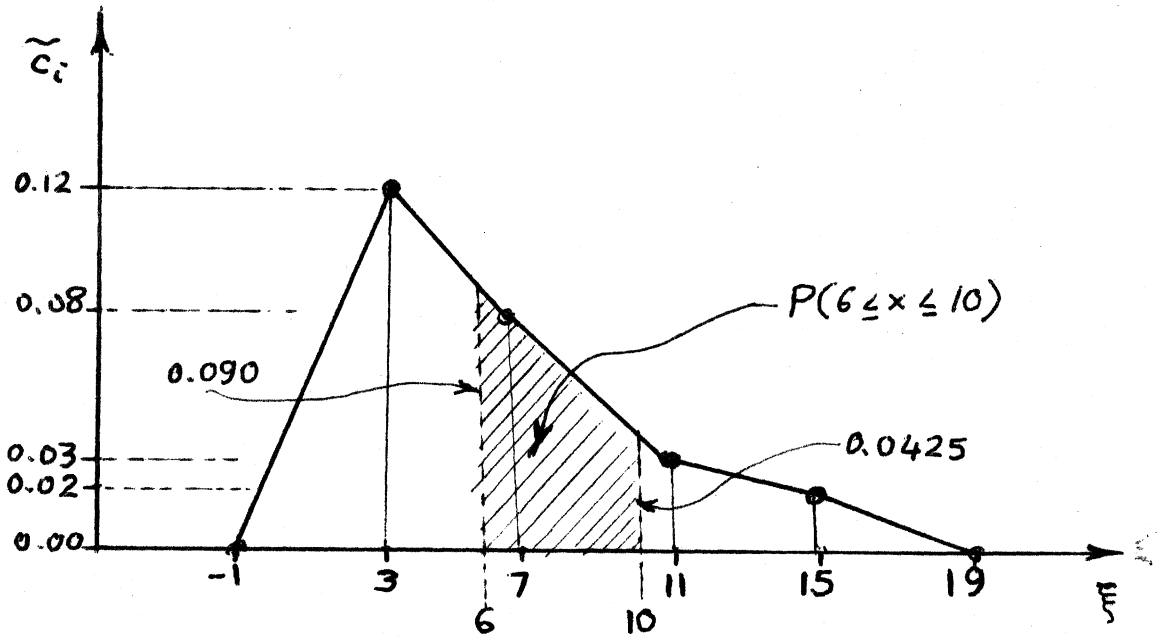


Figure 3.6

Similar to the histogram, the area"a"under the polygon should be equal to one. To show that this is the case, we compute"a"using Figure 3.6 as:

$$a = 4\left(\frac{1}{2} \cdot 0.12 + \frac{1}{2}(0.12 + 0.08) + \right.$$
$$\left. + \frac{1}{2}(0.08 + 0.03) + \frac{1}{2}(0.03 + 0.02) + \right.$$
$$\left. + \frac{1}{2} \cdot 0.02\right)$$
$$= 2(0.12 + 0.20 + 0.11 + 0.05 + 0.02)$$
$$= 2\ (0.50) = \underline{1.00}.$$

Let us compute the probability $P(6 \leq x \leq 10)$ using the polygon (the required probability is represented by the shaded area in Figure 3.6). To achieve this, we first have to interpolate the ordinates corresponding to 6 and 10, which are found to be 0.090 and 0.0425, respectively. Therefore, the required probability is:

$$P(6 \leq x \leq 10) = P(6 \leq x \leq 7) + P(7 \leq x \leq 10)$$
$$= 1.\frac{1}{2}(0.09+0.08)+3.\frac{1}{2}(0.08+0.0425)$$
$$= 1.0.085 + 3.0.06125$$
$$= 0.085 + 0.184 \doteq \underline{0.27},$$

which is the same as the value obtained when using the corresponding histogram.

So far, we have constructed the histogram and the polygon corresponding to the PDF of a sample. Completely analogously, we may construct the histogram and the polygon corresponding to the CDF of the sample which will be respectively called the cumulative histogram and the cumulative polygon. In this case, we will use a modified form of equation (3.2), namely

$$C(a) = P(x \leq a) = \sum_{x_i \leq a} P(x_{i-1} < x \leq x_i) \cdot \qquad (3.14)$$

Example 3.15: Let us plot the cumulative histogram and cumulative polygon of the sample ξ used in the examples of this section.

For the cumulative histogram, we get the following by using Figure 3.4:

$C(1) = P(1) = 0$ (remember that the probability of individual elements from the histogram or polygon is always zero).

$C(5) = P[1,5] = 4.0.12 = 0.48,$

$C(9) = C(5)+P(5,9] = 0.48 + 4.0.08$

$= 0.48 + 0.32 = 0.80,$

$C(13) = C(9)+P(9,13] = 0.80 + 4.0.03$

$= 0.80 + 0.12 = 0.92,$

$C(17) = C(13)+P(13,17] = 0.92 + 4.0.02$

$= 0.92 + 0.08 = \underline{1.00}.$

Figure 3.7 is a plot of the above computed cumulative histogram.

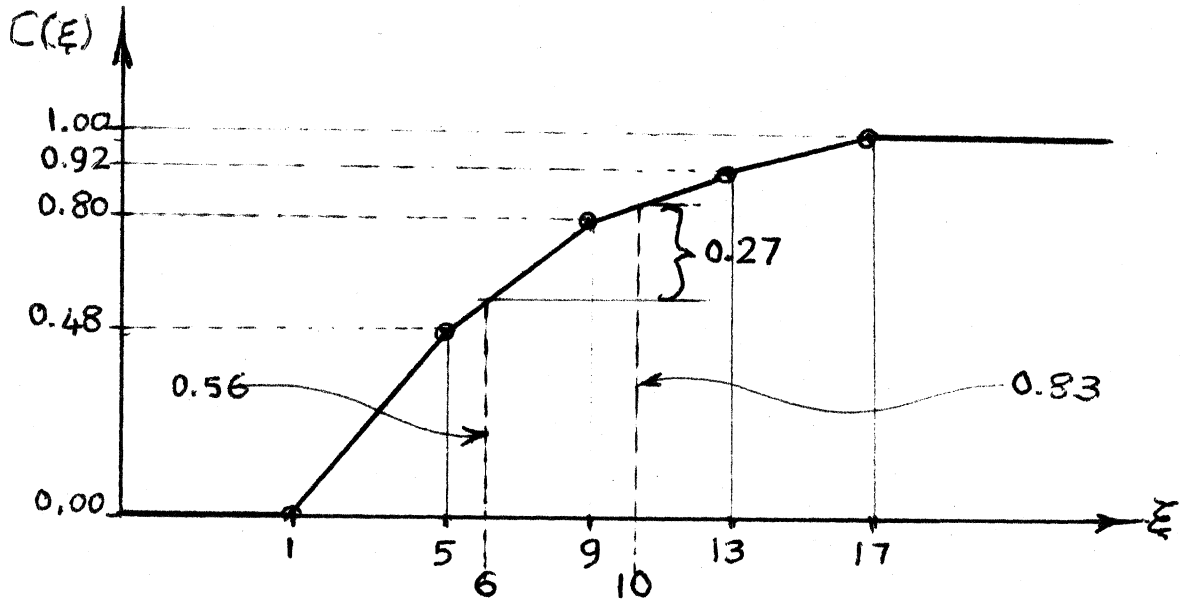


Figure 3.7

For the cumulative polygon, we get the following by using Figure 3.6:

$C(-1) = 0,$

$C(3) = P[-1,3] = \frac{1}{2}(4.0.12) = \frac{1}{2}(0.48) = 0.24,$

$C(7) = C(3)+P(3,7] = 0.24 + \frac{1}{2}.4(0.12 + 0.08)$

$= 0.24 + 0.40 = 0.64,$

$C(11) = C(7)+P(7,11] = 0.64 + \frac{1}{2}.4(0.08 + 0.03)$

$= 0.64 + 0.22 = 0.86,$

$C(15) = C(11)+P(11,15] = 0.86 + \frac{1}{2} .4(0.03 + 0.02)$

$= 0.86 + 0.10 = 0.96,$

$C(19) = C(15)+P(15,19] = 0.96 + \frac{1}{2} . 4.0.02$

$= 0.96 + 0.04 = 1.00$

Figure 3.8 is a plot of the above computed cumulative polygon (note here, as well as in Figure 3.7, the properties of the CDF mentioned in example 3.4).

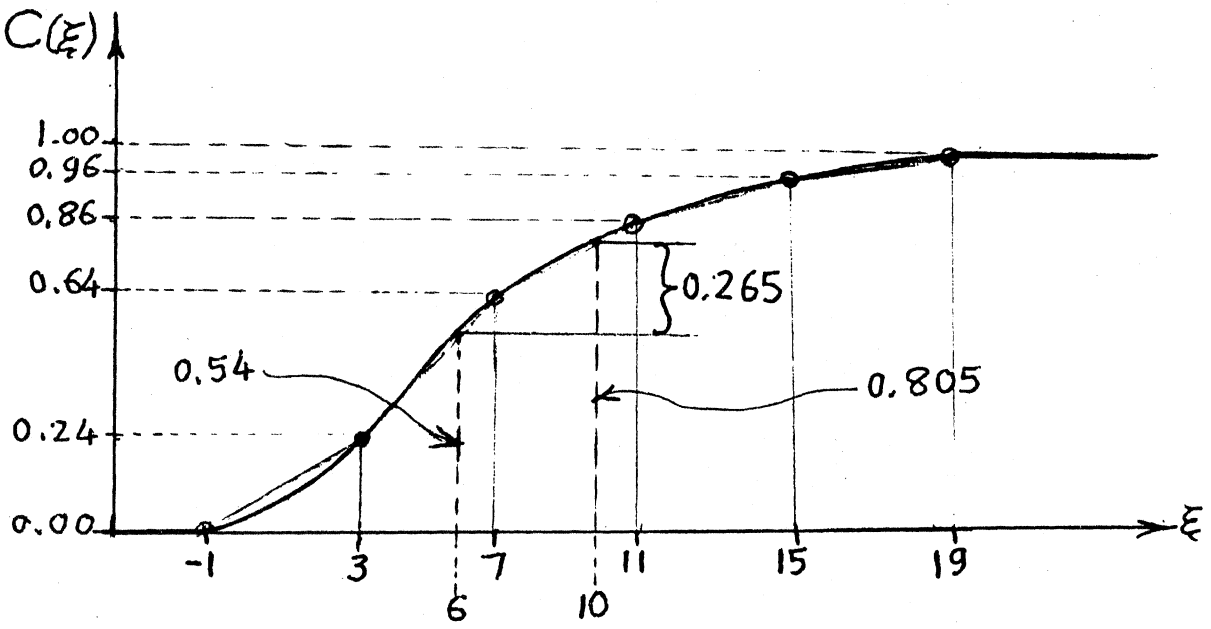


Figure 3.8

By examining Figures 3.7 and 3.8, we can see that the cumulative polygon uses the central point of each class-interval along with its ordinate from the corresponding cumulative histogram. Therefore, the relationship between the cumulative polygon and its corresponding cumulative histogram is exactly the same as the relationship between the polygon and its corresponding histogram.

Because of the nature of the CDF, we can see that the cumulative probability - represented by an area under the PDF extending to the left-most point - is represented just by an ordinate of the cumulative histogram or the cumulative polygon. Hence the cumulative histogram or the cumulative polygon can be used to determine the probability P[a,b], a<b, simply by subtracting the ordinate corresponding to a from the one corresponding to b.

Example 3.16: Let us compute the probability P[6,10] by using:

(i) the cumulative histogram of Figure 3.7,

(ii) the cumulative polygon of Figure 3.8.

First, we get the following by using Figure 3.7: The interpolated ordinates corresponding to 6 and 10 are found to be 0.56 and 0.83, respectively. Therefore, $P[6,10] = P(6 \leq x \leq 10) =$ $= 0.83-0.56 = \underline{0.27}$, which is the same value as the one obtained when using the histogram (example 3.12).

Second, we get the following by using Figure 3.8: The interpolated ordinates corresponding to 6 and 10 are found to be 0.54 and 0.805, respectively. Therefore:

$P[6,10] = P(6 \leq x \leq 10) = 0.805-0.54 = \underline{0.27}$,

which is again the same value as the one obtained when using the polygon (example 3.14).

To close this section, we should point out that both the histograms and the polygons (non-cumulative as well as cumulative) can be refined by refinning the classification of the sample. Note that this refinement makes the diagrams look smoother.

3.2 Statistics of a Random Variable

3.2.1 Random (Stochastic) Function and Random (Stochastic) Variable

In order to be able to solve the problems connected with interval probabilities (see the histograms and polygons of section 3.1.6) more easily and readily, the science of statistics has developed a more convenient approach. This approach is based on the replacement of the troublesome numerical functions defined on the discrete definition set of a random sample, by more suitable functions. To do so, we first define two idealizations of the real world: the random (stochastic) function and the random (stochastic) variable.

A random or stochastic function is defined as a function X mapping an unknown set U* into R, that is

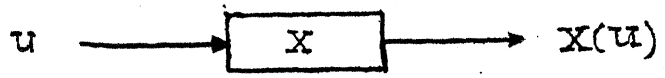
$$x \varepsilon \{U \rightarrow R\}$$

(Later on, concepts of multi-valued $X \varepsilon \{U \rightarrow R^m\}$ (where R^m is the Cartesian m-power of R, see section 1.3) are developed.)

This statement is to be understood as follows: For any value of the argument $u \in U$, the stochastic function x assumes a value $x(u) \varepsilon R$. But, because the set U is considered unknown, there is no way any formula for x can be written and we have to resort to the following "abstract experiment" to show that the concept of random functions can be used.

"Note that in experimental sciences the set U may be fully or at least partly known. The science of statistics however, assumes that it is either not known, or works with the unknown part of it only.

Suppose that the function x is realised by a device or a process (see the sketch) that produces a functional



value x(u) every time we trigger it. Knowing nothing about the inner workings of the process all we can do is to record the outcomes x(u). When a large enough number of values x(u) have been recorded, we can plot a histogram showing the relative count of the x(u) values within any interval $[x_o, x_1]$. In this abstraction we can imagine that we have collected enough values to be able to compute the relative counts for any arbitrarily small interval dx and thus obtain a 'smooth histogram'. Denoting the limit of the relative count divided by the width dx of the interval [x, x + dx], for dx going to zero, by $\phi(x)$ we end up with a function ϕ that maps $x \in R$ into R.

Going now back to the realm of mathematics, we see that the outcome of the stochastic function can be viewed as a pair $(x(u), \phi(x))$. This pair is known as the random (stochastic) variable. It is usual in literature to refer just to the values x(u) as random variable with the tacit understanding that the function ϕ is also known.

We note that the function ϕ is thus defined over the whole set of real numbers R and has either positive or zero values, i.e. ϕ is non-negative on all R. Further, we shall restrict ourselves to only such ϕ that are integrable on R in the Riemannian sense, i.e. are at least piece-wise continuous on R.

3.2.2 PDF and CDF of a Random Variable

The function ϕ described in 3.2.1, belonging to the random variable x, is called the probability distribution function (PDF) of the random variable. It can be regarded as equivalent to the experimental

PDF (see 3.1.2) of a random sample. From our abstract experiment it can be seen that

$$\int_{-\infty}^{\infty} \phi(x)\ dx = 1 \tag{3.15}$$

since the area under the "smooth histogram" must again equal to 1 (see 3.1.6). This is the third property of a PDF, the integrability and non-negativeness being the first two. We note that eq. 3.15 is also the necessary condition for $\phi(x)\ dx$ to be called probability (see 2.1).

Figure 3.9 shows an example of one such PDF, i.e. ϕ in which the integral (3.15) is illustrated by the shaded area under the ϕ.

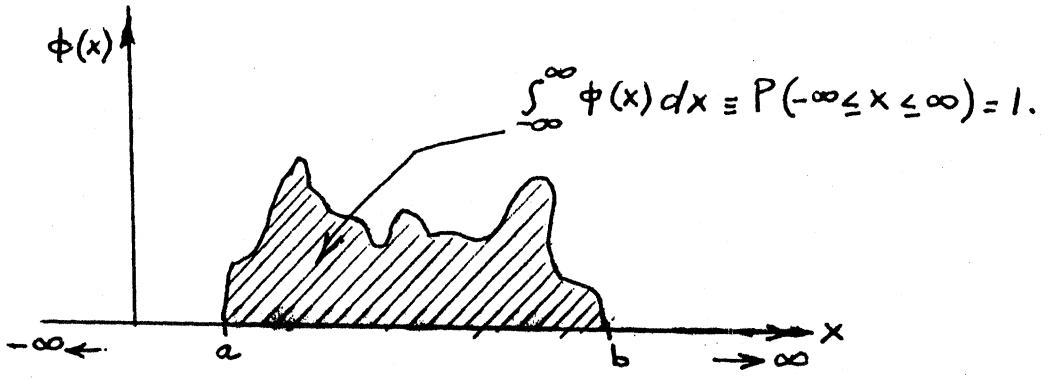


Figure 3.9

The definite integral of the PDF, ϕ, over an interval $D' \subset D$ is called the probability of D'. So, we have in particular:

$$\int_{-\infty}^{x_o} \phi\ (x)\ dx = P(x \leq x_o) \quad \varepsilon [0,\ 1]\ , \tag{3.16a}$$

$$\int_{x_o}^{\infty} \phi(x)\ dx = P(x \geq x_o) \quad \varepsilon [0,\ 1]\ , \tag{3.16b}$$

$$\int_{x_1}^{x_2} \phi(x)\ dx = P(x_1 \leq x \leq x_2) \quad \varepsilon [0,\ 1]\ . \tag{3.16c}$$

Consequently,

$$P(x \geq x_o) = 1 - P\ (x \leq x_o)\ . \tag{3.17}$$

The integrals (3.16a), (3.16b) and (3.16c) are represented by the corresponding shaded areas in Figure 3.10: a, b, and c, respectively.

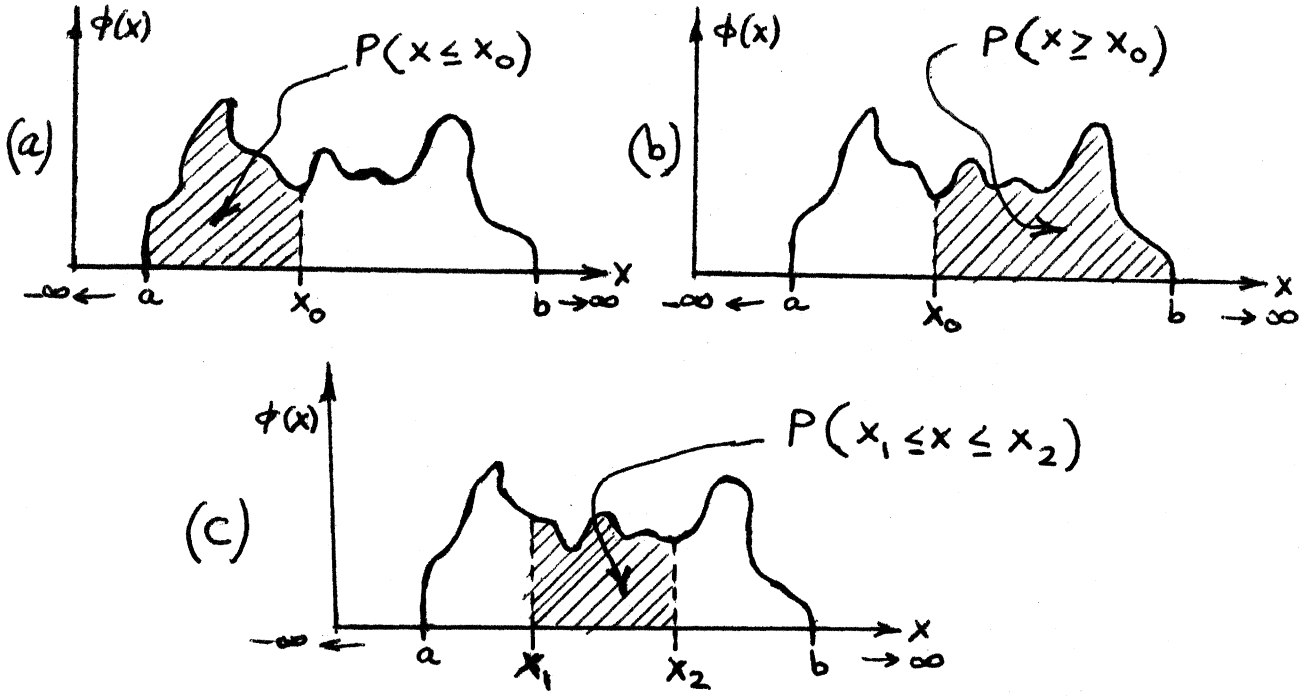


Figure 3.10

At this point, the difference between <u>discrete</u> and <u>compact</u> probability spaces, should be again born in mind. In the discrete space, the value of the PDF at any point, which is an element of the discrete definition set of the sample, can be interpreted as a probability (section 3.1.2). However in the compact space, it is only the area under the PDF, that has got the properties of probability.* We have already met this problem when dealing with histograms.

---

* The whole development for the discrete and the compact spaces could be made identical using either Dirac's functions or a more general definition of the integral.

Note further that:

$$P\ (x = x_o) = \int_{x_o}^{x_o} \phi(x)\ dx = 0^*).$$

Analogous to section 3.1.2, the function Ψ defined as

$$\Psi(x) = \int_{-\infty}^{x} \phi(y)\ dy,\ \varepsilon[x \to [0, 1]\} \quad , \tag{3.18}$$

where $y \varepsilon R$ is a dummy variable in the integration, is called a CDF provided that ϕ is a PDF. Ψ is again a non-negative, never decreasing function, and determines the probability $P(x \leq x_o)$. (Compare this with section 3.1.2); namely:

$$\Psi(x_o) = P(x \leq x_o)\ \varepsilon\ [0, 1]\ . \tag{3.19}$$

Figure 3.11 shows how the CDF (corresponding to the PDF in Figure 3.9) would look.

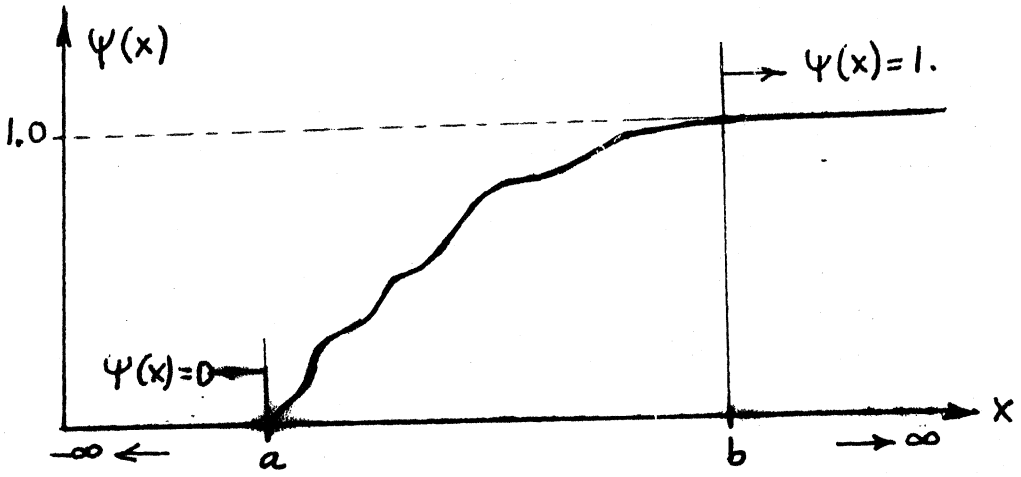


Figure 3.11

---

* This may not be the case for a more general definition of the integral, or for ϕ being the Dirac's function.

If ϕ is symmetrical, Ψ will be "inversely symmetrical" around the axis $\Psi(x) = 1/2$. Figure 3.12 is an example of such a case.

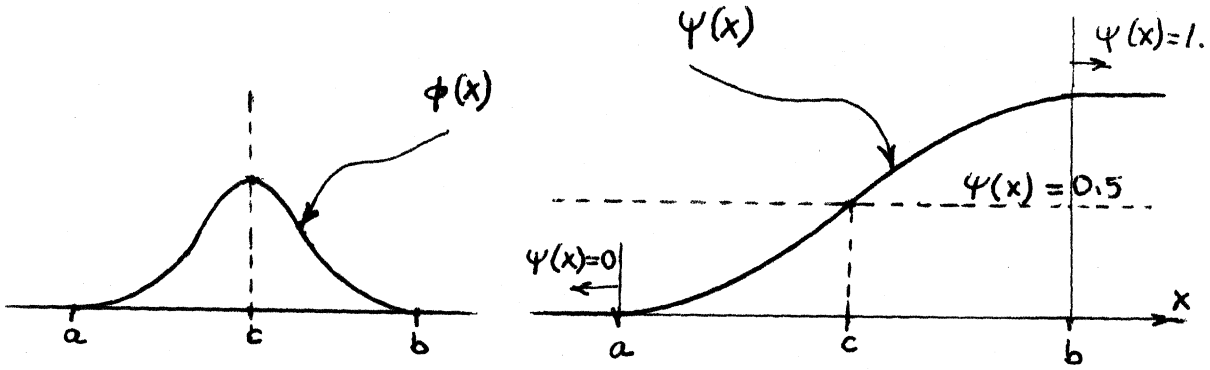


Figure 3.12

Note that Ψ is the primitive function of ϕ since we can write:

$$\phi(x) = \frac{d\,\Psi(x)}{dx} .$$

In addition, we can see that $\phi(x)$ has to disappear in the infinities in order to satisfy the basic condition:

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

Hence, we have:

$$\lim_{x\to-\infty} \Psi(x) = 0 , \quad \lim_{x\to\infty} \Psi(x) = 1 .$$

3.2.3 Mean and Variance of a Random Variable

It is conceivable that the concept of a random variable is useless if we do not know (or assume) its PDF. On the other hand, we do not have the one-to-one relation between the random variable and its PDF as we had with

the random samples (section 3.1.1 and 3.1.2). The random variable acts only as an argument for the PDF.

The random variable can be thus regarded as an argument of the function called PDF, that runs from minus infinity to plus infinity. Therefore, strictly speaking, we cannot talk about the "mean" and the "variance" of a random variable, in the same sense as we have talked about the "mean" and the "variance" of a random sample. On the other hand, we can talk about the value of the argument of the centre of gravity of the area under the PDF. Similarly, we can define the variance related to the PDF. It has to be stated, however, that it is a common practice to talk about the mean and the variance of the random variable; and this is what we shall do here as well.

The <u>mean</u> μ of the random variable x is defined as:

$$\boxed{\mu = \int_{-\infty}^{\infty} x\, \phi(x)\, dx\ .} \qquad (3.20)$$

Note the analogy of (3.20) with equation (3.4), section 3.1.3.

μ is often written again in terms of an operator E^*; usually we write

$$E^*(x) = \mu = \int_{-\infty}^{\infty} x\, \phi(x)\, dx\ .^{\dagger)} \qquad (3.21)$$

---

† E^* is again an abbreviation for the mathematical Expectation, similar to the operator E mentioned in section 3.1.3. However, we use the "asterisk" here to distinguish between both summation procedures, namely: E implies the summation using Σ; and E^* implies the summation using $\int$.

We can see that the argument in the operator E* is $x \cdot \phi(x)$ rather than x, x being just a dummy variable in the integration. However, we shall again use the customary notation to conform with the existing literature.

We have again, the following properties of E*, where k is a constant: (i) $E^*(kx) = k\,E^*(x)$;

(ii) $E^*\left(\sum_{j=1}^{r}(x^j)\right)=\sum_{j=1}^{r} E^*(x^j)$, where x^j, $j = 1, 2, \ldots r$, are r different "random variables", i.e., r random variables with appropriate PDF's;

(iii) and we also define:

$$E^*(E^*(x)) = E^*(x) = \mu^{\dagger\dagger}).$$

The variance σ^2 of a random variable x with mean μ, is defined as:

$$\boxed{\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2\, \phi(x)\, dx\ .} \qquad (3.22)$$

Note the analogy of (3.22) with equation (3.8), section 3.1.4. The square root of σ^2, i.e. σ, is again called the standard deviation of the random variable.

Carrying out the operation prescribed in (3.22) we get:

$$\sigma^2 = \int_{-\infty}^{\infty} [x^2\, \phi(x) - 2x\mu\phi(x) + \mu^2\phi(x)]\, dx$$

$$= \int_{-\infty}^{\infty} x^2\, \phi(x)\, dx - 2\mu \int_{-\infty}^{\infty} x\, \phi(x)\, dx + \mu^2 \int_{-\infty}^{\infty} \phi(x)\, dx\ .$$

---

††In order to prove this equation, one has to again use the Dirac's function as the PDF of E*(x).

In the above equation, we know that the integral in the second term equals μ (equation (3.20)), and the integral in the last term equals one (equation (3.15)). Therefore, by substituting we get:

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 \phi\ (x)\ dx - \mu^2. \tag{3.23}$$

Note the similarity of the first term in equation (3.23) with $E(\xi^2) = \sum_{j=1}^{m} d_j^2\ P(d_j)$ (section 3.1.4). This gives rise to an often used notation:

$$\sigma^2 = E^*\ (x-\mu)^2 = E^*\ (x-E^*(x))^2\ . \tag{3.23a}$$

We shall again accept this notation as used in the literature, bearing in mind that E* is <u>not</u> operating on the argument, but on the product of the argument with its PDF.

The expression

$$m_r = \int_{-\infty}^{\infty} x^r \phi(x)\ dx \tag{3.24}$$

is usually called the <u>r-th moment</u> of the PDF (random variable); more precisely; the r-th moment of the PDF <u>about zero</u>. On the other hand, the <u>r-th central moment</u> of the PDF is given by:

$$m'_r = \int_{-\infty}^{\infty} (x-\mu)^r\ \phi\ (x)\ dx\ . \tag{3.25}$$

By inspecting the above expressions for m_r and m'_r along with equations (3.20) and (3.22), we can see that:

$$\mu = m_1 \tag{3.26a}$$

and

$$\sigma^2 = m'_2 = m_2 - \mu^2 = m_2 - m_1^2 \quad . \tag{3.26b}$$

Compare the above result (3.26 a, b) with the analogy to mechanics mentioned in sections 3.1.3 and 3.1.4.

3.2.4 Basic Postulate (Hypothesis) of Statistics, Testing

The basic postulate of statistics is that "any random sample has got a parent random variable". This parent random variable $x \in R$ is usually called population and is considered to be infinite. It is common in statistics to postulate the PDF of the population for any random sample, and call it the postulated, or the underlying PDF. Such a postulate may be hence tested for statistical validity.

In order to be able to test the statistical validity we have to assume that the sample can be regarded as having been picked out, or drawn from the population, each element of the sample independently from the rest. This additional property of a sample is required by the standard definition of a random sample as used in statistical literature. However, since the present Introduction does not deal with statistical testing we shall keep using our original, more general definition.

There are infinitely many families of PDF's. Every such family is defined by one or more independent parameters, whose values characterize the shape of its PDF. The individual members of a family vary according to the value of these parameters. It is common to use if possible, the mean and the standard deviation as the PDF's parameters. The less parameters the family of PDF's contains the better; the easier it is to work with.

The usual technique is that we first select the "appropriate" family of PDF's on the basis of experience and then try to find such values

of its parameters that would fit the actual random sample the best. In other words, the shape of the postulated $\phi(x)$ is chosen first; then, its parameters are computed using some of the known techniques.

Since we shall be dealing with the samples and the random variables (populations) at the same time, we shall use, throughout these notes, the latin letters for the sample characteristics, and the corresponding greek letters for the population characteristics as we have done so far.

3.2.5 Two Examples of a Random Variable

Example 3.17: As the first example, let us investigate a random variable **x** with rectangular (uniform) PDF, which is symmetrical around a value $x = k$. Let the probability of $x < k-q$ and $x > k+q$, be zero. Obviously, this PDF has the following analytical form (see Figure 3.13):

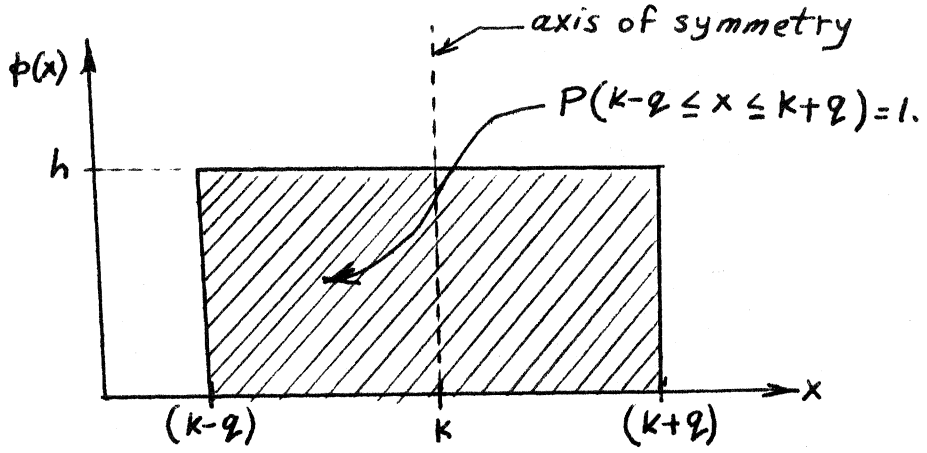


Figure 3.13

$$\phi(x) = \begin{cases} h, & \text{for } (k - q \leq x \leq k + q) \\ 0, & \text{for } (x < k - q) \text{ and } (x > k + q). \end{cases}$$

This can be written in an abbreviated form as:

$$\phi(x) = \begin{cases} h, \text{ for } (|x-k| \leq q). \\ 0, \text{ for } (|x-k| > q). \end{cases}$$

The above ϕ contains apparently three parameters k, q and h. However, only two are independent, since one can be eliminated from the condition (3.15), i.e.:

$$\int_{-\infty}^{\infty} \phi(x)\, dx = 1$$

that must be satisfied for any ϕ to be a PDF. Let us eliminate for instance the parameter h. We can write:

$$\int_{-\infty}^{\infty} \phi(x)\, dx = \int_{-\infty}^{k-q} \phi(x)\, dx + \int_{k-q}^{k+q} \phi(x)\, dx + \int_{k+q}^{\infty} \phi(x)\, dx$$

$$= 0 + \int_{k-q}^{k+q} h\, dx + 0$$

$$= h \int_{k-q}^{k+q} dx = h\,[x]_{k-q}^{k+q} = 2hq = 1.$$

This means that $h = \frac{1}{2q}$, and therefore:

$$\phi(x) = \begin{cases} 1/(2q), \text{ for } (|x-k| \leq q) \\ 0, \text{ for } (|x-k| > q) \end{cases}$$

The corresponding CDF to the above ϕ is:

$$\Psi(x) = \int_{-\infty}^{x} \phi(x)\, dx = \begin{cases} 0, \text{ for } (x \leq k-q) \\ \frac{1}{2q} \int_{k-q}^{x} dx = \frac{1}{2q}(x-k+q), \text{ for } (|x-k| \leq q). \\ 1, \text{ for } (x \geq k+q), \end{cases}$$

and is shown in Figure 3.14.

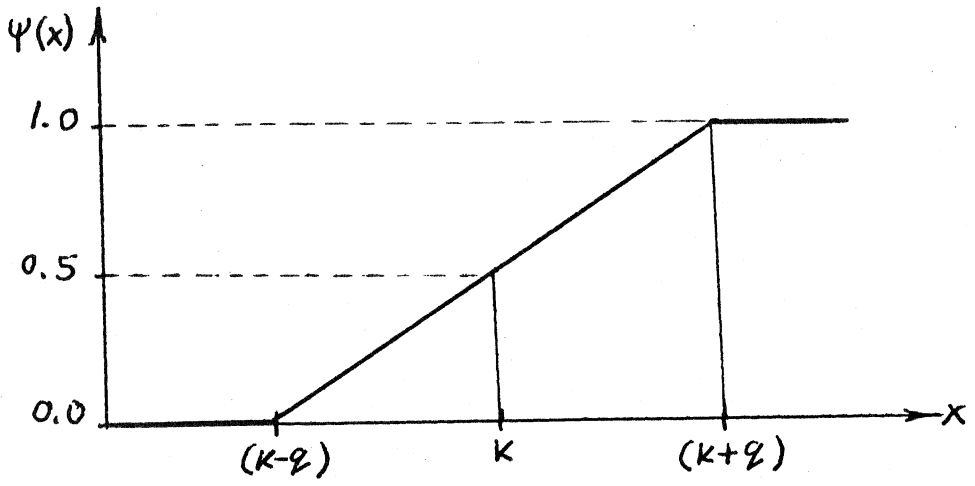


Figure 3.14

From the above figure we see that the function ψ is linear in the interval over which $\phi \neq 0$, and is constant everywhere else. Note that:

$$\phi(x) = \frac{d\,\psi\,(x)}{dx}$$

The mean of the given PDF is computed from equation (3.20) as follows:

$$\mu = \int_{-\infty}^{\infty} x\phi\,(x)dx = \frac{1}{2q}\int_{k-q}^{k+q} x dx = \frac{1}{2q}\left[\frac{x^2}{2}\right]_{k-q}^{k+q}$$

$$= \frac{1}{4q}\,(k^2 + 2kq + q^2 - k^2 + 2kq - q^2)$$

$$= \frac{4kq}{4q} = k.$$

This result satisfies our expectation for a symmetrical function around k. The variance of the given PDF can be obtained from equation (3.22), yielding

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2\ \phi(x)\,dx = \frac{1}{2q}\int_{k-q}^{k+q} (x-k)^2\,dx =$$

$$= \frac{1}{2q} \int_{k-q}^{k+q} x^2\,dx - \frac{2k}{2q} \int_{k-q}^{k+q} x\,dx + \frac{k^2}{2q} \int_{k-q}^{k+q} dx$$

$$= \frac{1}{2q} \left[\frac{x^3}{3}\right]_{k-q}^{k+q} - 2k^2 + k^2$$

$$= \frac{1}{6q} (k^3 + 3k^2q + 3kq^2 + q^3 - k^3 + 3k^2q - 3kq^2 + q^3) - k^2$$

$$= \frac{6k^2q}{6q} + \frac{2q^3}{6q} - k^2 = \frac{q^2}{3} \quad .$$

Since $k = \mu$ and $q = \sqrt{3}\sigma$, then $h = \frac{1}{2q} = \frac{1}{2\sqrt{3}\sigma}$, and we can express the given rectangular PDF, which we will denote by R, in terms of its mean μ and its standard deviation σ as follows:

$$R(\mu, \sigma\,; x) = \phi(x) = \begin{cases} \frac{1}{2\sqrt{3}\sigma}, & \text{for } (|x-\mu|) \leq \sqrt{3}\sigma) \\ 0, & \text{for } (|x-\mu|) > \sqrt{3}\sigma) \end{cases} .$$

Similarly, we can express its corresponding CDF, which we will denote by R_c, in terms of μ and σ, as follows:

$$R_c(\mu,\sigma;x) = \Psi(x) = \begin{cases} 0, & \text{for } (x \leq \mu - \sqrt{3}\sigma). \\ \frac{1}{2\sqrt{3}\sigma} (x-\mu+\sqrt{3}\sigma), & \text{for } (|x-\mu| \leq \sqrt{3}\sigma). \\ 1, & \text{for } (x \geq \mu + \sqrt{3}\sigma) . \end{cases}$$

Assume that we would like to compute the probability of $x \in [\mu-\sigma, \mu+\sigma]$, where x has the rectangular PDF. This can be done by using equation (3.16c) and Figure 3.15, as follows:

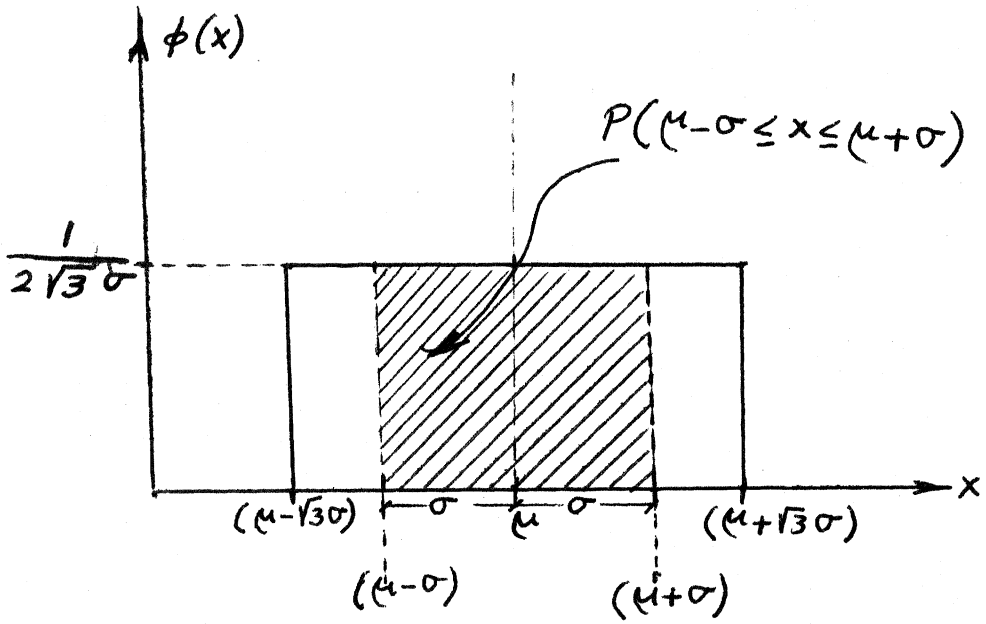


Figure 3.15

$$P(\mu-\sigma \leq x \leq \mu+\sigma) = \int_{\mu-\sigma}^{\mu+\sigma} \phi(x)\, dx$$

$$= \int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{2\sqrt{3}\sigma}\, dx = \frac{1}{2\sqrt{3}\sigma}\,[2\sigma]$$

$$= \frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3} = \frac{1.732}{3} \doteq 0.577 \doteq \underline{0.58}.$$

The above probability is given by the shaded area in Figure 3.15.

Similarly, for this particular uniform PDF, we find that:

$P(\mu-2\sigma \leq x \leq \mu+2\sigma) \equiv P(\mu-3\sigma \leq x \leq \mu+3\sigma) \equiv \underline{1.0}$

In statistical testing, we often need to compute the moments of the PDF (see section 3.2.3). Let us, for instance, compute the third moment m_3 about zero of the rectangular PDF. We will use equation (3.24), i.e.

$$m_3 = \int_{-\infty}^{\infty} x^3 \phi(x)\, dx = \int_{\mu-\sqrt{3}\sigma}^{\mu+\sqrt{3}\sigma} x^3 \frac{1}{2\sqrt{3}\sigma}\, dx$$

$$= \frac{1}{2\sqrt{3}\sigma} \left[\frac{x^4}{4}\right]_{\mu-\sqrt{3}\sigma}^{\mu+\sqrt{3}\sigma}$$

$$= \frac{1}{8\sqrt{3}\sigma} \left[(\mu + \sqrt{3}\sigma)^4 - (\mu - \sqrt{3}\sigma)^4\right]$$

$$= \frac{1}{8\sqrt{3}\sigma} \left(8\sqrt{3}\sigma\mu^3 + 24\sqrt{3}\,\sigma^3\mu\right)$$

$$m_3 = \underline{\mu^3 + 3\sigma^2\mu}$$

Example 3.18: As a second example let us investigate a random variable with a triangular PDF, which is symmetrical around $x = k$. Let us assume that the probability of $x < k - q$ and $x > k + q$, be zero. We may write (see Figure 3.16):

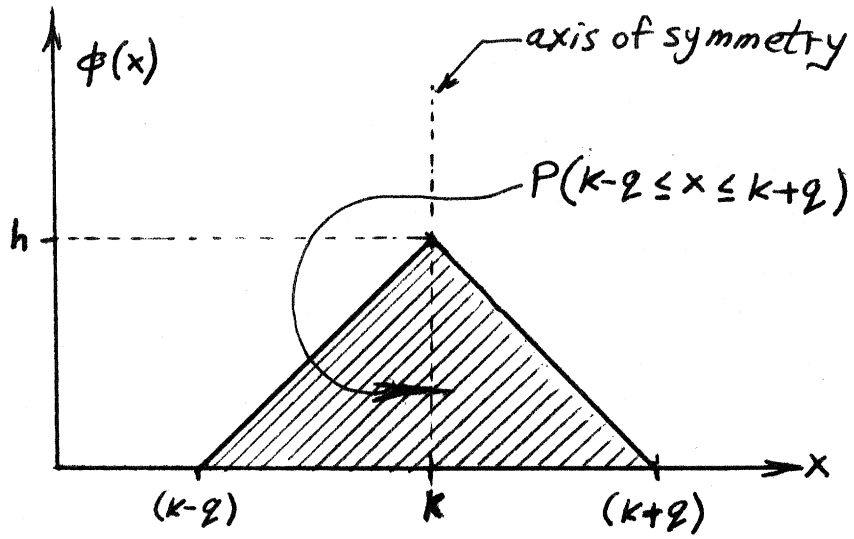


Figure 3.16

$$\phi(x) = \begin{cases} 0, \text{ for } (x \leq k - q). \\ \frac{h}{q}(x - k + q), \text{ for } (k - q \leq x \leq k). \\ \frac{h}{q}(-x + k + q), \text{ for } (k \leq x \leq k + q). \\ 0, \text{ for } (x \geq k + q). \end{cases}$$

This can be rewritten in the following abbreviated form as:

$$\phi(x) = \begin{cases} \frac{h}{q}(q - |x - k|), \text{ for } (|x - k| \leq q) \\ 0, \text{ for } (|x - k| \geq q). \end{cases}$$

From the above, we can see that the triangular PDF has the same parameters (k, q, h) as the uniform PDF of example 3.17. Let us again eliminate the parameter h from the condition $\int_{-\infty}^{\infty} \phi(x)\, dx = 1$. This integral is nothing else but the area of the triangle, so that we can write: $\frac{1}{2} \cdot 2q \cdot h = qh = 1$. This gives us: $h = \frac{1}{q}$, and hence,

$$\phi(x) = \begin{cases} \frac{1}{q} - \frac{|x-k|}{q^2}, \text{ for } (|x-k| \leq q). \\ 0, \text{ for } (|x-k| \geq q). \end{cases}$$

The computations of the mean and the variance of the triangular PDF can be performed by following the same procedure as we have done for the rectangular PDF in example 3.17. We state here the results without proof, and the verification is left to the student.

The mean μ of the given triangular PDF equals to k, and the variance σ^2 comes out as $\frac{1}{6} q^2$.

Since $k = \mu$ and $q = \sqrt{6}\sigma$, we can again express the triangular PDF, which we will denote by T, in terms of its mean μ and its standard deviation σ, as follows:

$$T(\mu, \sigma; x) = \phi(x) = \begin{cases} \frac{1}{\sqrt{6}\sigma} - \frac{|x-\mu|}{6\sigma^2}, & \text{for}(|x-\mu| \leq \sqrt{6}\sigma) \\ 0, & \text{for } (|x-\mu| \geq \sqrt{6}\sigma). \end{cases}$$

The corresponding CDF is given by:

$$\Psi(x) = \begin{cases} 0, & \text{for } (x \leq \mu - q). \\ \int_{\mu-q}^{x} \left(\frac{1}{q} - \frac{|x-\mu|}{q^2}\right) dx, & \text{for } (|x-\mu| \leq q) \\ 1, & \text{for } (x \geq \mu + q), \end{cases}$$

and is shown in Figure 3.17.

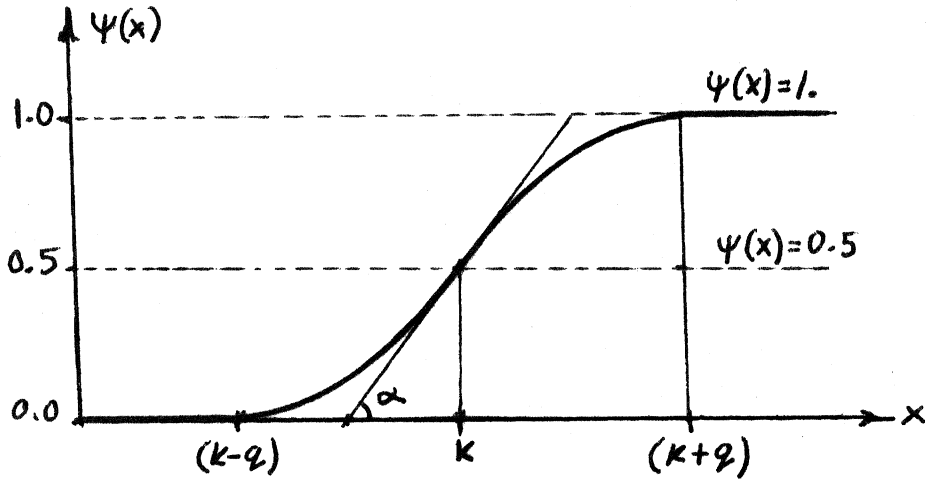


Figure 3.17

The integral in the above equation can be rewritten as:

$$\int_{k-q}^{x} \left(\frac{1}{q} - \frac{|x-\mu|}{q^2}\right) dx = \begin{cases} \displaystyle\int_{\mu-q}^{x} \frac{q + x - \mu}{q^2}\, dx, & \text{for } (x \leq \mu). \\ \displaystyle\int_{\mu-q}^{\mu} \frac{q + x - \mu}{q^2}\, dx + \int_{\mu}^{x} \frac{q - x + \mu}{q^2}\, dx, & \text{for } (x \geq \mu). \end{cases}$$

and we get:

$$\frac{1}{q^2} \int_{\mu-q}^{x} (q + x - \mu)\, dx = \frac{1}{q^2} \left\{\frac{1}{2} [x^2]_{\mu-q}^{x} + (q-\mu)\, [x]_{\mu-q}^{x}\right\}$$

$$= \frac{1}{q^2} \left\{\frac{1}{2}(x^2-\mu^2+2\mu q-q^2) + (q-\mu)\,(x-\mu+q)\right\}$$

$$= \frac{1}{2q^2} \{x^2-2\mu x+\mu^2+2q(x-\mu) + q^2\}$$

$$= \frac{(x-\mu)^2}{2q^2} + \frac{(x-\mu)}{q} + \frac{1}{2} \quad .$$

Similarly,

$$\frac{1}{q^2} \int_{\mu}^{x} (q-x+\mu)\, dx = -\frac{(x-\mu)^2}{2q^2} + \frac{(x-\mu)}{q},$$

and

$$\int_{\mu-q}^{\mu} \frac{q + x - \mu}{q^2}\, dx = \frac{1}{2} \quad .$$

Finally, we can express the CDF, which we are going to denote by T_c, in terms of the mean μ and the standard deviation σ, as follows:

$$T_c(\mu, \sigma; x) = \Psi(x) = \begin{cases} 0, \text{ for } (x \leq \mu - \sqrt{6}\sigma) \\ \dfrac{(x-\mu)^2}{12\sigma^2} + \dfrac{(x-\mu)}{\sqrt{6}\sigma} + \dfrac{1}{2}, \text{ for } (\mu - \sqrt{6}\sigma \leq x \leq u). \\ -\dfrac{(x-\mu)^2}{12\sigma^2} + \dfrac{(x-\mu)}{\sqrt{6}\sigma} + \dfrac{1}{2}, \text{ for } (u \leq x \leq \mu + \sqrt{6}\sigma). \\ 1, \text{ for } (x \geq \mu + \sqrt{6}\sigma). \end{cases}$$

By following the same procedure as in example 3.17, we can compute the probabilities: $P(\mu-\sigma \leq x \leq \mu + \sigma)$, $P(\mu-2\sigma \leq x \leq \mu + 2\sigma)$ and $P(\mu-3\sigma \leq x \leq \mu + 3\sigma)$ as well as the third moment m_3 about zero for the triangular PDF. Again, we give here the results, and the verification is left to the student:

$$P(\mu-\sigma \leq x \leq \mu + \sigma) \doteq 0.66 ,$$

$$P(\mu-2\sigma \leq x \leq \mu + 2\sigma) \doteq 0.97 ,$$

$$P(\mu-3\sigma \leq x \leq \mu + 3\sigma) = 1 \quad \text{and}$$

$$m_3 = \mu^3 + 3\sigma^2\mu .$$

3.3 Random Multivariate

3.3.1 Multivariate, its PDF and CDF

Analogically to the ideas of stochastic function and stochastic variable given in section 3.2.1, we introduce the concept of a multivalued stochastic function

$$X \varepsilon \{U \to R^s\}$$

in the s-dimensional space.

We note that X is a vector function, i.e., X(u) can be written as:

$$X(u) = (x^1(u), x^2(u), \ldots, x^s(u)) \varepsilon R^s, \; u \varepsilon U.$$

The individual components $x^j(u) \varepsilon R$, $j = 1, 2, \ldots, s$ are called components or constituents of X(u). We also note that each component x^j of the stochastic function X can be regarded as a random variable (univariate) of its own. One particular value of x^j may be denoted by x_i^j*) and similarly a particular value of X may be denoted by

$$X_i = (x_i^1, x_i^2, \ldots, x_i^s).$$

Note that a specific value of X is a sequence of real numbers (not a set), or a numerical vector.

The pair (X(u), ϕ(X)), where

$$\phi(X) = \phi(x^1, x^2, \ldots, x^s) \quad \varepsilon \{R^s \to R\} \tag{3-27}$$

is a non-negative, integrable function on R^s is called a random multivariate or simply a multivariate.

---

*The superscripts and subscripts here are found very useful to distinguish between the components x^j, $j = 1, 2, \ldots, s$ of the multivariate X, and the elements x_i^j, $i = 1, 2, \ldots, n_j$ of the univariate (random variable) x^j.

We can speak of a probability of $X \in [X_o, X_1] \subset R^s$, and define it as follows:

$$P(X_o \leq X \leq X_1) = \int_{X_o}^{X_1} \phi(X)\, dX \quad \varepsilon [0, 1] \; . \tag{3-28}$$

Here the integral sign stands for the s-dimensional integration; dX for an s-dimensional differential, i.e. $dX = (dx^1, dx^2, \ldots dx^s)$ and $X_o = (x_o^1, x_o^2, \ldots x_o^s)$, $X_1 = (x_1^1, x_1^2, \ldots x_1^s)$ are assumed to satisfy the following inequalities:

$$x_1^j \geq x_o^j \; , \qquad j = 1, 2, \ldots, s \; .$$

Note that in order to be able to call the function ϕ a PDF, the following condition has to be satisfied:

$$\int_{R^s} \phi\,(X)\, dX = 1 \; . \tag{3-29}$$

A complete analogy to the one-dimension 1 or univariate case (section 3.2.2) is the definition of the multivariate CDF. It is defined as follows:

$$\Psi(X) = \int_{-\infty}^{X} \phi(Y)\, dY \varepsilon \{R^s \rightarrow [0, 1]\} \tag{3-30}$$

where Y is an s-dimensional dummy variable in the integration.

Example 3.19: Consider the univariate PDF shown in Figure 3.12. This bell-shaped PDF is known as the normal or Gaussian PDF (to be discussed later in more details), and is usually denoted by N,

in terms of its μ and σ we have

$$\phi(x) = N(\mu, \sigma; x) \ .$$

Then the multivariate normal PDF in two-dimensional space, i.e. $\phi(X) = \phi(x^1, x^2)$, would appear as illustrated in Figure 3-18.

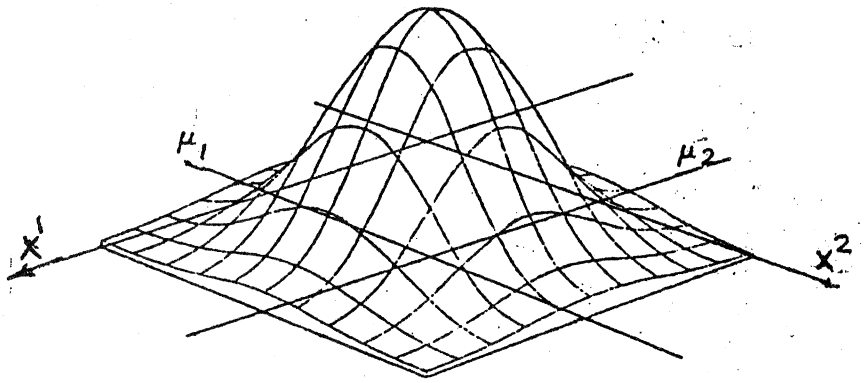


Figure 3-18.

In the two-dimensional space, $\phi(X)$ is called a <u>bivariate PDF,</u> and the bivariate normal PDF illustrated above can be expressed as

$$\phi(X) = N(\mu_1, \mu_2, \sigma_1, \sigma_2; X)$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\tfrac{1}{2}\frac{x^1 - \mu_1}{\sigma_1} - \tfrac{1}{2}\frac{x^2 - \mu_1}{\sigma_2}\right] .$$

$$\left[-\frac{1}{2}\frac{x^1-\mu_1}{\sigma_1} - \frac{1}{2}\frac{x^2-\mu_2}{\sigma_2}\right]$$

3.3.2 Statistical Dependence and Independence

The PDF, ϕ, of the multivariate X may have a special form, namely

$$\phi(X) = \phi_1(x^1)\cdot\phi_2(x^2) \ \ldots \ \phi_s(x^s) = \prod_{j=1}^{s} \phi_j\ (x^j) \quad .$$

In this case, the integral in equation (3-28) can be rewritten as:

$$\int_{X_o}^{X_1} \phi(X)\ dX = \int_{X_o}^{X_1} \prod_{j=1}^{s} \phi_j\ (x^j)\ dx^j$$

$$= \prod_{j=1}^{s} \int_{x_o^j}^{x_1^j} \phi_j\ (x^j)\ dx^j \quad . \qquad (3\text{-}31)$$

Remembering that each component x^j of the multivariate X can be regarded as a univariate, and regarding ϕ_j as the PDFs of the corresponding univariates we can rewrite equation (3-31) as:

$$\prod_{j=1}^{s} \int_{x_o^j}^{x_1^j} \phi_j(x^j)dx^j = \prod_{j=1}^{s} P(x_o^j \leq x^j \leq x_1^j\ ) \quad .$$

Comparing this result with equation (3-28), we get the relationship between the probabilities

$$\boxed{P(X_o \leq X \leq X_1) = \prod_{j=1}^{s} P(x_o^j \leq x^j \leq x_1^j) \quad .} \qquad (3\text{-}32)$$

This relation can be read as follows: "The combined probability of all the components satisfying the condition: $x_o^j \leq x^j \leq x_1^j$, equals to the product of the probabilities for the individual components", and obviously satisfies the definition of the combined probability of

independent events (section 2.3). Hence, the components x^j of such a multivariate X are called statistically independent. The PDF from example 3.19 is statistically independent.

If the PDF of a multivariate cannot be written as a product of the PDF's of its constituents, then these constituents are known as statistically dependent. In this case, the probability $P(X_o \leq X \leq X_1)$ is not equal to the product of the individual probabilities.

It can be shown that for statistically independent components we have

$$\int_R \phi_j (x^j)\, dx^j = 1, \quad j = 1, 2, \ldots, s.$$

3.3.3 Mean and Variance of a Multivariate

The sequence

$$\boxed{\tilde{\mu} = (u_1, u_2, \ldots, u_s) = \tilde{E}^*(X) \quad \varepsilon\, R^s\,,} \tag{3-33}$$

where

$$\mu_j = \int_{R^s} x^j\, \phi(X)\, dX = E^*(x^j) \varepsilon\, R, \quad j = 1, 2, \ldots, s \tag{3-34}$$

is called the mean of the multivariate X. The argument of the operator $\tilde{E}^*$ (i.e. the s-dimensional integral) is X. $\phi(X) = (x^1, x^2, \ldots, x^s) \cdot \phi(x^1, x^2, \ldots, x^s)$. Similarly the variance of the multivariate X is given by

$$\boxed{\begin{aligned} \tilde{\sigma}^2 &= (\sigma_1, \sigma_2^2, \ldots, \sigma_s^2) \\ &= \tilde{E}^*((X-\tilde{\mu})^2) \ \varepsilon\, R^s\,, \end{aligned}} \tag{3-35}$$

where

$$\begin{aligned} \sigma_j^2 &= \int_{R^s} (x^j - \mu_j)^2\, \phi(X)\, dX \\ &= E^*(x^j - \mu_j)^2) \ \varepsilon\, R, \quad j = 1, 2, \ldots, s. \end{aligned} \tag{3-36}$$

Note that we can write again

$$\tilde{E}^*(X-\tilde{\mu})^2 = \tilde{E}^*(X-\tilde{E}^*(X))^2 = \tilde{E}^*(X^2) - \tilde{\mu}^2 \ , \qquad (3\text{-}37)$$

and

$$E^*(x^j-\mu_j)^2 = E^*(x^j-E^*(x^j))^2 = E^*((x^j)^2) - \mu_j^2 \ . \qquad (3\text{-}38)$$

The variance of the multivariate does not express the statistical properties in the multi-dimensional space as fully as the variance of the univariate does in the one-dimensional space. For this reason, we extend the statistical characteristics of the random multivariate further and introduce the so-called variance-covariance matrix (see section 3.3.4).

Let us now turn our attention to what the mean and the variance of a "statistically independent" multivariate look like. For the statistically independent components x^j, $j = 1, 2, \ldots, s$ of a multivariate X, we obtain

$$\begin{aligned}
\mu_j &= \int_{R^s} x^j \prod_{\Pi=1}^{s} \phi_\ell(x^\ell)\, dx^\ell \\
&= \int_{R^s} [x^j \phi_j(x^j) \Big(\prod_{\substack{\ell=1 \\ \ell\neq j}}^{s} \phi_\ell(x^\ell) dx^\ell \Big) dx^j] \qquad (3\text{-}39) \\
&= \int_R x^j \phi_j(x^j) dx^j \cdot \prod_{\substack{\ell=1 \\ \ell\neq j}}^{s} \int_R \phi_\ell(x^\ell) dx^\ell \ .
\end{aligned}$$

Here, according to section 3.3.2, all the integrals in equation (3-39) after the Π-sign are equal to one, and thus we have

$$\mu_j = \int_R x^j \phi_j(x^j)\, dx^j, \quad j = 1, 2, \ldots, s \ . \qquad (3\text{-}40)$$

Similarly,

$$\sigma_j^2 = \int_R (x^j - \mu_j)^2 \, \phi_j(x^j) dx^j, \quad j = 1, 2, .., s. \qquad (3\text{-}41)$$

Thus for the statistically independent X, we can compute the mean and the variance of each component x^j separately, as we have computed μ_1, μ_2, σ_1, σ_2 of the PDF from example 3.19.

3.3.4 Covariance and Variance-Covariance Matrix

Before we start describing the variance-covariance matrix, let us define another statistical quantity needed for this matrix. This quantity is called covariance and it is defined for any two components x^j and x^k of a multivariate X as

$$\text{cov}\,(x^j x^k) = \sigma_{jk} = \int_{R^s} (x^j - \mu_j)\,(x^k - \mu_k)\,\phi(X)\,dX \qquad (3.42)$$

$$= E^*\,((x^j - \mu_j)(x^k - \mu_k)) = \sigma_{kj} \in R;\ k, j = 1,2,..,s$$

We note three things in equation (3-42). First, if $j = k$ we see that the expressions for the covariances become identical with those for the variances, namely:

$$\sigma_{jk} = \sigma_{kj} = \sigma_j^2, \quad \text{for } j = k.$$

Secondly, if the components of the multivariate are statistically independent, the covariances ($j \neq k$) are all equal to zero. To show this, let us write

$$\sigma_{jk} = \int_{R^s} (x^j - \mu_j)(x^k - \mu_k) \prod_{\ell=1}^{s} \phi_\ell(x^\ell)\, dx^\ell$$

$$= \int_R (x^j - \mu_j)\, \phi_j(x^j) dx^j \cdot \int_R (x^k - \mu_k)\, \phi_k(x^k) dx^k$$

$$= \left[\int_R x^j \phi_j\, (x^j) dx^j - \mu_j\right] \left[\int_R x^k \phi_k(x^k)\, dx^k - \mu_k\right]$$

$$= [\mu_j - \mu_j]\, [\mu_k - \mu_k] = 0 \quad .$$

Finally, noting that for a pair of components of a statistically independent multivariate we have

$$\sigma_{jk} = E^*((x^j - \mu_j)(x^k - \mu_k)) = 0 \quad , \tag{3.43}$$

we can write:

$$\sigma_{jk} = E^*(x^j x^k - x^j \mu_k - \mu_j x^k + \mu_j \mu_k)$$

$$= E^*(x^j x^k) - \mu_k E^*(x^j) - \mu_j E^*(x^k) + \mu_j \mu_k$$

$$= E^*(x^j x^k) - \mu_k \mu_j - \mu_j \mu_k + \mu_j \mu_k$$

$$= E^*(x^j x^k) - \mu_j \mu_k = E^*(x^j x^k) - E^*(x^j)\, E^*(x^k) = 0 \ .$$

Hence, for statistically independent components x^j and x^k, we get

$$E^*(x^j x^k) = E^*(x^j) \cdot E^*(x^k) \ , \tag{3-44}$$

or more generally, for r independent components we get

$$E^*\left(\prod_{\ell=1}^{r} x^\ell\right) = \prod_{\ell=1}^{r} E^*(x^\ell) \ . \tag{3-45}$$

Equation (3-45) completes the list of properties of the E* operator stated in section 3.2.3.

As we stated in section 3.3.3, the variance ($\tilde{\sigma}^2$) of a multivariate is not enought to fully characterize the statistical properties of the multivariate on the level of second moments. To get the same amount of statistical information as given by the variance alone (in the univariate case), we have to take into account also the covariances. The variances and covariances can be assembled into one matrix called the <u>variance-covariance matrix</u> or just the <u>covariance matrix</u>.

The variance-covariance matrix of a multivariate X is usually denoted by Σ_X^* and looks as follows:

$$\Sigma_X^* = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \cdots\cdots & \sigma_{1s} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \cdots\cdots & \sigma_{2s} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{s1} & \sigma_{s2} & & \sigma_s^2 \end{bmatrix}, \qquad (3\text{-}46)$$

It is not difficult to see that the variance-covariance matrix can also be written in terms of the mathematical expectation as follows:

$$\boxed{\Sigma_X^* = \tilde{E}^* \left[(X - \tilde{E}^*(X)) \; (X - \tilde{E}^*(X))^T \right]}, \qquad (3\text{-}47)$$

which is the expectation of a dyadic product of two vectors. Note that the superscript T in the above formula stands for the transposition in matrix operation. The proof of equation (3-47) is left to the student.

Note that the variance-covariance matrix is always symmetrical, the diagonal elements are the variances of the components and the off-diagonal elements are the covariances between the different pairs of components. The necessary and sufficient condition for the variance-covariance matrix to be diagonal, i.e. all the covariances to be zeros, is the statistical independence of the multivariate. The variance-covariance matrix is one of the most fundamental quantities used in adjustment calculus. It is positive - definite (with diagonal elements always positive) and the inverse exists if and only if there is no absolute correlation between components.

3.3.5 Random Multisample, its PDF and CDF

Like in the univariate case, we can also define here a quantity η corresponding to the random sample ξ, defined in section 3.1.1 as follows:

$$\eta = \begin{bmatrix} \xi^1 \\ \xi^2 \\ \cdot \\ \xi^s \end{bmatrix} \rightarrow \begin{bmatrix} (\xi_1^1, \quad \xi_2^1, \quad \xi_3^1, \ \ldots\ldots, \quad \xi_{n_1}^1) \ \varepsilon \ R^{n_1} \\ (\xi_1^2, \quad \xi_2^2, \quad \xi_3^2, \ \ldots\ldots, \quad \xi_{n_2}^2) \ \varepsilon \ R^{n_2} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \\ (\xi_1^s, \quad \xi_2^s, \quad \xi_3^s, \ \ldots\ldots, \quad \xi_{n_s}^s) \ \varepsilon \ R^{n_s} \end{bmatrix} \tag{3-48}$$

which is a straightforward generalization of a random sample, and will be called a random multisample. From the above definition, it is obvious that η has s components (constituents), ξ^j, each of which is a random sample on its own. The number of elements n_j in each component ξ^j may or may not be the same.

We can also define the definition set as well as the actual (experimental) PDF and CDF of a multisample in very much the same was as we have done for a random sample. Also, the distribution and cumulative distribution histograms and polygons can be used for two-dimensional multisamples. The development of these concepts, however, is left to the student.

3.3.6 Mean and Variance-Covariance Matrix of a Multisample

The mean of a multisample (3.48) is defined as

$$\boxed{\tilde{M} = (M_1, M_2, \ldots, M_s) = \tilde{E}(\eta) \ \varepsilon \ R^s \ ,} \tag{3-49}$$

where from equation (3-3) we get

$$M_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \xi_i^j = E(\xi^j) \ \varepsilon R, \quad j = 1, 2, \ldots, s \ . \tag{3-50}$$

Here, the operator $\tilde{E}$ is defined as a vector of operators E which is obvious from comparison of (3.49) with (3.50). Similarly,

$$\boxed{\tilde{S}^2 = (S_1^2, S_2^2, S_3^2, \ldots, S_s^2) = \tilde{E}(\eta - \tilde{M})^2 \varepsilon R^s ,} \qquad (3\text{-}51)$$

where from equation (3-6), we get

$$S_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (\xi_i^j - M_j)^2 = E(\xi^j - M_j)^2 \varepsilon R, \; j = 1, 2, .., s. \qquad (3\text{-}52)$$

We can also define the standard deviation $\tilde{S}$ of the multisample η as

$$\tilde{S} = (S_1, S_2, S_3, \ldots, S_s) \qquad (3\text{-}53)$$

Example 3.20: Let us determine the mean $\tilde{M}$, the variance $\tilde{S}^2$ and the standard deviation $\tilde{S}$ of a multisample $\eta = (\xi^1, \xi^2, \xi^3)$, where

$$\xi^1 = (2, 3, 4, 7, 4),$$

$$\xi^2 = (6, 4, 0, 3, 2) \text{ and}$$

$$\xi^3 = (5, 2, 5, 5, 8) .$$

Here we have $n_1 = n_2 = n_3 = 5$. The mean $\tilde{M}$ is given from equation (3-49) as

$$\tilde{M} = (M_1, M_2, M_3) .$$

The members M_j, j = 1, 2, 3 are computed from equation (3-50) as follows:

$$M_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \xi_i^1 = \frac{1}{5} \sum_{i=1}^{5} \xi_i^1$$

$$= \frac{1}{5} (2 + 3 + 4 + 7 + 4) = \frac{20}{5} = 4 ,$$

$$M_2 = \frac{1}{5} (6 + 4 + 0 + 3 + 2) = \frac{15}{5} = 3,$$

$$M_3 = \frac{1}{5} (5 + 2 + 5 + 5 + 8) = \frac{25}{5} = 5 ,$$

and we get

$$\underline{\tilde{M} = (4, 3, 5)} .$$

The <u>variance $\tilde{S}^2$</u> is given from equation (3-51) as

$$\tilde{S}^2 = (S_1^2, S_2^2, S_3^2) .$$

The members S_j^2, j = 1, 2, 3 are computed from equation (3-52) as follows:

$$S_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (\xi_i^1 - M_1)^2 = \frac{1}{5} \sum_{i=1}^{5} (\xi_i^1 - 4)^2$$

$$= \frac{1}{5} [(-2)^2 + (-1)^2 + (0)^2 + (3)^2 + (0)^2]$$

$$= \frac{1}{5} [4 + 1 + 0 + 9 + 0] = \frac{14}{5} = 2.8 ,$$

$$S_2^2 = \frac{1}{5}\,[(3)^2 + (1)^2 + (-3)^2 + (0)^2 + (-1)^2]$$

$$= \frac{1}{5}\,[9 + 1 + 9 + 0 + 1] = \frac{20}{5} = 4.0$$

$$S_3^2 = \frac{1}{5}\,[(0)^2 + (-3)^2 + (0)^2 + (0)^2 + (3)^2]$$

$$= \frac{1}{5}\,[0 + 9 + 0 + 0 + 9] = \frac{18}{5} = 3.6\ ,$$

and we get

$$\underline{\tilde{S}^2 = (2.8,\ 4.0,\ 3.6)\ .}$$

Taking the square root of the individual members S_j^2, $j = 1, 2, 3$, we obtain the standard deviation $\tilde{S}$ as

$$\tilde{S} = (S_1,\ S_2,\ S_3) = (1.67,\ 2.0,\ 1.9)\ .$$

If the jth and kth components of a multisample have the same number of elements, say n, we can write the covariance S_{jk} between these two components ξ^j and ξ^k as:

$$\boxed{S_{jk} = \frac{1}{n} \sum_{i=1}^{n} [(\xi^j - M_j)\ (\xi^k - M_k)]\ ,} \qquad (3\text{-}54)$$

which can be rewritten as:

$$S_{jk} = E((\xi^j - M_j)(\xi^k - M_k)) = S_{kj} \in R,\quad j, k = 1, 2, .., s\ .$$

Note that the covariance S_{jk}, as defined above, depends on the ordering of the elements in both components ξ^j and ξ^k, whereas the means M_j and M_k and the variances S_j^2 and S_k^2 do not. Therefore, to obtain a meaningful covariance S_{jk}, each of the components ξ^j and ξ^k should be in the same order as it was acquired. This can be visualized from the following example. Assuming that the elements of ξ^j are observations of one vertical angle, and the elements of ξ^k are the corresponding times of the observations. Clearly, to study the relationship (covariance) between the observation time and the value of the observed vertical angle, the matched pairs must be respected.

Example 3.21: Let us determine the covariances between the different pairs of components of the multisample η given in example 3.20. The covariances S_{jk} are computed from equation (3-54) as follows:

$$S_{12} = S_{21} = \frac{1}{5} \sum_{i=1}^{5} [(\xi_i^1 - 4)\ (\xi_i^2 - 3)]$$

$$= \frac{1}{5} [(-2)(3) + (-1)(1) + (0)(-3) +$$

$$+ (3)(0) + (0)(-1)]$$

$$= \frac{1}{5} [-6-1+0+0+0] = \frac{-7}{5} = \underline{-1.4},$$

$$S_{13} = S_{31} = \frac{1}{5}[(-2)(0) + (-1)(-3) + (0)(0) +$$

$$+ (3)(0) + (0)(3)]$$

$$= \frac{1}{5} [0+3+0+0+0] = \frac{3}{5} = \underline{0.6} \text{ and}$$

$$S_{23} = S_{32} = \frac{1}{5}[(3)(0) + (1)(-3) + (-3)(0) +$$

$$+ (0)(0) + (-1)(3)]$$

$$= \frac{1}{5} [0 - 3 + 0 + 0 - 3] = \frac{-6}{5} = \underline{-1.2}.$$

Finally, we can assemble the variance covariance matrix $\Sigma\eta$ of the multisample η:

$$\Sigma_\eta = \begin{bmatrix} S_1^2 & S_{12} & S_{13} & \cdots & S_{1s} \\ S_{21} & S_2^2 & S_{23} & \cdots & S_{2s} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ S_{s1} & S_{s2} & & & S_s^2 \end{bmatrix} . \qquad (3\text{-}54)$$

Having defined the mean and the variance-covariance matrix of a multisample let us stop and reflect for a while. We have stated in 3.3.3 that the expansion from one to s dimensions defied a straightforward generalisation of one dimensional variance. We had to introduce the variance-covariance matrix to describe the statistical properties

of a multivariate on the second moments level. Turning to the relationship sample - univariate we discover that this is not paralleled in the multi-dimensional case either. While formulae for the mean and the variance of a sample and a univariate were equivalent, those for a multisample and a multivariate are not. While equivalent formulae to (3-34), (3-35) and (3-42) can be devised for the multisample, the ones used mostly in practice ((3-49), (3-51) and (3-54)) correspond really to (3-40), (3-41), and (3-43) valid only for statistically independent multivariate.

This, together with the difficulty with the computation of multisample covariances, i.e., the necessity to have the same number of elements in any two components, leads often in practice to the adoption of an assumed variance-covariance matrix. Decisions connected with the determination of the multisample variance-covariance matrix are among the trickiest in adjustment calculus.

Example 3.22. Let us determine the variance-covariance matrix of the multisample η introduced in example 3.20. In this case, we have the variances computed in example 3.20, the results were:

$$S_1^2 = 2.8, \quad S_2^2 = 4.0 \quad \text{and} \quad S_3^2 = 3.6.$$

Also, we have the covariances computed in example 3.21, the results were:

$$S_{12} = S_{21} = -1.4, \quad S_{13} = S_{31} = 0.6 \text{ and}$$

$$S_{23} = S_{32} = -1.2.$$

Therefore, the required variance-covariance matrix will be:

$$\Sigma_\eta = \begin{bmatrix} 2.8 & -1.4 & 0.6 \\ -1.4 & 4.0 & -1.2 \\ 0.6 & -1.2 & 3.6 \end{bmatrix} .$$

3.3.7 Correlation

Although the covariances of a multisample do not play the same role as the covariances of a multivariate, they still can serve as a certain measure of statistical dependence. We say that they show the degree of correlation between the appropriate pairs of components.

The degree of correlation as a measure of statistical dependence, may, of course, vary. We can see that the covariance $S_{jk} \in R$ may attain any value. Hence it is not a very useful measure because we cannot predetermine the value of the covariance corresponding to the maximum or complete correlation. For this reason, we use another measure, the correlation coefficient, which is usually denoted by ρ, and is defined as

$$\boxed{\rho_{jk} = S_{jk} / (S_j \cdot S_k), \quad S_j, S_k \neq 0 \quad .} \qquad (3\text{-}57)$$

It can be shown that ρ_{jk} varies from -1 to+1.

Based on the use of the correlation coefficient is the correlation calculus, a separate branch of statistics. It will suffice here to say that we call two components ξ^j and ξ^k of a multisample η:

(i) totally uncorrelated, if $\rho_{jk} = 0$,

(ii) correlated, if $|\rho_{jk}| < 1$,

(iii) totally positively correlated, if $\rho_{jk} = 1$,

(iv) totally negatively correlated, if $\rho_{jk} = -1$.

Note that for the multivariate, the expression for ρ_{jk} is written completely analogous to equation (3-57).

Example 3.23: Let us discuss the degree of correlation between the different pairs of components of the multisample η which is used in examples 3.20 to 3.22 inclusive, and whose variance-covariance matrix is given in example 3.22.

The correlation coefficients ρ_{jk} are computed from equation (3-57) as follows:

$$\rho_{12} = \frac{S_{12}}{S_1 \cdot S_2} = \frac{-1.4}{1.67 \cdot 2} = -0.42$$

$$\rho_{13} = \frac{0.6}{1.67 \cdot 1.9} = 0.19 ,$$

$$\rho_{23} = \frac{-1.2}{2 \cdot 1.9} = -0.31 .$$

Note that:

$$\rho_{12} = \rho_{21}, \quad \rho_{13} = \rho_{31} \text{ and } \rho_{23} = \rho_{32} .$$

Since

$$|\rho_{jk}| < 1, \quad j, k = 1, 2, 3, \quad j \neq k,$$

thus the components ξ^1, ξ^2 and ξ^3 of the given multisample η are all correlated.

Example 3.24: Let us discuss the degree of correlation between the components ξ^1 and ξ^2, and between ξ^1 and ξ^3 of the multisample $\eta = (\xi^1, \xi^2, \xi^3)$, where:

$$\xi^1 = (2, 1, 3, 5, 4),$$
$$\xi^2 = (4, 2, 6, 10, 8),$$
$$\xi^3 = (-4, -2, -6, -10, -8).$$

By computing the means and variances of ξ^j, $j = 1, 2, 3$ similarly to example 3.20, and the covariances S_{12} and S_{13} similarly to example 3.21, we get the following results:

$$M_1 = 3, \quad M_2 = 6 \quad \text{and} \quad M_3 = -6,$$

$$S_1^2 = 2, \quad S_2^2 = 8 \quad \text{and} \quad S_3^2 = 8,$$

$$S_1 = \sqrt{2}, \; S_2 = S_3 = 2\sqrt{2},$$

$$S_{12} = 4 \quad \text{and} \quad S_{13} = -4.$$

Hence

$$\rho_{12} = \frac{S_{12}}{S_1 \cdot S_2} = \frac{4}{\sqrt{2} \cdot 2\sqrt{2}} = +1,$$

which means that ξ^1 and ξ^2 are totally positively correlated, and

$$\rho_{13} = \frac{S_{13}}{S_1 \cdot S_3} = \frac{-4}{\sqrt{2} \cdot 2\sqrt{2}} = -1,$$

which means that ξ^1 and ξ^3 are totally negatively correlated.

At this point it is worthwhile mentioning that the computations of the means, variances, covariances and correlation coefficients of the constituents of a multisample are always preferably performed in a tabular form for easier checking. The following table is an example of such an arrangment using the two constituents ξ^1 and ξ^2 of the multisample introduced in example 3.20.

	ξ^1			ξ^2			$(\xi_i^1-M_1)\cdot(\xi_i^2-M_2)$
	ξ_i^1	$(\xi_i^1-M_1)$	$(\xi_i^1-M_1)^2$	ξ_i^2	$(\xi_i^2-M_2)$	$(\xi_i^2-M_2)^2$	
	2	-2	4	6	3	9	-6
	3	-1	1	4	1	1	-1
	4	0	0	0	-3	9	0
	7	3	9	3	0	0	0
	4	0	0	2	-1	1	0
Σ	20		14	15		20	-7

$$M_1 = \frac{1}{5}(20) = 4, \qquad M_2 = \frac{1}{5}(15) = 3 ,$$

$$S_1^2 = \frac{1}{5}(14) = 2.8, \qquad S_2^2 = \frac{1}{5}(20) = 4 ,$$

$$S_1 = \sqrt{2.8} = 1.67 , \qquad S_2 = \sqrt{4} = 2 ,$$

$$S_{12} = \frac{1}{5}(-7) = -1.4,$$

and

$$\rho_{12} = \frac{-1.4}{1.67 \cdot 2} = - 0.42 .$$

61, 70, 102, 107, 113, 114, 117, 119, 120, 126, 126, 129, 129, 132, 137, 139, 139, 139, 142, 143, 146, 146, 147, 147, 148, 149, 149, 150, 150, 153, 153, 156, 157, 158, 159, 159, 159, 159, 162, 162, 164, 166, 166, 166, 167, 169, 169, 169, 169, 170, 170, 171, 172, 172, 172, 173, 173, 175, 175, 176, 176, 176, 177, 177, 178, 179, 180, 180, 181, 181, 181, 182, 183, 184, 184, 185, 186, 187, 188, 188, 190, 192, 192, 193, 194, 194, 194, 195, 195, 195, 196, 197, 198, 198, 200, 201, 201, 201, 202, 202, 203, 204, 206, 206, 209, 209, 209, 214, 216, 219, 219, 219, 221, 222, 223, 227, 233, 234, 236, 237, 246, 247, 254, 262, 270

Required: (i) Classify this sample according to your own choice, and then draw its: distribution histogram, distribution polygon, cumulative histogram, cumulative polygon.

(ii) Determine the mean, standard deviation, median and range of the sample; then plot these quantities on your histograms and polygons.

(iii) Determine the probability of the height being in between 121 and 174 cms, by using your distribution histogram, your distribution polygon, the cumulative histogram, the cumulative polygon, the actual sample. Then compare the results.

(3) Verify the results given in Example 3.18 for the mean, the variance and the third moment about zero of the triangular PDF.

3.4 Exercise 3

(1) The following table gives the weights as recorded to the nearest pound for a random sample of 20 high-school students:

138	150,	146,	158,	150
146	164	138	164	164
150	146	158	173	150
158	130	146	150	164

Required: (i) Compute the mean, the standard deviation, the median and the range of this random sample using both the original sample and its definition set.

(ii) Compute the experimental probabilities of the individual elements and then construct the corresponding discrete PDF and CDF of the sample.

(iii) Compute the probability that the weight of a high-school student is less than or equal to 150 pounds.

(iv) Compute the probability of the student weight to be in the interval [158, 173].

(2) The following table gives the observed heights in cm of a random sample of 125 nine years old pine trees.

(4) Verify the results given in Examples 3.17 and 3.18 for the probabilities $P(\mu-\sigma \leq x \leq \mu+\sigma)$, $P(\mu-2\sigma \leq x \leq \mu + 2\sigma)$ and $P(\mu-3\sigma \leq x \leq \mu + 3\sigma)$using the rectangular and the triangular CDF s respectively (rather than the corresponding PDF s.)

(5) Let x be a random variable whose PDF is given by:

$$\phi(x) = \begin{cases} h, \text{ for } (-3 \leq x \leq 7) \\ 0, \text{ everywhere else.} \end{cases}$$

Required: (i) Determine h.

(ii) Compute the mean and the standard deviation of x.

(iii) Construct the CDF of x.

(iv) Use both the PDF and CDF to determine the following probabilities: $P(x \leq 1.5)$,

$P(x \geq 2.5)$,

$P(-1 \leq x \leq 4)$,

$P(\mu-2\sigma \leq x \leq \mu+2\sigma)$.

(v) Compute the 3-rd and 4-th moments of the PDF about zero.

(6) Let x be a random variable having the following PDF:

$$\phi(x) = \begin{cases} k \cdot x \text{ , for } (0 \leq x \leq 2) \\ 0 \text{ , everywhere else.} \end{cases}$$

Required: (i) Determine the mean, the variance and the standard deviation of x.

(ii) Compute the probability $P(1 \leq x \leq 1.5)$.

(7) Let x be a random variable whose PDF is given as:

$$\phi(x) = \begin{cases} k + \frac{1}{50}x - \frac{3}{50}, & \text{for } (3 \leq x \leq 8) \\ k - \frac{1}{50}x + \frac{13}{50}, & \text{for } (8 \leq x \leq 13) \\ 0, & \text{everywhere else.} \end{cases}$$

Required: (i) Determine the mean and the standard deviation of x.

(ii) Compute the probabilities: $P(5.5 \leq x \leq 10.5)$, $P(x \leq 9)$, $P(x \geq 7)$, $P(\mu - \sigma \leq x \leq \mu + \sigma)$.

(8) Given a multisample $\eta = (\xi^1, \xi^2, \xi^3)$, where $\xi^1 = (4.2, 3.7, 4.1)$, $\xi^2 = (26.7, 26.3, 26.6)$, and $\xi^3 = (-17.5, -17.0, -18.0)$.

Required: (i) Compute the mean of η .

(ii) Compute the variance-covariance matrix of η.

(iii) Compute all the correlation coefficients between the different pairs of components of η.

(9) Given a bivariate $X = (x^1, x^2)$ with PDF

$$\phi(X) = \begin{cases} \frac{|x^1-q|}{s^2 t\ 12\sqrt{3}} + \frac{1}{st\ 6\sqrt{2}}, & \text{for } (|x^1-q| \leq s\sqrt{6}, \text{ and } |x^2-r| \leq t\sqrt{3}) \\ 0, & \text{everywhere else,} \end{cases}$$

where q, r are some real numbers and s, t are some positive real numbers.

Required: (i) Compute the mean of X.

(ii) Compute the variance-covariance matrix of X.

4. FUNDAMENTALS OF THE THEORY OF ERRORS

4.1 Basic Definitions

In practice we work with observations which are nothing else but numerical representation of some physical quantities, e.g. lengths, angles, weights, etc. These observations are obtained through measurements of some kind by comparison to predefined standards. In many cases we obtain several observations for the same physical quantity, which are usually postulated to represent this quantity.

There is a different school of thoughts claiming that no quantity can be measured twice. They say that if a quantity is measured for the second time, it becomes a different quantity. Philosophically, the two approaches are very different, however, in practice they coincide. They vary in assuming different things (hypotheses), but they lead to the same results.

The observations representing the same quantity may or may not have some spread or dispersion (by spread we mean that not all the observations are identical). For instance, when we measure the length of the side of a rectangle using a graduated ruler, we will have two possibilities (see Figure 4.1a, b).

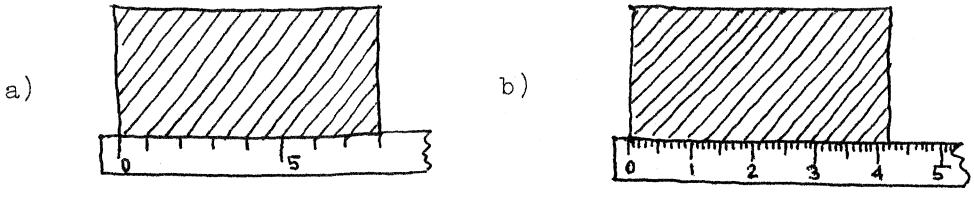


Figure 4.1

First, if the length of that side is exactly equivalent to an integer number of graduations (divisions) on the ruler, the measurement of it will not produce any spread. This is simply because the beginning of the side will be at a graduation line of the ruler, and at the same time the end of the side will be at another graduation line, and hence we get always the same result. On the other hand, if the end of the side is located between two division lines on the ruler, there will be a fraction of the smallest division on the ruler to be estimated. The estimates (observations) will differ, say due to different observers, and hence we shall get a spread.

Usually, the spread and its presence depend on many other things like: the design of the experiment, measuring equipment, precision required, atmospheric conditions, etc. If we know the causes that influence the spread, we can try to account for them in one way or the other. In other words, we will apply certain corrections to eliminate such unwanted influences which are usually called systematic errors. Examples of systematic errors are numerous like: variation of the length of a tape with temperature, variation of atmospheric conditions with time, etc.

In practice, this is possible if we can express such corrections mathematically as functions of some measurable physical quantities. In some cases, the systematic errors remain constant in both magnitude and sign during the time of observations, e.g. most of the instrumental systematic errors. In such cases, we can eliminate these systematic errors by following certain techniques in making the observations. For example, the error in the rod reading due to the inclination of the line

of sight of the level, with respect to the bubble axis, can be eliminated by taking the backsight and the foresight at equal distances from the level.

Further, we shall assume that there are no blunders (mistakes) in the observations. These blunders are usually gross errors due to the carelessness of the observer and/or the recorder. The elimination of blunders has to be carried out before starting to work with the observations. The ways for intercepting blunders are numerous and are as different as the experiments may be. We are not going to venture into this here.

4.2 Random (Accidental) Errors

Even after eliminating the blunders and applying the appropriate corrections to eliminate the systematic errors, the observations representing a single physical quantity usually still have a remaining spread, i.e. are still not identical, and we begin to blame some unknwon or partly unknown reasons for it. Such remaining spread is practically inevitable and we say that the observations contain random or accidental errors.

The above statement should be understood as follows: given a finite sequence L of observations of the same physical quantity ℓ', i.e.

$$L = (\ell_1, \ell_2, \ldots, \ell_n),$$

we assume that the individual elements ℓ_i, $i = 1, 2, \ldots, n$ represent the same quantity ℓ', where ℓ' is the unknown value, and can be written as:

$$\ell_i = \ell' + \varepsilon_i, \quad i = 1, 2, \ldots, n . \qquad (4\text{-}2)$$

The quantities ε's are the so-called random (accidental) errors*.

The sequence

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n) , \qquad (4\text{-}3)$$

(or the sequence L, equation (4-1), for this matter) is declared a random sample as defined earlier in section 3.1.1. This random sample has a parent random variable, as defined in section 3.1.2.

It should be noted that the term "random error" is used rather freely in practice.

4.3 Gaussian PDF, Gauss Law of Errors

The histograms (polygons) of the random samples representing observations encountered in practice generally show a tendency towards being bell-shaped, as shown in Figure 4.2 a,b.

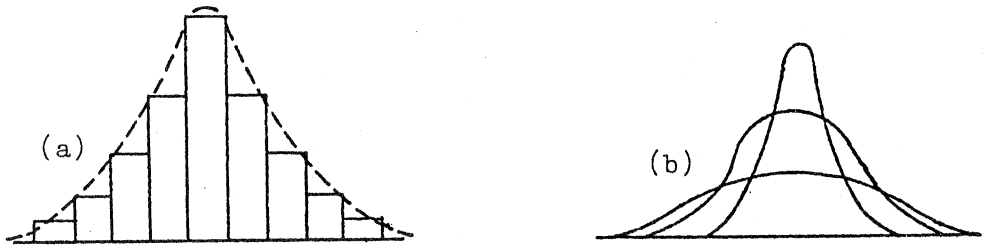


Figure 4.2

---

* It may happen, and as a matter of fact often does happen, that we are able to spot some dependence of ε (for whatever this means) on one or more parameters, e.g. temperature, pressure, time, etc., that had not been suspected and eliminated before. Then we say that the ε's change systematically or predictably with the parameter in question, or we say that there is a correlation between the ε's and the parameter. Here, we may say that the observations still contain systematic errors. In such a case we may try to eliminate them again, after establishing the law governing their behaviour.

Various people throughout the history have thus tried to explain this phenomenon and establish a theory describing it. The commonly accepted explanation is due to Gauss and Laplace independently. This explanation leads to the derivation of the well known model - the Gaussian PDF. The assumptions, due to Hagen, necessary to be taken into account, along with the derivation of the law, due to de Moivre, are given in Appendix I. Here we state only the result.

The Gaussian PDF, $G(C;\varepsilon)$ is found to be (equation (I-11), Appendix I):

$$\boxed{G(C;\ \varepsilon) = \sqrt{\frac{2}{C\pi}}\ \exp\ (-2\varepsilon^2/C),} \qquad (4\text{-}4)$$

where its argument ε is the random error, i.e. a special type of random variable with mean equal to zero, and C is the only parameter of the distribution. The Gaussian PDF is continuous and is shown in Figure 4.3.

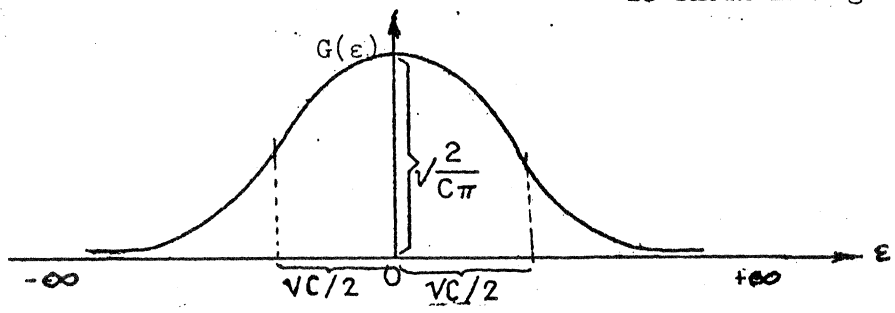


Figure 4.3.

From the above Figure we note the following characteristics of the Gaussian PDF,

(i) G is symmetrical around 0.

(ii) The maximum ordinate of G is at $\varepsilon = 0$, and equals $\sqrt{(2/C\pi)}$, which varies with the parameter C, see Figure 4.2b.

(iii) G approaches the ε axis asymptotically as ε goes to $\pm\ \infty$.

(iv) G has two points of inflextion at $\varepsilon = \pm \sqrt{C/2}$.

The shape of G reflects what is known as the "Gauss law of a large sample of errors", which states that:

(i) smaller errors are more probable than the larger errors,

(ii) positive and negative errors have the same probability.*

Note that since G is a PDF it satisfies the following condition:

$$\int_{-\infty}^{\infty} G(C;\varepsilon)\, d\varepsilon = \sqrt{\frac{2}{C\pi}} \int_{-\infty}^{\infty} \exp(-2\varepsilon^2/C)\, d\varepsilon = 1 \qquad (4.5)$$

4.4 Mean and Variance of the Gaussian PDF

Since G is symmetrical around zero, it is obvious that its mean μ_ε equals zero (see section 3.2.5).

The variance σ_ε^2 of G is again obtained from

$$\sigma_\varepsilon^2 = E^*(\varepsilon - \mu_\varepsilon)^2 = \int_{-\infty}^{\infty} \varepsilon^2 G(C;\varepsilon)\, d\varepsilon$$

$$= \sqrt{\frac{2}{C\pi}} \int_{-\infty}^{\infty} \varepsilon^2 \exp(-2\varepsilon^2/C)\, d\varepsilon. \qquad (4.6)$$

Recalling that

$$\int_0^{\infty} t^2 \exp(-a^2t^2)\, dt = \frac{\sqrt{\pi}}{4a^3}, \quad (a > 0), \qquad (4.7)$$

we get from equations (4.6) and (4.7)

---

* The same result can be obtained using slightly weaker (more general) assumptions through the "central limit theorem".

$$\int_{-\infty}^{\infty} \varepsilon^2 \exp(-2\varepsilon^2/C)d\varepsilon = 2\int_{0}^{\infty} \varepsilon^2 \exp(-a^2\varepsilon^2)d\varepsilon$$

$$= \frac{\sqrt{\pi}}{2a^3} ,$$

where

$$a = \sqrt{\frac{2}{C}} .$$

Hence,

$$\sigma_\varepsilon^2 = \sqrt{\frac{2}{C\pi}} \cdot \frac{\sqrt{\pi}}{2} \left[\sqrt{\frac{C}{2}}\right]^3 = \frac{\sqrt{2} \cdot C\sqrt{C}}{2\sqrt{C} \cdot 2\sqrt{2}} = \frac{C}{4} ,$$

and we get

$$C = 4\sigma_\varepsilon^2 . \qquad (4\text{-}8)$$

Consequently, the variance σ_ε^2, or rather the standard deviation σ_ε, can be considered the only parameter of G. Substituting equation (4-8) into equation (4-4) we get:

$$\boxed{G(\sigma_\varepsilon;\ \varepsilon) = \frac{1}{\sigma_\varepsilon \sqrt{(2\pi)}} \exp\left(-\varepsilon^2/(2\sigma_\varepsilon^2)\right) .} \qquad (4\text{-}9)$$

Note from equation (4-8) that $\sigma_\varepsilon = \sqrt{C}/2$, which equals to the abscissas of the two points of inflextion of G.

Example 4.1. Let us compute, approximately, the probability $P(-\sigma_\varepsilon \leq \varepsilon \leq \sigma_\varepsilon)$ assuming that ε has a Gaussian PDF. We first expand the function $\exp(-\varepsilon^2/2\sigma_\varepsilon^2)$ to be able to integrate equation (4-9). Recall that:

$$\exp(y) = e^y = 1+y + \frac{y^2}{2!} + \frac{y^3}{3!} + \ldots .$$

Hence

$$\exp\left(-\varepsilon^2/(2\sigma_\varepsilon^2)\right) = 1 - \frac{\varepsilon^2}{2\sigma_\varepsilon^2} + \frac{\varepsilon^4}{8\sigma_\varepsilon^4} - \frac{\varepsilon^6}{48\sigma_\varepsilon^6} + \cdots$$

and

$$P(-\sigma_\varepsilon \leq \varepsilon \leq \sigma_\varepsilon) = \int_{-\sigma_\varepsilon}^{\sigma_\varepsilon} G(\sigma_\varepsilon; \varepsilon)\,d\varepsilon$$

$$= \frac{1}{\sigma_\varepsilon\sqrt{(2\pi)}} \int_{-\sigma_\varepsilon}^{\sigma_\varepsilon} \exp\left(-\varepsilon^2/(2\sigma_\varepsilon^2)\right) d\varepsilon$$

$$= \frac{1}{\sigma_\varepsilon\sqrt{(2\pi)}} \left[\int_{-\sigma_\varepsilon}^{\sigma_\varepsilon} d\varepsilon - \frac{1}{2\sigma_\varepsilon^2} \int_{-\sigma_\varepsilon}^{\sigma_\varepsilon} \varepsilon^2\,d\varepsilon + \right.$$

$$\left. + \frac{1}{8\sigma_\varepsilon^4} \int_{-\sigma_\varepsilon}^{\sigma_\varepsilon} \varepsilon^4\,d\varepsilon \ldots.\right]$$

$$= \frac{1}{\sigma_\varepsilon\sqrt{(2\pi)}} \left[2\sigma_\varepsilon - \frac{1}{2\sigma_\varepsilon^2} \cdot \frac{2\sigma_\varepsilon^3}{3} + \frac{1}{8\sigma_\varepsilon^4} \cdot \frac{2\sigma_\varepsilon^5}{5} - \right.$$

$$\left. - \frac{1}{48\sigma_\varepsilon^6} \cdot \frac{2\sigma_\varepsilon^7}{7} + \cdots \right]$$

$$= \frac{2\sigma_\varepsilon}{\sigma_\varepsilon\sqrt{(2\pi)}} [1. - 0.167 + 0.025 - 0.003]$$

$$= \frac{2}{\sqrt{(2\pi)}} [0.855] \doteq 0.683$$

Thus:

$$P(-\sigma_\varepsilon \leq \varepsilon \leq \sigma_\varepsilon) \doteq \underline{0.683}\,.$$

By following the same procedure, we can find that:

$$P(-2\sigma_\varepsilon \leq \varepsilon \leq 2\sigma_\varepsilon) \doteq \underline{0.954},$$

$$P(-3\sigma_\varepsilon \leq \varepsilon \leq 3\sigma_\varepsilon) \doteq \underline{0.997}.$$

4.5 Generalized or Normal Gaussian PDF

The Gaussian PDF (equation (4.9)) can be generalized to have an arbitrary mean μ_y. This is achieved by the transformation

$$y = \varepsilon + \mu_y \ , \tag{4-10}$$

in equation (4-9), where y is the argument of the new PDF - the generalized Gaussian. Such generalized Gaussian PDF is usually called normal PDF and is denoted by N, where:

$$N(\mu_y, \sigma_y; y) = \frac{1}{\sigma_y \sqrt{(2\pi)}} \exp\left(-\frac{(y-\mu_y)^2}{2\sigma_y^2}\right) . \tag{4-11}$$

The name "normal" reflects the trust which people have, or used to have, in the power of the Gaussian law (also called the "normal law") which is mentioned in section 4.3 . If the errors behave according to this law and display a histogram conforming to the normal PDF, they are normal. On the other hand, if they do not, they are regarded as abnormal and strange things are suspected to have happened.

The normal PDF contains only two parameters - the mean μ_y and the standard deviation σ_y. Hence, it is well suited for computations.

Note here that the family of $G(\sigma_\varepsilon; \varepsilon)$ is a subset of the family of $N(\mu_y, \sigma_y; y)$. Also note that the following condition has to be satisfied by N:

$$\int_{-\infty}^{\infty} N(\mu_y, \sigma_y; y)dy = \frac{1}{\sigma_y \sqrt{(2\pi)}} \int_{-\infty}^{\infty} \exp\left(-\frac{(y-\mu_y)^2}{2\sigma_y^2}\right) dy = 1.$$

The formula for the normal CDF corresponding to N is given as:

$$\Psi_N(y) = \frac{1}{\sigma_y \sqrt{(2\pi)}} \int_{-\infty}^{y} \exp\left(-\frac{(x-\mu_y)^2}{2\sigma_y^2}\right) dx, \tag{4-12}$$

where x is a dummy variable in the integration.

For the generalized (normal) Gaussian PDF, it can be again shown that:

$$P(\mu_y - \sigma_y \leq y \leq \mu_y + \sigma_y) \doteq 0.683 ,$$

$$P(\mu_y - 2\sigma_y \leq y \leq \mu_y + 2\sigma_y) \doteq 0.954$$

and

$$P(\mu_y - 3\sigma_y \leq y \leq \mu_y + 3\sigma_y) \doteq 0.997 .$$

(Compare the values to the corresponding results of the triangular PDF in example 3.18).

4.6 Standard Normal PDF

The outcome t of the following linear transformation

$$t = \frac{x - \mu_x}{\sigma_x} \tag{4-13}$$

is often called the standardized random variable, where x is a random variable with mean μ_x and standard deviation σ_x. Note that the above standardization process does not require any specific distribution for x.

The transformation of the normal variable y (equation (4-10)) to a standardized normal variable $t = \frac{y-\mu_y}{\sigma_y}$ results in a new PDF

$$\boxed{N(\mu_t, \sigma_t; t) = \frac{1}{\sqrt{(2\pi)}} \exp(-t^2/2) = N(0, 1; t) = N(t),} \tag{4-14}$$

whose mean μ_t is zero and whose standard deviation σ_t is one. This PDF is called the standard normal PDF, a particular member of the family of all normal distributions.

Since both the parameters $\mu_t = 0$ and $\sigma_t = 1$ are determined once for all, the standard normal PDF is particularly suitable for tabulation due to the fact that it is a function of t only. An example of such tabulation is given in Appendix II-A, which gives the ordinates of the standard normal PDF for different values of t. Note again that

$$\int_{-\infty}^{\infty} N(t)dt = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt = 1.$$

The CDF corresponding to N (t) is given by

$$\Psi_N(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{t} \exp\left(-\frac{x^2}{2}\right) dx, \qquad (4\text{-}15)$$

or

$$\boxed{\Psi_N(t) = \frac{1}{2} + \frac{1}{\sqrt{(2\pi)}} \int_{o}^{t} \exp\left(-\frac{x^2}{2}\right)dx,} \qquad (4\text{-}16)$$

where x is a dummy variable in the integration. Again, the CDF of the standard normal PDF is tabulated to facilitate its use in probability computations. Appendix II-B is an example of tabulated $\Psi_N(t)$ using equation (4-15), which gives the accumulated areas (probabilities) under the standard normal PDF for different positive* values of t. Appendix II-C contains a similar table, but it gives the values of the second term in equation (4-16) only, for different values of t. Hence, care must be taken when using different tables for computations.

---

* For negative values of the argument t the cumulative probability $P(t \leq -t_o) = \Psi_N(-t_o)$ is computed from $\Psi_N(t_o)$ through the condition:

$$\Psi_N(-t_o) = 1 - \Psi_N(t_o) .$$

The second term in equation (4-16) is usually known as $\frac{1}{2}$ erf (t), i.e.

$$\Psi_N(t) = \frac{1}{2} + \frac{1}{2}\ \text{erf}\ (t), \qquad (4\text{-}17)$$

where, erf (t) is known as the error function, and is obviously given by

$$\boxed{\text{erf}\ (t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp\left(-\frac{x^2}{2}\right)dx\ .} \qquad (4\text{-}18)$$

This erf (t) is also tabulated*.

In order to be able to use the tables of the standard normal PDF and CDF for computations concerning a given normal random variable x, we first have to standardize x, i.e. to transform x to t using equation (4-13), then enter these tables with t. Thus, if we want, for instance, to determine the probability $P(x \leq x_o)$ we have to write:

$$P(x \leq x_o) = P\left(\frac{x-\mu_x}{\sigma_x} \leq \frac{x_o-\mu_x}{\sigma_x}\right)\ . \qquad (4\text{-}19)$$

This is identical to the probability $P(t \leq t_o)$ that can be obtained from the standard normal tables.

Example 4.2

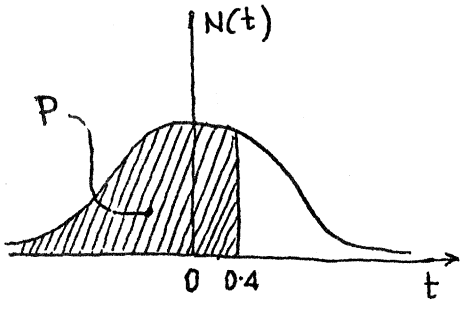


Figure 4.4 - i

Suppose that the height h of a student is a normally distributed random variable with mean $\mu_h = 66$ inches and standard deviation $\sigma_h = 5$ inches. Find the approximate number K out of 1000 students h inches tall:

(i) $h \leq 68$ inches (Figure 4.4-i);

(ii) $h \leq 61$ inches (Figure 4.4-ii);

---

* In most of the computer languages, this error function, erf (t) is a built-in function. Hence, it can be called as any library subroutine and evaluated more precisely than using the corresponding tables.

(iii) $h \geq 74.6$ inches (Figure 4.4-iii);

(iv) $[64.3 \leq h \leq 70]$ inches (Figure 4.4-iv).

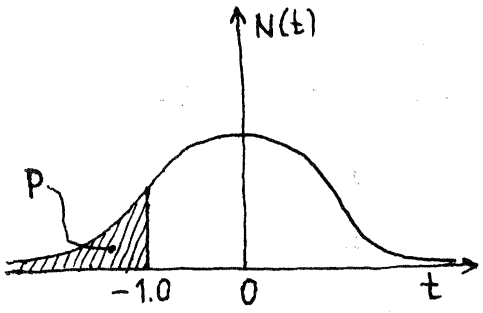


Figure 4.4-ii

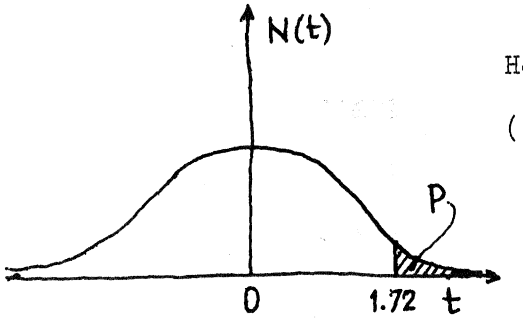


Figure 4.4-iii

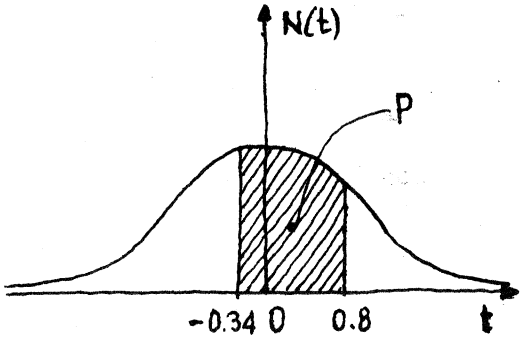


Figure 4.4-iv

Solution: We are going to use the Table in Appendix II-B.

(i) $P(h \leq 68) = P(t \leq \frac{68-66}{5})$

$= P(t \leq 0.40) = \underline{0.6554}$.

Hence, $K_1 = (0.6554)(1000) \doteq 655$ students.

(ii) $P(h \leq 61) = P(t \leq \frac{61-66}{5})$

$= P(t \leq -1) = 1-P(t \leq 1)$

$= 1. - 0.8413 = \underline{0.1587}$.

Hence, $K_2 = (0.1587)(1000) \doteq 159$ students.

(iii) $P(h \geq 74.6) = P(t \geq \frac{74.6-66}{5})$

$= P(t \geq 1.72)$

$= 1. - P(t \leq 1.72)$

$= 1. - 0.9573 = \underline{0.0427}$.

Hence, $K_3 = (0.0427)(1000) \doteq 43$ students.

(iv) $P(64.3 \leq h \leq 70) =$

$= P\left(\frac{64.3-66}{5} \leq t \leq \frac{70-66}{5}\right)$

$= P(-0.34 \leq t \leq 0.80)$

$= P(t \leq 0.80) - P(t \leq -0.34)$

$= P(t \leq 0.80) - (1-P(t \leq 0.34)$

$= 0.7881 - [1-0.6331]$

$= 0.7881 - 0.3669 = \underline{0.4212}$.

Hence $K_4 = (0.4212)(1000) \doteq 421$ students.

Example 4.3

For the normal random variable h given in example 4.2, determine the student's height H such that:

(i) $P(h \leq H_1) = 0.6554$ (Figure 4.5-i);

(ii) $P(h \geq H_2) = 0.25$ (Figure 4.5-ii);

(iii) $P(h \leq H_3) = 0.20$ (Figure 4.5-iii);

(iv) $P(H_4 \leq h \leq H_5) = 0.95$,

where $H_4 = \mu_h - K$ and $H_5 = \mu_h + K$. (Figure 4.5-iv).

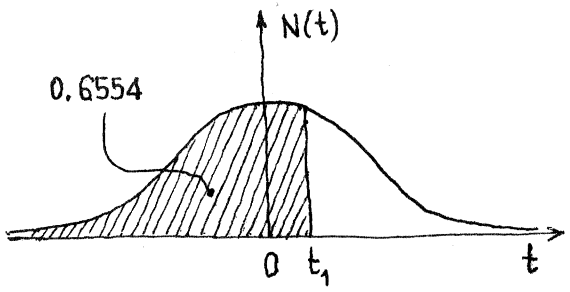


Figure 4.5-i

Solution: Again in this example, we are going to use the standard normal CDF table given in Appendix II-B.

(i) $P(h \leq H_1) = P(t \leq t_1) = 0.6554$.

From the above mentioned table, we get $t_1 = 0.4$, that corresponds to probability $P = 0.6554$. But we know that $t_1 = \frac{H_1 - \mu_h}{\sigma_h}$.

From example 4-2 we have $\mu_h = 66$ inches and $\sigma_h = 5$ inches. Hence,

$$t_1 = \frac{H_1 - 66}{5} = 0.4$$

from which we get

$$H_1 - 66 = 5(0.4) = 2,$$

i.e.

$$H_1 = 66 + 2 = \underline{68} \text{ inches},$$

which is identical to the first case in example 4.2; however, what we are doing here is nothing else but the inverse solution.

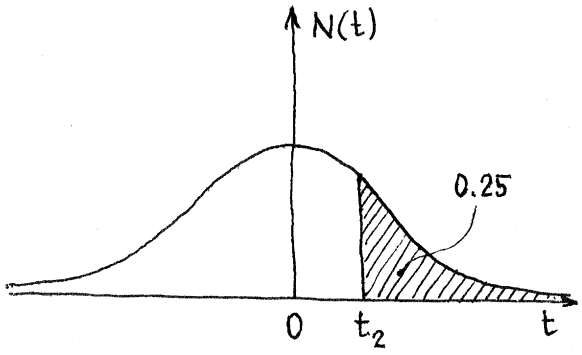


Figure 4.5-ii

(ii) $P(h \geq H_2) = P(t \geq t_2) = 0.25.$

But:

$P(t \geq t_2) = 1-P\ (t \leq t_2) = 0.25,$

and we get

$P(t \leq t_2) = 1 - 0.25 = 0.75.$

By interpolation in the above mentioned table we get $t_2 \doteq 0.675$ which corresponds to P = 0.75. Hence,

$$t_2 = \frac{H_2-\mu_h}{\sigma_h} = \frac{H_2-66}{5} = 0.675.$$

i.e. $H_2 = 66 + 5\ (0.675)$

$= 66 + 3.375 = \underline{69.375}$ inches .

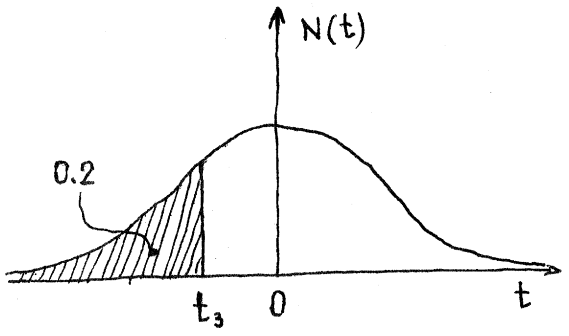


Figure 4.5-iii

(iii) $P(h \leq H_3) = P(t \leq t_3) = 0.20.$

By examining the above mentioned table we discover that the smallest probability reading is 0.50, since it considers only the positive values of t. Therefore we have to write:

$P(t \leq t_3) = 1 - P(t \leq -t_3) = 0.20,$

and we get

$P(t \leq -t_3) = 1 - 0.20 = 0.80.$

By interpolation in the above mentioned table we get: $(-t_3) = 0.842$, which corresponds to P = 0.80. Then we have:

$$t_3 = \frac{H_3-66}{5} = -0.842$$

and, $H_3 = 66-5(0.842)$

$= 66-4.210 = \underline{61.79}$ inches.

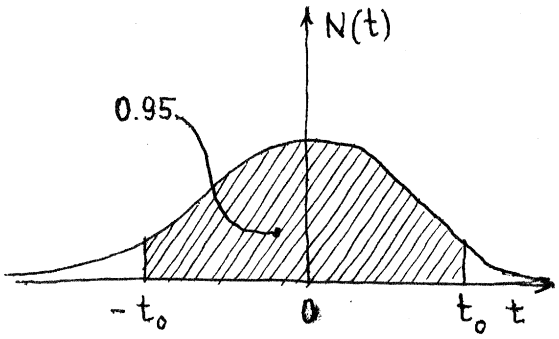


Figure 4.5-iv

(iv) $P(H_4 \leq h \leq H_5)$

$$= P\left(\frac{H_4-\mu_h}{\sigma_h} \leq t \leq \frac{H_5-\mu_h}{\sigma_h}\right)$$

$$= P\left(\frac{-K}{\sigma_h} \leq t \leq \frac{K}{\sigma_h}\right)$$

$$= P(-t_o \leq t \leq t_o) = 0.95,$$

where $t_o = \frac{K}{\sigma_h} = \frac{K}{5}$.

The above statement means that:

$P(t \leq t_o) - P(t \leq -t_o) = 0.95.$

However, from the symmetry of the normal PDF we get:

$P(t \geq t_o) = P(t \leq -t_o) = \frac{1. - 0.95}{2} = 0.025$

and we get:

$P(t \leq t_o) = 0.95 + 0.025 = 0.975,$

<u>or</u> $P(t \leq t_o) = 1. - 0.025 = 0.975.$

From the above mentioned table we get:

$t_o = 1.96$, which corresponds to

$P = 0.975$, and we have:

$$t_o = \frac{K}{5} = 1.96,$$

i.e. $K = 5(1.96) = 9.80$. Consequently:

$H_4 = \mu_h - K = 66 - 9.80 = \underline{56.2}$ inches

and

$H_5 = \mu_h + K = 66 + 9.80 = \underline{75.8}$ inches.

Example 4.4:

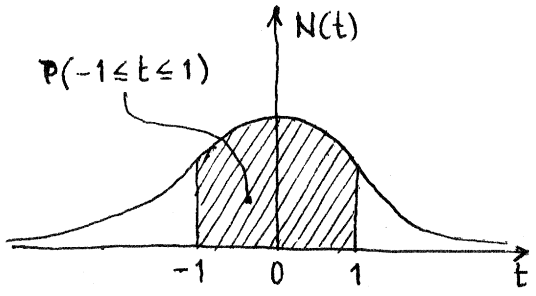


Figure 4.6

Let us solve example 4.1 again by using the standard normal CDF tables. Recall that it was required to compute $P(-\sigma_\varepsilon \leq \varepsilon \leq \sigma_\varepsilon)$, where ε has a Gaussian PDF (i.e. its $\mu_\varepsilon = 0$). We can write:

$$P(-\sigma_\varepsilon \leq \varepsilon \leq \sigma_\varepsilon) = P\left(\frac{-\sigma_\varepsilon - \mu_\varepsilon}{\sigma_\varepsilon} \leq t \leq \frac{\sigma_\varepsilon - \mu_\varepsilon}{\sigma_\varepsilon}\right)$$

$$= P\left(\frac{-\sigma_\varepsilon - 0}{\sigma_\varepsilon} \leq t \leq \frac{\sigma_\varepsilon - 0}{\sigma_\varepsilon}\right)$$

$= P(-1 \leq t \leq 1)$, see Figure 4.6.

Further we can write:

$P(-1 \leq t \leq 1) = 2P(0 \leq t \leq 1)$.

From the table given in Appendix II-C, we get:

$P(0 \leq t \leq 1) = 0.3413$.

Hence,

$$P(-\sigma_\varepsilon \leq \varepsilon \leq \sigma_\varepsilon) = 2(0.3413)$$

$$= 0.6826 \doteq \underline{0.683},$$

which is the same result as obtained in example 4.1.

4.7 Basic Hypothesis (Postulate) of the Theory of Errors, Testing

We have left the random sample of observations L behind in section 4.2 while we developed the analytic formulae for the PDF's mostly used in the error theory. Let us get back to it and state the following basic postulate of the error-theory. A finite sequence of observations L representing the same physical quantity is declared a random sample with parent random variable distributed according to the normal PDF $N(\mu_\ell, \sigma_\ell; \ell)$. Other PDF's are used rather seldom. The validity of this hypothesis may or may not be tested, on which topic we shall not elaborate here.

The mean M_L of the sample L is said to approximate (the word estimate is often used in this context) the mean μ_ℓ of the parent PDF, i.e. $N(\mu_\ell, \sigma_\ell; \ell)$. Also, the variance S_L^2 of the sample L is said to estimate the variance σ_ℓ^2 of the parent PDF.

Considering the original sample

$$L = (\ell_i) = (\ell' + \varepsilon_i) \ , \quad i = 1, 2, \ldots, n,$$

we get:

$$M_L = \frac{1}{n} \sum_{i=1}^{n} \ell_i = \frac{1}{n} \sum_{i=1}^{n} (\ell' + \varepsilon_i) = \frac{1}{n} (n\ell') + \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i = \ell' + M_\varepsilon \ . \quad (4\text{-}20)$$

Since the random errors ε's are postulated to have a parent Gaussian PDF $N(0, \sigma_\varepsilon; \varepsilon)$, which implies that $\mu_\varepsilon = 0$, then we should expect that $M_\varepsilon \to 0$ and we can write equation (4-20) as:

$$M_L \doteq \ell' = \mu_\ell \ , \quad (4\text{-}21)$$

keeping in mind that by the unknown value ℓ' we mean the unknown mean μ_ℓ of the parent PDF of ℓ. We say that the mean M_L of the sample L approximates (estimates) the value of the mean μ_ℓ of the parent PDF of ℓ.

Similarly, we get

$$S_L^2 = \frac{1}{n}\sum_{i=1}^{n}(\ell_i - M_L)^2 = \frac{1}{n}\sum_{i=1}^{n}[\ell_i - (\ell' + M_\varepsilon)]^2 = \frac{1}{m}\sum_{i=1}^{m}(\varepsilon_i - M_\varepsilon)^2 = S_\varepsilon^2 \qquad (4.22)$$

The above result indicates that the variance S_L^2 of the sample L is identical to the variance S_ε^2 of its corresponding sample of random errors ε. This is actually why S_L^2 is sometimes called the mean square error of the sample, and is abbreviated by MSE. Also, S_L is known as the root mean square error of the sample, and is abbreviated by RMS. According to the basic hypothesis of the error-theory we can write equation (4-22) as:

$$S_L^2 = S_\varepsilon^2 \doteq \sigma_\ell^2 = \sigma_\varepsilon^2 \quad , \qquad (4\text{-}23)$$

which states that S_L^2 estimates the variance σ_ε^2 of the parent PDF of $\varepsilon \equiv (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$.

Example 4.5 Assume that the sample L ≡ (2, 7, 6, 4, 2, 7, 4, 8, 6, 4) is postulated to be normally distributed. Let us transform this sample in such a way that the transformed sample will have:

(i) Gaussian distribution ;

(ii) Standard normal distribution.

Solution: First we compute the mean M_L and the variance S_L^2 of the given sample as follows:

$$M_L = \frac{1}{10} \sum_{i=1}^{10} \ell_i = \frac{1}{10} (50) = 5,$$

$$S_L^2 = \frac{1}{10} \sum_{i=1}^{10} (\ell_i - 5)^2 = \frac{1}{10} (40) = 4.$$

According to the basic postulate of the error-theory we can say that:

$$\mu_\ell \doteq M_L = 5 \text{ and } \sigma_\ell = \sigma_\varepsilon \doteq S_L = 2,$$

where μ_ℓ and σ_ℓ are respectively the mean and the standard deviation of the parent normal PDF $N(\mu_\ell, \sigma_\ell; \ell)$ assumed for the given sample. The parameters μ_ℓ and σ_ℓ will be used for the required transformations as follows:

(i) The Gaussian distribution $G(\sigma_\varepsilon; \varepsilon)$, where $\sigma_\varepsilon = \sigma_\ell = 2$, has an argument ε obtained from equation (4-10) as:

$$\varepsilon_i = \ell_i - \mu_\ell = \ell_i - 5, \quad i = 1, 2, \ldots, 10.$$

Hence the transformed sample that has a Gaussian PDF is:

$$\varepsilon \equiv (\varepsilon_i), \quad i = 1, 2, \ldots, 10, \text{ i.e.:}$$

$$\varepsilon \equiv (-3, 2, 1, -1, -3, 2, -1, 3, 1, -1).$$

(ii) The standard normal distribution $N(t)$, has an argument t obtained from equation (4-13) as:

$$t_i = \frac{\ell_i - \mu_\ell}{\sigma_\ell} = \frac{\ell_i - 5}{2} \quad , \; i = 1, 2, \ldots, 10.$$

Hence, the transformed sample that has a standard normal PDF is $T \equiv (t_i)$, $i = 1, 2, \ldots, 10$, i.e.:

$T \equiv (-1.5, 1, 0.5, -0.5, -1.5, 1, -0.5, 1.5, 0.5, -0.5)$.

4.8 Residuals, Corrections and Discrepencies

As we have seen, we are not able to compute the unknown value ℓ 'or μ_ℓ. All we can get is an estimate $\bar{\ell}$ for it from the following equation

$$\bar{\ell} = M_L = \ell' + M_\varepsilon = \ell' + \bar{\varepsilon}, \; *) \qquad (4\text{-}24)$$

and hope that $\bar{\varepsilon}$, in accordinace with the basic postulate of the error-theory, will really go to zero.

The residual r_i is defined as the difference between the observation ℓ_i and the sample mean $\bar{\ell}$, i.e.

---

* From now on, we shall use the symbol $\bar{\ell}$ for the mean M_L of the sample L. The "bar" above the symbol will indicate the sample mean to make the notation simpler.

$$\boxed{r_i = \ell_i - \bar{\ell} = (\ell' + \varepsilon_i) - (\ell' + \bar{\varepsilon}) = \varepsilon_i - \bar{\varepsilon}.} \qquad (4.25)$$

Residuals with inverted signs are usually called corrections. It should be noted that a residual, as defined above, is a uniquely determined value and not a variable. The observed value ℓ_i is fixed and so is the mean $\bar{\ell}$ for the particular sample. In other words, for a given sample, the residuals can be computed in one way only. Note that the differences $(\ell_i - \bar{\ell}) = r_i$ are called residuals and not errors, because errors are defined as $\varepsilon_i = (\ell_i - \mu_\ell)$ and μ_ℓ may be different from $\bar{\ell}$.

In practice, one often hears talks about "minimized residuals", "variable residuals" etc. which are not strictly correct. If one wants to regard the "residuals as variables" the problem has to be stated differently. The difference v_i between the observed value ℓ_i and any arbitrarily assumed (or computed) value ℓ°, i.e.

$$\boxed{v_i = \ell_i - \ell^\circ, \quad i = 1, 2, \ldots, n} \qquad (4.26)$$

should be called discrepency, or misclosure, to distinguish it from the residual. These discrepencies are obviously linear functions of ℓ°; their values vary with the choice of ℓ°. Hence one can talk about "minimization of discrepencies", "variation of discrepencies" etc. Evidently, residuals and discrepencies are very often mixed up in practice.

At this point it is worthwhile to mention yet another pair of formulae for computing the sample mean $\bar{\ell}$ and the sample variance S^2_L. Such simplified formulae facilitate the computations especially for large samples

whose elements have large numerical values. The development of these formulae is done analogically to the formulation of equations (4.20), (4.22), (4.25) and (4.26). Here we state only the results, and the elaboration is left to the student.

$$\bar{\ell} = \ell^\circ + \bar{v} \,, \tag{4.27}$$

and

$$S_L^2 = \frac{1}{n} \sum_{i=1}^{n} r_i^2 \,, \tag{4.28}$$

where: ℓ° is an arbitrarily chosen value, usually close to $\bar{\ell}$,

$$\bar{v} = \frac{1}{n} \sum_{i=1}^{n} v_i \,, \tag{4.29}$$

$$v_i = \ell_i - \ell^\circ$$

and

$$r_i = \ell_i - \bar{\ell} = v_i - \bar{v} \,. \tag{4.30}$$

Example 4.6: The second column of the following table is a sample of 10 observations of the same distance. It is required to compute the sample mean and variance using the simplified formulae given in this section.

We take $\ell^\circ = 972.0$ m,

$$\bar{v} = \frac{1}{10} \sum_{i=1}^{10} v_i = \frac{1}{10} (10.50) = 1.05 \text{ m},$$

$$\bar{\ell} = \ell^\circ + \bar{v} = 972.0 + 1.05 = \underline{973.05} \text{ m},$$

$$\text{MSE} = S_L^2 = \frac{1}{10} \sum_{i=1}^{10} r_i^2 = \frac{1}{10} (0.5730) = \underline{0.0573}. \text{ m}^2$$

and RMS $= S_L \doteq \underline{0.24}$ m.

One of the checks on the computations is that $\sum_{i=1}^{n} r_i = 0$, see the fourth column of the given table.

No.	ℓ_i (m)	$v_i = \ell_i - \ell^\circ$ (m)	$r_i = \ell_i - \bar{\ell} = v_i - \bar{v}$	$r_i^2 \cdot 10^4$ (m^2)
1	972.89	0.89	-0.16	256
2	973.46	1.46	0.41	1681
3	973.04	1.04	-0.01	1
4	972.73	0.73	-0.32	1024
5	972.63	0.63	-0.42	1764
6	973.01	1.01	-0.04	16
7	973.22	1.22	0.17	289
8	973.10	1.10	0.05	25
9	973.30	1.30	0.25	625
10	973.12	1.12	0.07	49
$\Sigma =$		10.50	-0.95 +0.95 $= 0.00$	5730

4.9 Other Possibilities Regarding the Postulated PDF

The normal PDF (or its relatives) are by no means the only bell-shaped PDF's that can be postulated. Under different assumptions, one can derive a whole multitude of bell-shaped curves. Generally, they would contain more than two parameters which is an advantage from the point of view of fitting them to any experimental PDF. In other words the additional parameters provide more flexibility. On the other hand, the computations

with such PDF's are more troublesome. In this context let us just mention that some recent attempts have been made to design a family of PDF's that are more peaked than the normal PDF in the middle. Such PDF's are called "Leptokurtic". This more pronounced peakedness is a feature that quite a few scholars claim to have spotted in the majority of observational samples. We shall have to wait for any definite word in this domain for some time.

Hence, the normal is still the most popular PDF and likely to remain so because it is relatively simple and contains the least possible number of parameters - the mean and the standard deviation.

4.10 Other Measures of Dispersion

So far, we have dealt with two measures of dispersion of a sample namely: The root mean square error (RMS) mentioned in section 4.7, and the range (Ra) mentioned in section 3.1.5. Besides the RMS and the range of a sample the following measures of dispersion (spread) are often used.

The average or mean error a_e of the sample L is defined as

$$a_e = \frac{1}{n} \sum_{i=1}^{n} |\ell_i - \bar{\ell}| = \frac{1}{n} \sum_{i=1}^{n} |r_i| \,, \qquad (4.31)$$

which is the mean of the absolute values of the residuals.

The most probable error p_e, of the sample L, is defined as the error for which:

$$P(|r| < p_e) = P(|r| > p_e) = 0.50 \qquad (4.31)$$

which means that there is 50% probability that the residual is smaller and 50% probability that the residual is larger than p_e.

The most probable error of a random sample can be computed by constructing the CDF of the corresponding absolute values of the sample residuals, and take the value of r which corresponds to the CDF = 0.5 as the value of p_e.

Both a_e and p_e can be defined for the continuous distributions as well. For instance, by considering the normal PDF, $N(\mu_x,\sigma_x; x)$, we can write:

$$a_e = \int_{-\infty}^{\infty} |x| \, \phi(x) \, dx$$

$$= \frac{1}{\sigma_x \sqrt{(2\pi)}} \int_{-\infty}^{\infty} |x| \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right) dx . \qquad (4.33)$$

Similarly for p_e, by taking the symmetry of the normal curve into account, we can write:

$$P(x \leq \mu_x - p_e) = \Psi_N(\mu_x - p_e) =$$

$$= \frac{1}{\sigma_x \sqrt{(2\pi)}} \int_{-\infty}^{(\mu_x - p_e)} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right) dx = 0.25 \qquad (4.34)$$

and

$$P(x \leq \mu_x + p_e) = \Psi_N(\mu_x + p_e) = 0.75 \qquad (4.35)$$

where Ψ_N is the normal CDF.

It can be shown for the normal PDF $N(\mu_x, \sigma_x; x)$ that "σ_x", "a_e" and "p_e" are related to each other by the following approximate relation:

$$\sigma_x \doteq 1.25\ a_e \doteq 1.5\ p_e,$$

or

$$\sigma_x : a_e : p_e \doteq 1.0 : 0.80 : 0.67. \tag{4.36}$$

The relative or proportional error r_e, of the sample L, is defined as the ratio between the sample RMS and the sample mean, i.e.

$$r_e = S_L / \bar{\ell}. \tag{4.37}$$

In practice, the relative error is usually used to describe the uncertainty of the result, i.e. the sample mean. In that case, the relative error is defined as:

$$\boxed{\bar{r}_e = S_{\bar{\ell}} / \bar{\ell}} \tag{4.38}$$

where $S_{\bar{\ell}}$ is the standard deviation of the mean $\bar{\ell}$ and will be derived later in Chapter 6. In this respect, one often hears expressions like "proportional accuracy 3 ppm (parts per million)", which simply means that the relative error is $3/10^6 = 3 \cdot 10^{-6}$. It should be noted that unlike the other measures of dispersion, the relative error is unitless.

The idea of the confidence intervals is based on the assumption of normality of the sample, i.e. the postulated parent normal PDF $(N(\bar{\ell}, S_L; \ell))$ for the random sample L. It is very common to represent the sample L by its mean $\bar{\ell}$ and its standard deviation S_L as

$$\boxed{[\bar{\ell} \pm S_L]}$$

or

$$[\bar{\ell} - S_L \leq \ell \leq \bar{\ell} + S_L], \tag{4.39}$$

and refer to it as the "68% confidence interval" of ℓ. This is based on the fact that the probability $P(\mu_\ell - \sigma_\ell \leq \ell \leq \mu_\ell + \sigma_\ell)$ is approximately 0.68 for

the normal PDF (see section 4.5).

Similarly, one can talk about the "95% confidence interval", the "99% confidence interval", etc. In general, the confidence interval of ℓ is expressed as:

$$[\bar{\ell} \pm K S_L] \tag{4.40}$$

where K is determined in such a way as to make

$$P(\mu_\ell - K\sigma_\ell \leq \ell \leq \mu_\ell + K\sigma_\ell) \text{ equal to } 0.95,\ 0.99,\ \text{etc.}$$

The values $(\bar{\ell} - KS_L)$ and $(\bar{\ell} + KS_L)$ are called the lower and the upper confidence limits.

Example 4.7: Let us compute the average error, the relative error and the 95% confidence interval for the sample of observations L given in example 4.6.

The average error is computed using equation (4.31) and the fourth column of the given table in example 4.6 as:

$$a_e = \frac{1}{10} \sum_{i=1}^{10} |r_i| = \frac{1}{10}(1.90) = \underline{0.19} \text{ m} .$$

The relative error of the sample is computed from equation (4.37) and the results obtained in example 4.6 as:

$$r_e = S_L / \bar{\ell} = \frac{0.24}{973.05} \doteq \underline{247 \text{ ppm}} .$$

The 95% confidence interval of ℓ is

$$[\bar{\ell} - K S_L \leq \ell \leq \bar{\ell} + K S_L] .$$

where the number K is computed so that

$$P(\mu_\ell - K\sigma_\ell \leq \ell \leq \mu_\ell + K\sigma_\ell) = 0.95 .$$

This is identical to the probability $P(-K \leq t \leq K)$ obtained from the standard normal tables (see example 4.3, the last case). Hence we can write:

$$P(-K \leq t \leq K) = P(t \leq K) - P(t \leq -K) = 0.95$$

from which we get

$$P(t \leq K) = 0.975.$$

Using the table for the standard normal variable of Appendix II - B we get:

$$K = \underline{1.96}.$$

(In practice K = 2 is usually used for the 95% confidence interval.) The 95% confidence interval of ℓ then becomes

$$[973.05 - 1.96\ (0.24) \leq \ell \leq 973.05 + 1.96(0.24)],$$

that is:

$$[972.58 \leq \ell \leq 973.52]\ \text{m}$$

or

$$\underline{[973.05 \pm 0.47]\ \text{m}}.$$

Example 4.8: Given a random variable x assumed to have a normal distribution N (35, 4; $\mathbf{x}$), compute the most probable error.

From the assumed PDF we have:

$$\mu_x = 35 \text{ and } \sigma_x = 4.$$

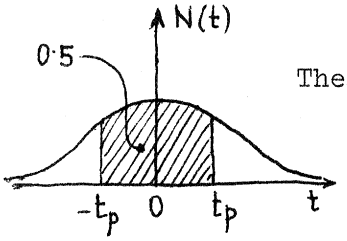


Figure 4.7a

The most probable error p_e is computed so that

$$P(\mu_x - p_e \leq x \leq \mu_x + p_e) =$$

$$= P(-t_p \leq t \leq t_p) = 0.50, \text{ (Figure 4.7a)}$$

where $t_p = \dfrac{p_e}{\sigma_x} = \dfrac{p_e}{4}$.

The above probability statement can be rewritten as (equation (4.35)):

$P(x \leq \mu_x + p_e) = P(t \leq t_p) = 0.75$ (Figure 4.7b).

From the table in Appendix II - B, we obtain $t_p \doteq 0.675$ corresponding to P = 0.75. Hence,

$$p_e = 4\, t_p = 4\,(0.675) = \underline{2.7}\ .$$

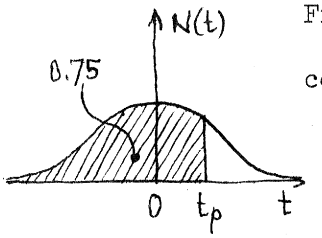


Figure 4.7b

Note that in the second case of example 4.3, the value 3.375 is nothing else but the most probable error of the given random variable h.

4.11 Exercise 4

1. Prove that the Gaussian PDF given by equation (4.4), has two points of inflection at abscissas $\pm \sqrt{C}/2$.

2. For the Gaussian PDF given by equation (4.8), determine approximately the probabilities: $P(-2\sigma_\varepsilon \leq \varepsilon \leq 2\sigma_\varepsilon)$ and $P(-3\sigma_\varepsilon \leq \varepsilon \leq 3\sigma_\varepsilon)$ by integrating the PDF, then check your results by using the standard normal tables.

3. Prove by direct evaluation that the standard normal PDF has a standard deviation equals to one.

4. Show that the standard deviation σ, average error a_e and the most probable error p_e of the normal PDF satisfy the following approximate relations:

$$\sigma : a_e : p_e \doteq 1.0 : 0.80 : 0.67.$$

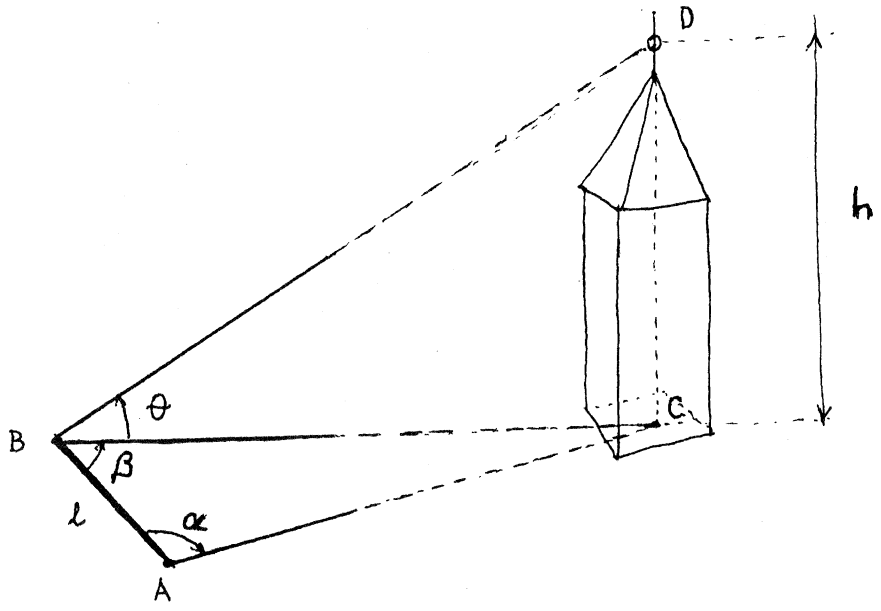
5. Determine: the average error, the most probable error, the relative error and the 90% confidence interval of the random sample given in the second problem of exercise 3, section 3.4.

6. Assume that the sample $H \equiv (-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5)$ is hypothesized (postulated) to have a Gaussian distribution. Transform this sample so that the transformed (new) sample will have:
   (i) Normal distribution with mean equal 10.
   (ii) Standard normal distribution.

7. Given a random variable x distributed as $N(25, 10; x)$, determine the following probabilities:
   (i) $P(x \leq 28.5)$, (ii) $P(x \leq 22.5)$,
   (iii) $P(x \geq 27.5)$, (iv) $P(16.75 \leq x \leq 23.82)$,
   (v) $P(|x-25| \leq 1.25)$.

8. For the random variable in the previous problem, determine the values Z_i such that
   (i) $P(x \leq Z_1) = 0.65$, (ii) $P(x \geq Z_2) = 0.025$,
   (ii) $P(x \leq Z_3) = 0.33$, (iv) $P(|x-25| \leq Z_4) = 0.33$,
   (v) $P(|x-25| \geq Z_5) = 0.50$.

9.



The above figure shows a surveying technique to determine the height h of a tower CD, which cannot be measured directly. The observed quantities are:

ℓ = the horizontal distance AB ,

α , β = the horizontal angles at A and B ,

θ = the vertical angle of D at B .

The field results of these observations are given in the following table:

Field Observations			
ℓ(m)	α	β	θ
145.63	65° 32' 03"	37° 13' 08"	42° 53' 15"
.55	32 04	13 11	52 30
.59	31 59	13 10	53 00
.65	32 01	13 13	51 00
.58	31 58	13 06	52 15
		13 12	52 45
			51 15
			53 00
			51 45
			52 15

Average temperature during the observations time was T = 20° F.

The following information was given to the observer:

(i) The micrometer of the vertical circle of the used theodolite was not adjusted to read 00' 00" when the corresponding bubble axis is horizontal; it reads - (00' 30").

(ii) The nominal length of the used tape is 20 m at the calibration temperature T_0 = 60° F, and the coefficient of expansion of the tape material is $\gamma = 5 \cdot 10^{-5}$ / 1° F .

Required

(i) Compute the estimated values for the quantities ℓ, α, β and θ.

(ii) For each of the above observed quantities compute its standard deviation and its average error.

(iii) Compare the precision of these observed quantities (by comparing the respective relative errors).

(iv) Assume that each of these observed quantities has a postulated normal parent PDF, construct the 95% confidence interval for each quantity.

(v) Compute the estimated value of the tower's height h to the nearest centimeter.

5. LEAST-SQUARES PRINCIPLE

5.1 The Sample Mean as "The Least Squares Estimator"

One may now ask oneself a hypothetical question: given the sample $L \equiv (\ell_i)$, $i = 1, 2, \ldots, n$, what is the value ℓ^o that makes the summation of the squares of the discrepancies

$$v_i = \ell_i - \ell^o \quad , \quad i = 1, 2, \ldots, n, \tag{5-1}$$

the smallest (i.e. minimum)?

The above question may be stated more precisely as follows: Defining a "new variance" S^{*2} as

$$S^{*2} = \frac{1}{n} \sum_{i=1}^{n} (\ell_i - \ell^o)^2 = \frac{1}{n} \sum_{i=1}^{n} v_i^2 \quad , \tag{5-2}$$

find the value ℓ^o that is going to give us the smallest (minimum) value of S^{*2}.

Obviously, such a question can be answered mathematically. From equation (5-2), we notice that S^{*2} is a function of ℓ^o, which is the only free variable here and can be written as

$$S^{*2} = S^{*2}(\ell^o) \quad . \tag{5-3}$$

We know that:

$$\min_{\ell^o \in R} [S^{*2}(\ell^o)] \text{ implies that } \frac{\partial S^{*2}}{\partial \ell^o} = 0 \quad .$$

Hence, by differentiating equation (5-2) with respect to ℓ^o and equating it to zero, we get:

$$\frac{\partial S^{*2}}{\partial \ell^{o}} = \frac{1}{n} \frac{\partial}{\partial \ell^{o}} \left[\sum_{i=1}^{n} (\ell_i - \ell^{o})^2 \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial}{\partial \ell^{o}} (\ell_i - \ell^{o})^2 \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} [2(\ell_i - \ell^{o})(-1)] = \frac{-2}{n} \sum_{i=1}^{n} (\ell_i - \ell^{o}) = 0 ,$$

that is:

$$\sum_{i=1}^{n} (\ell_i - \ell^{o}) = 0 .$$

The above equation can be rewritten as:

$$\sum_{i=1}^{n} \ell_i = \sum_{i=1}^{n} \ell^{o} = n\ell^{o} ,$$

which yields

$$\ell^{o} = \frac{1}{n} \sum_{i=1}^{n} \ell_i \equiv \bar{\ell} . \qquad (5\text{-}4)$$

The result (5-4) is nothing else but the "sample mean" $\bar{\ell}$ again. In other words, the mean of the sample is the value that minimizes the sum of the squares of the discrepancies making them equal to the residuals, (see section 4.8).

This is the reason why the mean $\bar{\ell}$ is sometimes called the least-squares estimation (estimator) of ℓ', i.e. of μ_ℓ ; the name being derived from the process of minimization of the squares of the discrepancies. We also notice that $\bar{\ell}$ minimizes the variance of the sample if we want to regard the variance as a function of the mean.

Note that the above property of the mean is completely independent of the PDF of the sample. This means that the sample mean $\bar{\ell}$ is <u>always</u> "the minimum variance estimator of ℓ" whatever the PDF may be.

5.2 The Sample Mean as "The Maximum Probability Estimator"

Let us take our sample L again, and let us postulate an underlying parent PDF to be normal (see section 4.5) with a mean $\mu_\ell = \ell^\circ$ and a variance σ_ℓ^2 given by:

$$\sigma_\ell^2 = S^{*2} = \frac{1}{n} \sum_{i=1}^{n} (\ell_i - \ell^\circ)^2 = \frac{1}{n} \sum_{i=1}^{n} v_i^2 . \qquad (5.5)$$

We say that the normal PDF, $N(\mu_\ell, \sigma_\ell; \ell) \equiv N(\ell^\circ, S^*; \ell)$ is the most probable underlying PDF for our sample L ($L = (\ell_i)$, $i = 1, 2, \ldots n$) if the combined probability of simultaneous occurrence of n elements, that have the normal distribution $N(\ell^\circ, S^*; \ell)$, at the same places as L is maximum. In other words, we ask that:

$$P[(\ell_i \leq \ell \leq \ell_i + \delta\ell_i),\ i = 1, 2.. n] =$$

$$= \prod_{i=1}^{n} N(\ell^\circ, S^*; \ell_i) \cdot \delta\ell_i \qquad (5.6)$$

be maximum with respect to the existing free parameters. By examining equation (5.6), we find that the only free parameter is ℓ° (note that S* is a function of ℓ°), and hence we can write the above combined probability as a function of ℓ° as follows:

$$P[(\ell_i \leq \ell \leq \ell_i + \delta\ell_i),\ i = 1, 2, \ldots, n] = \lambda\ (\ell^\circ) \qquad (5.7)$$

Note that $\delta\ell$'s are some values depending on L and therefore are determined uniquely by L.

We shall show that the value of ℓ° satisfying the above condition is (for the postulated normal PDF) again the value rendering the smallest value of S*. We can write:

$$\max_{\ell^\circ \varepsilon R} [\lambda(\ell^\circ)] = \max_{\ell^\circ \varepsilon R} [\prod_{i=1}^{n} N(\ell^\circ, S^*; \ell_i)\, \delta\ell_i]$$

$$= \max_{\ell^\circ \varepsilon R} [\prod_{i=1}^{n} \frac{1}{S^*\sqrt{(2\pi)}} \prod_{i=1}^{n} \exp\left(-\frac{(\ell_i - \ell^\circ)^2}{2S^{*2}}\right) \delta\ell_i]$$

$$= \max_{\ell^\circ \varepsilon R} [\left(\frac{1}{S^*\sqrt{(2\pi)}}\right)^n \prod_{i=1}^{n} \exp\left(-\frac{(\ell_i - \ell^\circ)^2}{2S^{*2}}\right) \delta\ell_i] . \quad (5.8)$$

Here $\prod_{i=1}^{n} \delta\ell_i$ is determined by L, and hence does not lend itself to maximization. It thus can be regarded as a constant, i.e.

$$\max_{\ell^\circ \varepsilon R} [\lambda(\ell^\circ)] = \max_{\ell^\circ \varepsilon R} [\left(\frac{1}{S^*\sqrt{(2\pi)}}\right)^n \prod_{i=1}^{n} \exp\left(-\frac{(\ell_i - \ell^\circ)^2}{2S^{*2}}\right)] . \quad (5.9)$$

Let us denote the second term in the RHS of equation (5.9) by Q, which can be expressed as:

$$Q = \prod_{i=1}^{n} \exp(-x_i), \text{ where } x_i = \frac{(\ell_i - \ell^\circ)^2}{2S^{*2}} . \quad (5.10)$$

This implies that:

$$\ell n\ Q = \ell n \left(\prod_{i=1}^{n} \exp(-x_i)\right) = \sum_{i=1}^{n} \ell n\,(\exp(-x_i)) ,$$

or

$$Q = \exp\left(\sum_{i=1}^{n} (-x_i)\right) . \quad (5.11)$$

From equations (5-9), (5-10) and (5-11) we get:

$$\prod_{i=1}^{n} \exp\left(-\frac{(\ell_i-\ell^o)^2}{2S^{*2}}\right) = \exp\left[-\frac{1}{2S^{*2}} \sum_{i=1}^{n} (\ell_i-\ell^o)^2\right] . \qquad (5\text{-}12)$$

The condition (5-9) can be then rewritten as:

$$\max_{\ell^o \varepsilon R} [\lambda(\ell^o)] = \max_{\ell^o \varepsilon R} \left[\left(\frac{1}{S^*\sqrt{(2\pi)}}\right)^n \exp\left(-\frac{1}{2S^{*2}} \sum_{i=1}^{n} (\ell_i-\ell^o)^2\right)\right]. \qquad (5\text{-}13)$$

From equation (5-5), we have:

$$\sum_{i=1}^{n} (\ell_1-\ell^o)^2 = nS^{*2}.$$

Hence by substituting this value into equation (5-13) we get:

$$\max_{\ell^o \varepsilon R} [\lambda(\ell^o)] = \max_{\ell^o \varepsilon R} \left[\left(\frac{1}{S^*\sqrt{(2\pi)}}\right)^n \exp\left(-\frac{n}{2}\right)\right]. \qquad (5\text{-}14)$$

Since the only quantity in equation (5-14) that depends on ℓ^o is S*, we can write:

$$\max_{\ell^o \varepsilon R} [\lambda(\ell^o)] = \max_{\ell^o \varepsilon R} \left[\left(\frac{1}{S^*}\right)^n\right] = \max_{\ell^o \varepsilon R} [(S^*)^{-n}]$$

$$= \min_{\ell^o \varepsilon R} [(S^*)^n]. \qquad (5\text{-}15)$$

Because S* is a non-negative (quadratic) function of ℓ^o, the minimum of $(S^*)^n$ will be attained for the same argument as the minimum of S* (see Figure 5-1.

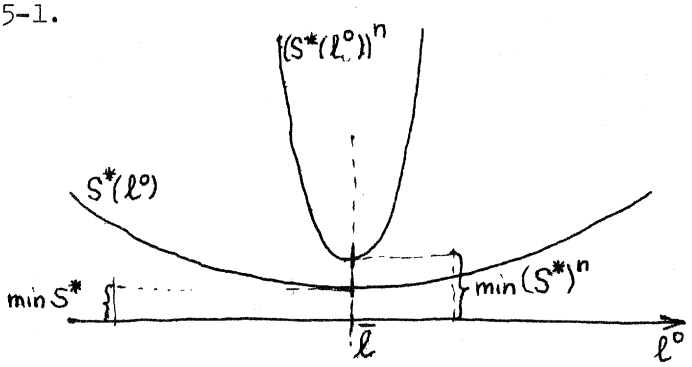


Figure 5-1

Finally, our original condition (equation (5-9) can be restated as:

$$\max_{\ell^o \epsilon R} [\lambda(\ell^o)] = \min_{\ell^o \epsilon R} [S^*(\ell^o)] \equiv \min_{\ell^o \epsilon R} [S^{*2}(\ell^o)] \qquad (5\text{-}16)$$

which implies that

$$\frac{\partial S^*}{\partial \ell^o} = \frac{\partial S^{*2}}{\partial \ell^o} = 0 \ ,$$

that is:

$$\frac{\partial}{\partial \ell^o} \sum_{i=1}^{n} v_i^2 = 0 \ . \qquad (5\text{-}17)$$

Obviously, the condition (5-17) is the same condition as that of the "minimum variance" discussed in the previous section, and again we have

$$\ell^o = \bar{\ell} \ .$$

We have thus shown that under the postulate for the underlying PDF, the mean $\bar{\ell}$ of the sample L is the maximum probability estimator for ℓ. As a matter of fact, we would find that the requirement of maximum probability leads to the condition

$$\frac{\partial S^*}{\partial \ell^o} = 0 \qquad (5.18)$$

for quite a large family of PDF's, in particular the symmetrical PDF's. If one assumes the additional properties of the random sample as mentioned in 3.2.4. then additional features of the sample mean can be shown. This again is considered beyond the scope of this course.

5.3 Least-Squares Principle

We have shown that the sample mean renders always the minimum sum of squares of discrepancies and that this property is required, for a large family of postulated PDF's, to yield the maximum probability for the underlying

PDF. Hence the sample mean $\bar{\ell}$, which automatically satisfies the condition of the least sum of squares of discrepancies, is at the same time the most probable value of the mean μ_ℓ of the underlying PDF under the condition that the underlying PDF is symmetrical. This is the necessary and sufficient condition for the sample mean to be both the least squares and the maximum probability estimator, i.e. for both estimators to be equivalent.

The whole development we have gone through does not say anything about the most probable value of the standard deviation σ_ℓ of the underlying PDF*). σ_ℓ has to be postulated according to equation (4.23).

The idea of minimizing the sum of squares of the discrepancies is known as the least-squares principle, and has got a fundamental importance in the adjustment calculus. We shall show later how the same principle is used for all kinds of estimates (not only the mean of a sample) and how it is developed into the least-squares method. However, the basic limitations of the least-squares principle should be born in mind, namely

(i) A normal PDF (or some other symmetrical PDF) is postulated.

(ii) The least-squares principle does not tell anything about the best estimator of σ_ℓ with respect to the mean μ_ℓ of the postulated PDF.

---

*) Some properties of the standard deviation S can be revealed if the additional properties of the random sample are assumed (see 3.2.4).

5.4 Least-Sqaures Principle for Random Multivariate

So far, we have shown that the least-squares principle spells out the equivalence between the sample mean $\bar{\ell}$ and the estimate for the parent population mean μ_ℓ determined from the condition that the sum of discrepancies be minimum. We have also shown that $\bar{\ell}$ is the most probable estimate for μ_ℓ providing the parent population is postulated to have normal or any other symmetrical PDF. We shall show now that the same principle is valid even for random multisample if we postulate the underlying PDF to be statistically independent (see Section 3.3.2).

Denoting the multisample by $\tilde{L}$ and its components by L^j, $j = 1, 2, \ldots, s$, and remembering that each L^j is a sample on its own, we can write:

$$\left.\begin{aligned} \tilde{L} &= (L^1, L^2, \ldots, L^s) \\ L^j &= (\ell_1^j, \ell_2^j, \ldots, \ell_{n_j}^j) \ \varepsilon R^s \quad . \end{aligned}\right\} \tag{5-19}$$

Assuming a particular value L_o for the multisample $\tilde{L}$, where

$$L_o = (\ell_o^1, \ell_o^2, \ldots, \ell_o^s) \ \varepsilon R^s \tag{5-20}$$

is a numerical vector (sequence of real numbers), the associated discrepancies $\tilde{V}$, which can be regarded as a multisample as well, are:

$$\tilde{V} = (\tilde{L}-L_o) \equiv (V^1, V^2, \ldots, V^s) \ . \tag{5-21}$$

Here, each V^j, $j = 1, 2, \ldots, s$ is a sample of discrepancies on its own, i.e.

$$V^j \equiv (V_1^j, V_2^j, \ldots, V_{n_j}^j) \ \varepsilon R^s \ . \tag{5-22}$$

Making use of formula (3-52), we can write analogically to (5.2)

$$S^{*2}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (\ell^j_i - \ell^j_o)^2 = E[(V^j)^2]. \qquad (5\text{-}23)$$

The minimization of the variances, i.e. minimization of each $E[(V^j)^2]$, is equivalent to the minimization of each S^{*2}_j, or as we usually write:

$$(\min_{L_o \varepsilon R^s} [E(V^j)^2],\ j = 1, 2, .., s) \equiv \min_{L_o \varepsilon R^s} [\text{trace } \Sigma_{\tilde{L}}]\ ^{*)}, \qquad (5\text{-}24)$$

where $\Sigma_{\tilde{L}}$ is the variance-covariance matrix of the multisample $\tilde{L}$ (see section 3.3.6). By carrying out this operation, similar to section 5.1, we will find that the vector

$$L_o = (\bar{\ell}^1, \bar{\ell}^2, \ldots, \bar{\ell}^s) \varepsilon R^s \qquad (5\text{-}25)$$

satisfies the condition (5-24). On the other hand, the result (5-25) is nothing else but the mean $\bar{L}$ of the multisample, i.e.:

$$L_o = \bar{L} \varepsilon R^s . \qquad (5\text{-}26)$$

Postulating a normal PDF, $N(\ell^j_o, S^*_j; \ell^j)$, for each component L^j of the multisample $\tilde{L}$, the multivariate PDF of the parent population can be written as:

$$\phi(\tilde{\ell}) = \prod_{j=1}^{s} N(\ell^j_o, S^*_j; \ell^j)$$

$$= \prod_{j=1}^{s} \frac{1}{S^*_j \sqrt{(2\pi)}} \exp\left[-\frac{(\ell^j - \ell^j_o)^2}{2S^{*2}_j}\right], \qquad (5\text{-}27)$$

where ℓ^j is the random variable having mean ℓ^j_o and standard deviation S^*_j. Following a similar procedure as in section 5.2, we end up again with with the discovery that the vector

$$L_o \equiv \bar{L} \equiv (\bar{\ell}^1, \bar{\ell}^2, \ldots, \bar{\ell}^s) \varepsilon R^s \qquad (5\text{-}28)$$

---

*) Trace of a matrix is the sum of its diagonal elements.

maximizes the probability that the members of the parent population will occur at the same places as the members of the multisample $\tilde{L}$.

Hence $\bar{L} \epsilon R^s$ is, under the above conditions,*) the maximum probable estimator for the mean $\tilde{\mu}$ of the postulated parent multivariate PDF, where

$$\tilde{\mu} = (\mu_1, \mu_2, \ldots, \mu_s) \; \epsilon R^s \; . \qquad (5\text{-}29)$$

5.5 Exercise 5

1. Prove that the mean μ of a continuous PDF, $\phi(x)$, defined as:

$$\mu = \int_{-\infty}^{\infty} x \; \phi \; (x) \; dx$$

minimizes the PDF variance σ^2, defined as:

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \; \phi(x) \; dx.$$

2. Prove that $\frac{\partial S^*}{\partial \ell^o} = 0$, is the necessary and sufficient condition for the rectangular (uniform) PDF, $R(\ell^o, S^*; \ell)$, to be the most probable underlying PDF for a sample L with mean $\bar{\ell}$ and variance S^2. Note that the analytic expression for the uniform PDF is given in example 3.17, section 3.2.5.

3. Prove that the same holds for the triangular PDF, $T(\ell^o, S^*; \ell)$, using its analytic expression given in example 3.18, section 3.2.5.

---

*) It can be shown that $\bar{L}$ is the maximum probability estimator of μ even when we postulate a statistically dependent multi dimensional PDF from a certain family of PDF's.

6. FUNDAMENTALS OF ADJUSTMENT CALCULUS

6.1 Primary and Derived Random Samples

So far, we have been dealing with random samples (multisamples) that had been obtained through some measurement or through any other data collecting process. These samples may all be regarded as primary or original random samples (multisamples).

In practice, we are often interested in other samples that would be derived from the primary samples by means of a computation of some kind. Such samples may be called derived random samples (multisamples).

From the philosophical point of view, there is not much difference between these two, since even the "primary" samples may be regarded as derived from the samples of physical influences or physical happenings. However, it is necessary to distinguish between them to be able to speak about the transition from one to the other.

6.2 Statistical Transformation, Mathematical Model

The transition from a primary to a derived sample (multisample) along with the associated variances and covariances may be called statistical transformation. We have already met two examples of such transformation although applied to random variable rather than sample (see sections 4.5 and 4.6), namely the transformation of the Gaussian PDF to the normal and to the standard normal PDF's, respectively.

Such statistical transformation may not always be as simple as in the above two cases. As a matter of fact, it may not be even possible to derive the sample at all from the primary sample which is usually

the case with multisamples. In other words, it might not be possible to express the derived sample explicitly in terms of the primary sample.

Let us consider a primary multisample $L \equiv (L^i)$, $i = 1, 2, \ldots, s$, that has s constituents. Each constituent $L^i = (\ell_k{}^i)$, $k = 1, 2, \ldots, n_i$, is a random sample on its own and represents a distinct physical quantity ℓ_i (i.e. the observations $\ell_k{}^i$, $k = 1, 2, \ldots, n_i$ are all representing the same physical quantity ℓ_i). Now, we may be interested in deriving a multisample X having n constituents, ie.

$$X = (X^j) \ , \ j=1, 2, \ldots, n,$$

from the original multisample L; noting again that each constituent X^j represents a distinct physical quantity x_j, $j = 1, 2, \ldots, n$. The formulae (relationships) relating the physical quantities ℓ and x, where

$$\left.\begin{aligned} \ell &= (\ell_1, \ell_2, \ldots, \ell_s), \\ \text{and} \qquad x &= (x_1, x_2, \ldots, x_n) \end{aligned}\right\} \qquad (6.1)^*$$

are called the mathematical model for the statistical transformation; and is usually expressed as:

$$\boxed{F(\ell, x) = 0} \qquad (6.2)$$

where F denotes the vector of functions f_i, $i = 1, 2, \ldots, r$ (having r components) that can be established between ℓ and x.

To be able to derive x from ℓ, the mathematical model (6.2) should be formulated as:

$$x = F(\ell), \qquad (6.3)$$

---

* Note that ℓ and x are nothing else but the multivariates corresponding to the multisamples L and X respectively.

which gives x as explicit function of ℓ.

Example 6.1: After having measured the two perpendicular edges a and b of a rectangular desk (see Figure 6.1), suppose that we are interested in learing something about the length of the diagonal d, and about the surface area α of this disk. In this case, the mathematical model will be written as:

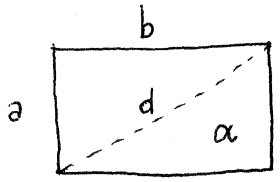


Figure 6.1

$$\mathbf{x} = F(\ell), \text{ where}$$

$$x = (x_1, x_2) = (d, \alpha), \text{ and}$$

$$\ell = (\ell_1, \ell_2) = (a, b) .$$

To derive the components of x from ℓ we write:

$$d = f_1 (a, b) = \sqrt{(a^2 + b^2)} , \text{ and}$$

$$\alpha = f_2 (a, b) = ab . \quad .$$

In vector notation, we can write:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} d \\ \alpha \end{bmatrix} = \begin{bmatrix} \sqrt{(a^2 + b^2)} \\ ab \end{bmatrix} .$$

The possibility of carrying out the statistical transformation depends basically on three factors:

(i) complexity of the mathematical model, i.e., the possibility of expressing x explicitly in terms of ℓ ($x = F(\ell)$);

(ii) "completeness" of the primary multisample L, i.e. whether all its constituents have the same number of elements in order to deduce the variance-covariance matrix Σ_L;

(iii) our willingness to match the individual s-tuples of elements from the primary multisample L with the n-tuples of elements from the derived multisample X, which creates much of a problem.

Particularly the last two factors are so troublesome that we usually do not even try to carry out the transformation and put up with some statistical estimates, i.e. representative values $\tilde{E}(X)$ and Σ_X for the derived multisample instead. To do so, we first evaluate $E(L)$ and Σ_L for the primary multisample L, from which we then compute the statistical estimates $\tilde{E}(X)$ and Σ_X for X.

According to the basic postulate of the error theory and to make the subsequent development easier, we generally postulate at this stage the PDF of the parent multivariate to the multisample L and assume

$$\tilde{E}(L) = \tilde{E}^*(\ell), \quad \Sigma_L = \Sigma^*_\ell, \quad \tilde{E}(X) = \tilde{E}^*(x), \text{ and } \Sigma_X = \Sigma^*_x \qquad (6.4)$$

in very much the same way as we postulated

$$\bar{\ell} = \mu_\ell \text{ and } S_L = \sigma_\ell$$

for the univariate case as discussed in section 4.7. This postulate allows us to work with continuous variables in the mathematical model and write it as:

$$F(L, X) = 0 \qquad (6.5)$$

understanding tacitly that each value X has its counterpart L.

From now on, we shall write $\bar{L}$ for $\tilde{E}(L)$, and $\hat{X}$ for the statistical estimate of X. Hence the mathematical model (6.5) becomes

$$F(\bar{L}, \hat{X}) = 0 , \qquad (6.6)$$

which consists of r functional relationships between $\bar{L}$ and $\hat{X}$.

From the point of view of the mathematical model $F(\bar{L}, \hat{X}) = 0$, the statistical transformation can be either solvable (if $s \geq n$) or unsolvable (if $s < n$). If it is solvable then we may still have <u>two</u> distinctly different cases:

(i) either the model yields only one solution $\hat{X}$ (when $r = s = n$) by using the usual mathematical tools, i.e., $\hat{X}$ is uniquely derived from $\bar{L}$;

(ii) or the mathematical model is overdetermined (when $r, s > n$) and cannot be resolved for $\hat{X}$ at all by using the ordinary mathematical tools, since an infinite number or different solutions for X can be found.

The first case we have met in example (6.1) where the determination of $\hat{X}$ from L does not present any problem from the statistical point of view. The only problem is to obtain Σ_X from L and Σ_L. This problem, known as propagation of errors, will be the topic of the next section.

If the model is overdetermined, or as we often say, if there are redundancies, (redundant or surplus observations) then the problem of transforming $(\bar{L}, \Sigma_L) \rightarrow (\hat{X}, \Sigma_X)$ constitutes the proper problem of adjustment.*)

6.3 Propagation of Errors

6.3.1 Propagation of Variance-Covariance Matrix, Covariance Law

The relationship between Σ_X and Σ_L for a mathematical model

$$F(\bar{L}, \hat{X}) = 0$$

is known as the propagation of variance-covariance matrix. Such relationship can be deduced explicitly only for explicit relations

$$\hat{X} = F(\bar{L}) \ .$$

To make things easier, let us deduce it first for one particular explicit relation, namely the linear relation between $\hat{X}$, and $\bar{L}$, i.e.

$$\boxed{\hat{X} = B\,\bar{L} + C}$$

---

* It has to be mentioned here that in practice we are in both cases working with $\Sigma_{\bar{L}}$ and $\Sigma_{\hat{X}}$, the variance-covariance matrices of $\bar{L}$ and X rather than Σ_L, Σ_X belonging to the samples L and X. The expressions for $\Sigma_{\bar{L}}$, $\Sigma_{\hat{X}}$ are derived in 6.4.4.

where B is indeed an n by s matrix composed of known elements *). Note that X is determined uniquely, as required. We want to establish the transition

$$\Sigma_L = \tilde{E}((L-\bar{L})\ (L-\bar{L})^T) \rightarrow \Sigma_X \text{ ,where } \bar{L} = E(L). \tag{6.8}$$

We can write:

$$\Sigma_X = \tilde{E}((X-\tilde{E}(X))\ (X-\tilde{E}(X))^T). \tag{6.9}$$

Here X = BL, and according to the postulate introduced in section 6.2 we can write:

$$\tilde{E}(X) = \tilde{E}(B\ L + C) = B\ \tilde{E}\ (L) + C = B\ \bar{L} + C.$$

Hence

$$\Sigma_X = \tilde{E}\ ((BL - B\bar{L})\ \ (BL - B\bar{L})^T)$$

$$= \tilde{E}\ (B(L - \bar{L})\ (B(L - \bar{L}))^T)$$

$$= \tilde{E}\ (B(L - \bar{L})\ (L - \bar{L})^T\ B^T)$$

$$= B\tilde{E}\ ((L - \bar{L})\ (L - \bar{L})^T)\ B^T = B(\Sigma_L)\ B^T,$$

$$\boxed{\Sigma_X = B\ \Sigma_L\ B^T.} \tag{6.10}$$

This formula (6.10) is known as the law of propagation of variance -- covariance matrix, or simply the covariance law.

---

*) This matrix B, which determines the linear relationship between X and L is sometimes called the "design matrix","the matrix of the coefficients" of the constituents of L in the linearized model, or simply the "coefficients matrix".

Example 6.2: Assume that the variance-covariance matrix of a given multisample $L = (\ell_1, \ell_2, \ell_3)$ was found to be

$$\Sigma_L = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 3 & 1 \\ 0 & 1 & 4 \end{bmatrix}.$$

If a multisample $X = (x_1, x_2)$ is to be derived from L according to the following relationships:

$$x_1 = \ell_1 - 3\ell_3,$$

$$x_2 = 2\ell_1 + \ell_2,$$

determine the variance-covariance matrix Σ_X of X.

It can be seen that the above relationships between the components of X and L are linear, and our mathematical model can be expressed as:

$$\underset{2,1}{X} = \underset{2,3}{B}\ \underset{3,1}{L},$$

i.e.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -3 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{bmatrix}.$$

This indicates that the coefficients matrix B is given by:

$$B = \begin{bmatrix} 1 & 0 & -3 \\ 2 & 1 & 0 \end{bmatrix}.$$

The variance-covariance matrix Σ_X of X is given by equation (6.10) i.e., in our case:

$$\underset{2,2}{\Sigma_X} = \underset{2,3}{B} \quad \underset{3,3}{\Sigma_L} \quad \underset{3,2}{B^T}$$

$$\Sigma_X = \begin{bmatrix} 1 & 0 & -3 \\ 2 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 3 & 2 & 0 \\ 2 & 3 & 1 \\ 0 & 1 & 4 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -3 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & -1 & -12 \\ 8 & 7 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -3 & 0 \end{bmatrix} = \begin{bmatrix} 39 & 5 \\ 5 & 23 \end{bmatrix}$$

$$\text{i.e.} \quad \Sigma_X = \begin{bmatrix} 39 & 5 \\ 5 & 23 \end{bmatrix}.$$

Now we shall show that the propagation of variance-covariance matrix can be deduced even for a more general case, namely the <u>non-linear</u> relation between X and $\bar{L}$, i.e.

$$X = F(L) \tag{6.11}$$

when F is a function with at least the first order derivative. Here we have to adopt another approximation yet. We have to <u>linearize</u> the relation (6.11) using, for instance, Taylor's series expansion around an approximate value L^o for L .

$$X = F(L^o) + \frac{dF}{dL}\Big|_{L = L^o} (L - L^o) + \text{higher order terms},$$

where

$$\frac{dF}{dL}\Big|_{L = L^o} (L - L^o) = \sum_{i=1}^{s} \frac{\partial F}{\partial \ell_i}\Big|_{\ell_i = \ell_i^o} (\ell_i - \ell_i^o) .$$

Taking the first two terms only, which is permissible when the values of the elements in Σ_L are much smaller than the values of ℓ_i, we can write:

$$X \doteq F(L^\circ) + B(L - L^\circ) \tag{6.12}$$

where B is again an n by s matrix but this time composed from all the partial derivatives $\left.\frac{\partial x_i}{\partial \ell_j}\right|_{\ell_j = \ell_j^\circ}$ *). Applying the expectation operator we obtain realizing that $\tilde{E}(F(L^\circ)) = F(L^\circ)$ and $\tilde{E}(L^\circ) = L^\circ$ (because L° is a selected vector of constant values) :

$$\begin{aligned} \tilde{E}(X) &\doteq \tilde{E}(F(L^\circ) + B(L - L^\circ)) \\ &= F(L^\circ) + B(\tilde{E}(L) - L^\circ). \end{aligned} \tag{6.13}$$

Subtracting (6.13) from (6.12) we get:

$$X - \tilde{E}(X) \doteq B(L - \tilde{E}(L)) = B(L - \bar{L}) \tag{6.14}$$

and we end up again with

$$\Sigma_X \doteq B\Sigma_L B^T, \tag{6.15}$$

realizing that $\Sigma_{X-\tilde{E}(X)} = \Sigma_X$.

---

* Explicitly, if we have

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} x_1(\ell_1, \ell_2, \ldots, \ell_s) \\ x_2(\ell_1, \ell_2, \ldots, \ell_s) \\ \\ x_n(\ell_1, \ell_2, \ldots, \ell_s) \end{bmatrix}$$

then the matrix B will take the form:

$$B_{n,s} = \begin{bmatrix} \frac{\partial x_1}{\partial \ell_1} & \frac{\partial x_1}{\partial \ell_2} & \cdots & \frac{\partial x_1}{\partial \ell_s} \\ \frac{\partial x_2}{\partial \ell_1} & \frac{\partial x_2}{\partial \ell_2} & \cdots & \frac{\partial x_2}{\partial \ell_s} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_n}{\partial \ell_1} & \frac{\partial x_n}{\partial \ell_2} & \cdots & \frac{\partial x_n}{\partial \ell_s} \end{bmatrix}.$$

Hence the linear case may be regarded as one particular instance (special case) of the more general explicit relation, yielding therefore the same law for the propagation of variance-covariance matrix, i.e., the same covariance law. It should be noted that the physical units of the individual elements of both matrices B and Σ_L must be considered and selected in such a way to give the required units of the matrix Σ_X .

Example 6.3: Let us take again the example 6.1 and form the variance covariance matrix Σ_X for the diagonal d and the area α of the desk in question. We have:

$$\Sigma_L = \begin{bmatrix} S_a^2 & S_{ab} \\ S_{ab} & S_b^2 \end{bmatrix}$$

and the model is non-linear, although explicit, i.e.

$X = F(L)$, or $(d, \alpha) = F(a, b)$.

We have to linearize it as follows:

$$X = (d, \alpha) = (d^o, \alpha^o) + B\,[(a, b) - (a^o, b^o)],$$

where $(d^o, \alpha^o) = F(a^o, b^o)$, and

$$B = \begin{bmatrix} \frac{\partial d}{\partial a} & \frac{\partial d}{\partial b} \\ \frac{\partial \alpha}{\partial a} & \frac{\partial \alpha}{\partial b} \end{bmatrix}.$$

Here:

$$\frac{\partial d}{\partial a} = \frac{1}{2} \frac{1}{\sqrt{(a^2+b^2)}}\, 2a = \frac{a}{d}\,, \quad \frac{\partial d}{\partial b} = \frac{b}{d}\,, \quad \frac{\partial \alpha}{\partial a} = b, \quad \frac{\partial \alpha}{\partial b} = a\,.$$

Hence, the matrix B in this case takes the form:

$$B_{2,2} = \begin{bmatrix} a/d & b/d \\ b & a \end{bmatrix}$$

and by applying the covariance law (equation (6.15)) we get:

$$\Sigma_X = \begin{bmatrix} S_d^2 & S_{d\alpha} \\ S_{\alpha d} & S_\alpha^2 \end{bmatrix} = B\,\Sigma_L\,B^T$$

$$= \begin{bmatrix} a/d & b/d \\ b & a \end{bmatrix} \begin{bmatrix} S_a^2 & S_{ab} \\ S_{ba} & S_b^2 \end{bmatrix} \begin{bmatrix} a/d & b \\ b/d & a \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{d^2}(a^2 S_a^2 + 2abS_{ab} + b^2 S_b^2), & \frac{1}{d}(ab(S_a^2 + S_b^2) + (a^2 + b^2))\,Sab) \\ \frac{1}{d}(ab(S_a^2 + S_b^2) + (a^2 + b^2)\,S_{ab}), & b^2 S_a^2 + 2abS_{ab} + a^2 S_b^2 \end{bmatrix}.$$

Example 6.4: Let us assume that the primary multisample $L \equiv (a, b)$ which we have dealt with in Examples 6.1 and 6.3 is given by:

$L = \{a, b\} = \{(128.1, 128.1, 128.2, 128.0, 128.1), (62.5, 62.7, 62.6, 62.6, 62.5)\}$, in **centimetres.**

Accordingly, the statistical estimate of the derived quantities will be

$$\hat{X} = \begin{bmatrix} \hat{d} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} \sqrt{[(\bar{a})^2 + (\bar{b})^2]} \\ \bar{a} \quad \bar{b} \end{bmatrix}$$

where $\bar{a}$ and $\bar{b}$ are the estimates (means) of the two **measured** sides of the desk. From the given data we get

$\bar{a} = 128.1$ cm. and $\bar{b} = 62.58$ cm.

Hence

$$\hat{X} = \begin{bmatrix} \hat{d} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} \sqrt{[(128.1)^2 + (62.58)^2]} \\ (128.1) \cdot (62.58) \end{bmatrix}$$

$$= \begin{bmatrix} 142.57 \text{ cm} \\ 8016.50 \text{ cm}^2 \end{bmatrix}.$$

After computing the variance-covariance matrix Σ_L we get

$$\Sigma_L = \begin{bmatrix} 0.004 & 0 \\ 0 & 0.0056 \end{bmatrix} \text{ cm}^2$$

which indicates that the constituents a and b are being taken as statistically independent.

Evaluating the elements of the B matrix (as given in Example 6.3) we get:

$$B = \begin{bmatrix} \frac{\bar{a}}{\hat{d}} & \frac{\bar{b}}{\hat{d}} \\ \bar{b} & \bar{a} \end{bmatrix} = \begin{bmatrix} 0.898 & 0.439 \\ 62.58 & 128.1 \end{bmatrix}$$

in which the elements of the first row are unitless, and of the second row are in cm.

Finally Σ_X is computed as follows:

$$\Sigma_X = B \, \Sigma_L \, B^T$$

$$= \begin{bmatrix} 0.898 & 0.439 \\ 62.58 & 128.1 \end{bmatrix} \begin{bmatrix} 0.004 & 0 \\ 0 & 0.0056 \end{bmatrix} \begin{bmatrix} 0.898 & 62.58 \\ 0.439 & 128.1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0043 & 0.5397 \\ 0.5397 & 107.5627 \end{bmatrix}, \text{ with units } \begin{bmatrix} \text{cm}^2 & \text{cm}^3 \\ \text{cm}^3 & \text{cm}^4 \end{bmatrix}.$$

Furthermore

$$S_d = \sqrt{(0.0043)} = \underline{0.066 \text{ cm.}},$$

$$S_\alpha = \sqrt{(107.5627)} = \underline{10.37 \text{ cm}^2}.$$

6.3.2 Propagation of Errors, Uncorrelated Case

If X contains one component only, i.e. x, the matrix B in the formulae (6.10) or (6.15) degenerates into a 1 by s matrix, i.e. into a row vector $B \equiv [B_1, B_2, \ldots, B_s]$, and

$$\Sigma_X = B\Sigma_L B^T$$

becomes a quadratic form which has dimensions 1 by 1. Then

$$\Sigma_X = S_x^2, \tag{6.16}$$

If, moreover, L is assumed uncorrelated, we have

$$\Sigma_L = \text{diag}\,(S_{\ell_1}^2, S_{\ell_2}^2, \ldots, S_{\ell_s}^2), \tag{6.17}$$

which is a diagonal matrix, and we can write

$$\boxed{S_x^2 = \sum_{i=1}^{s} B_i^2 S_{\ell_i}^2, \quad B_i = \frac{\partial F}{\partial \ell_i} = \frac{\partial x}{\partial \ell_i}} \quad . \tag{6.18}$$

This formula is known as the law of propagation of MSE's or simply the law of propagation of errors. The law of propagation of errors is hence nothing else, but a special case of the propagation of variance-covariance matrix.

The law of propagation of errors has many applications in surveying practice as well as in many other experimental sciences.

Example 6.5: In figure 6.2, we assume a plane triangle in which the angles α and β whose estimated values are:

$\bar{\alpha} = 32^\circ\ 15'\ 20''$, with $S_\alpha = 4''$,

$\bar{\beta} = 75^\circ\ 43'\ 32''$, with $S_\beta = 3''$, are observed .

Also, assume that α and β are independent, i.e. $S_{\alpha\beta} = 0$.

Let us estimate the third angle γ, along with its standard error S_γ, as follows:

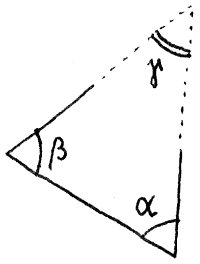


Figure 6.7

$$\hat{\gamma} = 180^\circ - (\bar{\alpha} + \bar{\beta}) = 72^\circ\ 1'\ "08 \ ,$$

$$S_\gamma^2 = (\frac{\partial\gamma}{\partial\alpha})^2\ S_\alpha^2 + (\frac{\partial\gamma}{\partial\beta})^2\ S_\beta^2$$

$$= (-1)^2 \cdot (4)^2 + (-1)^2 \cdot (3)^2 = 16 + 9 = 25 \ ,$$

that is: $S_\gamma = 5"$.

Example 6.6: Figure 6.3 shows a levelling line between two bench marks A, C, with observed level differences h_i of the individual sections with length ℓ_i, i = 1, 2, . . ., s. Assume that all the h_i's are uncorrelated and the MSE of h_i is proportional to ℓ_i, i.e. $S_{h_i}^2 = k\,\ell_i$, where k is a constant.

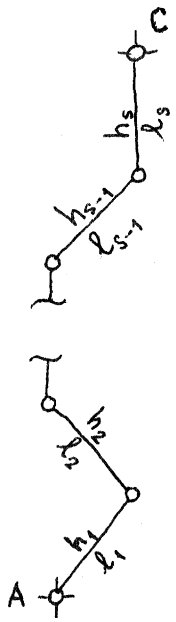


Figure 6.3

Let us deduce the expression for the MSE of the overall level difference ΔH between A and C where:

$$\Delta H = H_C - H_A = \sum_{i=1}^{s} h_i \ .$$

The mathematical model in this case is

$$\Delta H = h_1 + h_2 + h_3 + \ldots + h_s \ .$$

Hence:

$$S_{\Delta H}^2 = (\frac{\partial\Delta H}{\partial h_1})^2\ S_{h_1}^2 + (\frac{\partial\Delta H}{\partial h_2})^2\ S_{h_2}^2 + \ldots$$

$$= (1)^2\ (k\ell_1) + (1)^2\ (k\ell_2) + \ldots$$

$$= \sum_{i=1}^{s} k\,\ell_i = k \sum_{i=1}^{s} \ell_i ,$$

which means that the MSE of ΔH equals to the constant of proportionality k multiplied by the total (overall) length of the levelling line A – C.

Let us consider the example 6.3 and assume that the errors in a, b are uncorrelated i.e. $S_{ab} = 0$ as we did in Example 6.4. Then we can treat d and α separately (if we are interested in their individual MSE's alone) and we get by applying the law of propagation of errors:

$$S_d^2 = \left(\frac{\partial d}{\partial a}\right)^2 S_a^2 + \left(\frac{\partial d}{\partial b}\right)^2 S_b^2 = \frac{1}{a^2}\left(a^2 S_a^2 + b^2 S_b^2\right) ,$$

$$S_\alpha^2 = \left(\frac{\partial \alpha}{\partial a}\right)^2 S_a^2 + \left(\frac{\partial \alpha}{\partial b}\right)^2 S_b^2 = b^2 S_a^2 + a^2 S_b^2 \quad .$$

Note that the same results can be obtained from Example 6.3 immediately by putting $S_{ab} = 0$.

On the other hand, if we are interested in the covariance $S_{d\alpha}$ between the two derived quantities d and α, we have to apply the covariance law (equation 6.15) and we will end up with

$$S_{d\alpha} = \frac{ab}{d}\left(S_a^2 + S_b^2\right) ,$$

that is $S_{d\alpha} \neq 0$, and Σ_X $(X \equiv (d, \alpha))$ is not a diagonal matrix, even though the Σ_L of the primary multisample is diagonal i.e. $S_{ab} = 0$, see the results obtained in Example 6.4 . This is a very important discovery and should be taken into consideration when using the derived multisample $X \equiv (d, \alpha)$ for any further treatment in which case we cannot assume that d and α are uncorrelated any more and we must take the entire Σ_X into account.

Example 6.7: Let us solve Example 6.2 again, but this time we will consider the primary multisample $L \equiv (\ell_1, \ell_2, \ell_3)$ as uncorrelated and its Σ_L is:

$$\Sigma_L = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix} = \text{diag } (3, 3, 4).$$

From example 6.2 we have:

$$B = \begin{bmatrix} 1 & 0 & -3 \\ 2 & 1 & 0 \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Hence:

$$\Sigma_X = B\Sigma_L B^T,$$

$$\Sigma_X = \begin{bmatrix} 1 & 0 & -3 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -3 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 39 & 6 \\ 6 & 15 \end{bmatrix} = \begin{bmatrix} S^2_{x_1} & S_{x_1x_2} \\ S_{x_2x_1} & S^2_{x_2} \end{bmatrix},$$

which again verifies the fact that even when Σ_L is diagonal the Σ_X is not.

On the other hand we can treat x_1 and x_2 separately by using the law of propagation of errors (since L is uncorrelated) to get $S^2_{x_1}$ and $S^2_{x_2}$ separately; for instance,

$$S^2_{x_1} = \left(\frac{\partial x_1}{\partial \ell_1}\right)^2 S^2_{\ell_1} + \left(\frac{\partial x_1}{\partial \ell_2}\right)^2 S^2_{\ell_2} + \left(\frac{\partial x_1}{\partial \ell_3}\right)^2 S^2_{\ell_3}$$

$$= (1)^2 (3) + (0)^2 (3) + (-3)^2 (4)$$

$$= 3 + 0 + 36 = 39,$$

which is the same value as we got by applying the covariance law above.

Example 6.8: To determine the two sides AC = z and BC = y of the plane triangle shown in Figure 6.4, the length AB = x along with the two horizontal angles α and β were observed and their estimates were found to be:

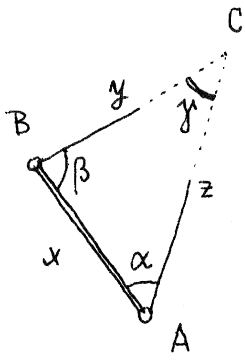


Figure 6.4

$$\bar{x} = 10 \text{ m, with } S_x = 3 \text{ cm},$$

$$\bar{\alpha} = 90^\circ \text{ , with } S_\alpha = 2'' \text{ ,}$$

$$\bar{\beta} = 45^\circ \text{ , with } S_\beta = 4'' \text{ ,}$$

$$S_{\alpha\beta} = -1 \text{ arc sec}^2 \text{ and } S_{x\alpha} = S_{x\beta} = 0 .$$

It is required to compute the statistical estimates for y and z along with their associated variance-covariance matrix Σ_X in cm^2, where

$$X = (y, z).$$

First, we establish the mathematical model which relates the primary and derived samples, i.e.,

$$X = F(L), \text{ where}$$

$$L = (\ell_1, \ell_2, \ell_3) = (\alpha, \beta, x) ,$$

$$X = (x_1, x_2) = (y, z) .$$

From the sine law of the given triangle we get:

$$\frac{y}{\sin \alpha} = \frac{z}{\sin \beta} = \frac{x}{\sin \gamma} \quad ;$$

however the angle γ is not observed, i.e. it is not an element of the primary sample, therefore we have to substitute for it in terms of the observed quantities, say α and β by putting

$$\sin \gamma = \sin (\alpha + \beta),$$

and we get:

$$y = x \frac{\sin \alpha}{\sin (\alpha + \beta)} \quad ,$$

$$z = x \frac{\sin \beta}{\sin (\alpha + \beta)} \quad .$$

By substituting for α, β and x, we get

$$\hat{y} = 10 \sqrt{2} = 10 \ (1.414) = \underline{14.14} \text{ m} \quad ,$$

$$\hat{z} = \underline{10 \text{ m.}}$$

Our mathematical model then can be written as:

$$X = \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} y \ (\alpha, \beta, x) \\ z \ (\alpha, \beta, x) \end{bmatrix} = \begin{bmatrix} x \sin \alpha / \sin (\alpha + \beta) \\ x \sin \beta / \sin (\alpha + \beta) \end{bmatrix} .$$

To compute $\Sigma_X = B \ \Sigma_L \ B^T$, we have to evaluate the matrix B which is of the form

$$B = \begin{bmatrix} \frac{\partial y}{\partial \alpha} , & \frac{\partial y}{\partial \beta} , & \frac{\partial y}{\partial x} \\ \\ \frac{\partial z}{\partial \alpha} , & \frac{\partial z}{\partial \beta} , & \frac{\partial z}{\partial x} \end{bmatrix} ,$$

$$= \begin{bmatrix} \frac{z}{\sin(\alpha + \beta)} , & \frac{-y}{\tan (\alpha + \beta)} , & \frac{y}{x} \\ \\ \frac{-z}{\tan (\alpha + \beta)} , & \frac{y}{\sin (\alpha + \beta)} , & \frac{z}{x} \end{bmatrix} .$$

From the given data, the matrix Σ_L takes the form

$$\Sigma_L = \begin{bmatrix} S_\alpha^2 & S_{\alpha\beta} & S_{\alpha x} \\ S_{\beta\alpha} & S_\beta^2 & S_{\beta x} \\ S_{x\alpha} & S_{x\beta} & S_x^2 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & -1 & 0 \\ -1 & 16 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

(It is very important to maintain the same sequence of the elements of the primary sample in both matrices B and Σ_L to give a meaningful Σ_X.)

Now matching the units of the individual elements of B and Σ_L, keeping in mind that Σ_X is required in cm^2, results in scaling the B matrix to

$$B = \begin{bmatrix} \frac{z\,(100)}{\rho''\,\sin(\alpha+\beta)} & , & \frac{-y(100)}{\rho''\tan(\alpha+\beta)} & , & \frac{y}{x} \\ \\ \frac{-z(100)}{\rho''\tan(\alpha+\beta)} & , & \frac{y(100)}{\rho''\sin(\alpha+\beta)} & , & \frac{z}{x} \end{bmatrix}$$

where $\rho'' = 206265 \doteq 2.10^5$ arc sec.

Evaluating the elements of the above B matrix we get:

$$B = \begin{bmatrix} 0.007 & 0.007 & 1.414 \\ 0.005 & 0.010 & 1.000 \end{bmatrix}$$

and consequently

$$\Sigma_X = \begin{bmatrix} .007 & .007 & 1.414 \\ .005 & .010 & 1.000 \end{bmatrix} \begin{bmatrix} 4 & -1 & 0 \\ -1 & 16 & 0 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} .007 & .005 \\ .007 & .010 \\ 1.414 & 1.000 \end{bmatrix} ,$$

i.e.

$$\Sigma_X = \begin{bmatrix} 18.0009 & 12.7272 \\ 12.7272 & 9.0016 \end{bmatrix} \doteq \begin{bmatrix} 18 & 13 \\ 13 & 9 \end{bmatrix} \text{ cm}^2 ,$$

and

$$S_y = \sqrt{18} = 4.2 \text{ cm} ,$$

$$S_z = \sqrt{9} = 3 \text{ cm}.$$

The results of the above example show that the high precision in measuring the angles α and β has insignificant effect on the estimated standard errors of the derived y and z lengths as compared to the effect of the precision of the measured length x. Hence, one can use the error propagation to detect the main deciding factors in the primary sample on the accuracy of the derived quantities and decide on the needed accuracy of the observations. This process is usually known as pre-analysis which is done before taking any actual measurements by using very approximate values for the observed quantities. This results in accepting specifications concerning the observations techniques to achieve the required accuracy. Some more details about it are given in section 6.3.5.

6.3.3 Propagation of Non-Random Errors, Propagation of Total Errors

The idea of being able to foretell the expected magnitude of the MSE (as a measure of random errors) of a function of observations - this is essentially what the law of propagation of errors is all about - is often extended to non-random errors. These non-random errors are sometimes called systematic errors, for which the law governing their behaviour is not known. Hence, the values of such non-random errors used in the subsequent development are rather hypothesized (postulated) for the analysis and specification purposes.

The problem may be now stated as follows: let us have an explicit mathematical model

$$x = f(L) , \qquad (6\text{-}19)$$

in which x is a single quantity, f is a single-valued function and $L \equiv (\ell_1, \ell_2, \ldots, \ell_s)$ is the vector of the different observed quantities that are assumed to be uncorrelated. We are seeking to determine the influence of small, non-random errors $\delta\ell_i$ in each observation ℓ_i on the result x. This influence will be denoted by δ_x.

The problem is readily solved using again the truncated Taylor's series expansion, around the approximate values $L^o = (\ell_1^o, \ell_2^o, \ldots, \ell_s^o)$, from which we get:

$$x \doteq f(L^o) + \left.\frac{\partial f}{\partial L}\right|_{L = L^o} (L - L^o)$$

$$= x^o + \sum_{i=1}^{s} \left.\frac{\partial f}{\partial \ell_i}\right|_{\ell_i = \ell_i^o} (\ell_i - \ell_i^o) . \qquad (6\text{-}20)$$

By substituting $\delta\ell_i$ for $(\ell_i - \ell_i^o)$ and δ_x for $(x - x^o)$ in equation (6-20) we get:

$$\boxed{\delta_x = \sum_{i=1}^{s} \left.\frac{\partial f}{\partial \ell_i}\right|_{\ell_i = \ell_i^o} \delta\ell_i \text{ *)}\,,} \qquad (6\text{-}21)$$

which is the formula for the propagation of non-random error.

Note in formula (6-21), the signs of both the partial derivatives $(\frac{\partial f}{\partial \ell i})$ and the non-random errors $\delta\ell_i$, have to be considered. (Compare this to formula (6-18).)

We may also ask what incertitude can we expect in x if the observations ℓ_i are burdened with both random and non-random errors. In such a case we define the total error as:

$$\boxed{T = \sqrt{(\delta^2 + S^2)}} \qquad (6\text{-}22)$$

with δ being the non-random error and S being the MSE. Combining the two errors in x as given above and using equations (6-18) and (6-21) we get:

$$T_x = \sqrt{\left[\left(\sum_{i=1}^{s} \frac{\partial f}{\partial \ell_i}\,\delta\ell_i\right)^2 + \sum_{i=1}^{s} \left(\frac{\partial f}{\partial \ell_i}\right)^2 S_i^2\right]}$$

$$= \sqrt{\left[\sum_{i=1}^{s}\left\{\left(\frac{\partial f}{\partial \ell_i}\right)^2 (\delta\ell_i^2 + S_i^2)\right\} + \sum_{\substack{j=1 \\ i \neq j}}^{s} \frac{\partial f}{\partial \ell_i}\,\frac{\partial f}{\partial \ell_j}\,\delta\ell_i\,\delta\ell_j\right]}$$

or

$$T_x = \sqrt{\left[\sum_{i=1}^{s} \left(\frac{\partial f}{\partial \ell_i}\right)^2 T_i^2 + q\right]}\,, \qquad (6\text{-}23)$$

where q may be regarded as a kind of "covariance" between individual non-random errors, and T_i is the total error in the observation ℓ_i.

---

* For the validity of the Taylor's series expainsion, we can see that the requirement of $\delta\ell_i$ being small in comparison to ℓ_i is obviously essential.

As we mentioned in section 4.2, the non-random (systematic) errors may be known or assumed functions of some parameters. In this case their influence δ_x on x can be also expressed as a function of the same parameters.

Example 6.9: Let us solve Example 6.2 again considering the primary multisample $L \equiv (\ell_1, \ell_2, \ell_3)$ to be uncorrelated with variance-covariance matrix:

$$\Sigma_L = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix} ,$$

and having also non-random (systematic) errors given as:

$\delta L = (\delta\ell_1, \delta\ell_2, \delta\ell_3) = (-1.5, 2, 0.5)$, in the same units as the given standard errors.

It is required to compute the total error in the derived quantities: x_1 and x_2 according to the mathematical model given in Example 6.2.

The total errors are given by equation (6-22) as:

$$T_{x_1} = \sqrt{(\delta^2_{x_1} + S^2_{x_1})} ,$$

$$T_{x_2} = \sqrt{(\delta^2_{x_2} + S^2_{x_2})} .$$

We have:

$$S^2_{x_1} = \sum_{i=1}^{3} \left(\frac{\partial x_1}{\partial \ell_i}\right)^2 S^2_{\ell_i} = 39 ,$$

$$S^2_{x_2} = \sum_{i=1}^{3} \left(\frac{\partial x_2}{\partial \ell_i}\right)^2 S^2_{\ell_i} = 15 .$$

The influences δ_{x_1} and δ_{x_2} due to the given non-random errors in L are computed from equation (6-21) as follows:

$$\delta_{x_1} = \sum_{i=1}^{3} \frac{\partial x_1}{\partial \ell_i} \delta\ell_i$$

$$= (1)(-1.5) + (0)(2) + (-3)(0.5)$$

$$= -1.5 + 0 - 1.5 = -3 ,$$

$$\delta_{x_2} = \sum_{i=1}^{3} \frac{\partial x_2}{\partial \ell_i} \delta\ell_i$$

$$= (2)(-1.5) + (1)(2) + (0)(0.5)$$

$$= -3 + 2 + 0 = -1 .$$

Hence, the required total errors will be:

$$T_{x_1} = \sqrt{[(-3)^2 + 39]} = \sqrt{[48]} \doteq \underline{6.93} ,$$

$$T_{x_2} = \sqrt{[(-1)^2 + 15]} = \sqrt{[16]} = \underline{4}.$$

Example 6.10: Consider again Example 6.6. In addition to the given information, assume that each height difference h_i has got a non-random (systematic) error expressed as $\delta_{h_i} = k'h_i$, where k' is another constant, a constant of proportionality between h_i and δ_{h_i}. Determine the total error in ΔH where

$$\Delta H = H_C - H_A = \sum_{i=1}^{s} h_i = h_1 + h_2 + \ldots + h_s .$$

The total error in ΔH is given by:

$$T_{\Delta H} = \sqrt{(\delta^2_{\Delta H} + S^2_{\Delta H})} .$$

In Example 6.6, we found that:

$$S^2_{\Delta H} = k\, \ell_{AC} \quad ,$$

where k was a constant and $\ell_{AC} = \sum_{i=1}^{s} \ell_i$ is the entire length of the levelling line AC.

We can now compute $\delta_{\Delta H}$ as follows:

$$\delta_{\Delta H} = \sum_{i=1}^{s} \frac{\partial \Delta H}{\partial h_i}\, \delta h_i \quad ,$$

where

$$\frac{\partial \Delta H}{\partial h_1} = \frac{\partial \Delta H}{\partial h_2} = \ldots = \frac{\partial \Delta H}{\partial h_s} = 1 \quad ,$$

and

$$\delta_{h_i} = k' h_i$$

Then we get

$$\delta_{\Delta H} = \sum_{i=1}^{s} k' h_i = k' \sum_{i=1}^{s} h_i = k'\, \Delta H .$$

Finally, the expression for the total error in ΔH will be:

$$T_{\Delta H} = \sqrt{[k'^2\, \Delta H^2 + k\, \ell_{AC}]}.$$

6.3.4 Truncation and Rounding

In any computation we have to represent the numbers we work with, which may be either irrational like π, e, √2, or rational with very many decimal places like 1/3, 5/11, etc., by rational numbers with a fixed number of figures.

The representation can be made in basically two different ways. We either truncate the original number after the required number

of figures or we round off the original number to the required length.

The first process can be mathematically described as:

$$\boxed{a \doteq a_T = \text{Int } (a \cdot 10^n)/10^n} \qquad (6\text{-}24)$$

where a is the original number assumed normalized*), n is the required number of decimal places and Int stands for the integer value.

Example 6.11: $\pi = 3.141592 \ldots\ldots$, $n = 3$ and we get:

$$\pi \doteq \pi_T = \text{Int } (\pi \cdot 10^3)\ 10^{-3}$$

$$= \text{Int } (3141.592 \ldots)\ 10^{-3}$$

$$= 3141 \cdot 10^{-3}$$

$$= \underline{3.141.}$$

The second process, i.e. the rounding-off, can be described by the formulae:

$$\boxed{a \doteq a_R = \text{Int } (a \cdot 10^n + 0.5)/10^n} \qquad (6\text{-}25)$$

in which all terms are as described above.

Example 6.12: π, $n = 3$ and we get:

$$\pi \doteq \pi_R = \text{Int } (\pi \cdot 10^3 + 0.5)\ 10^{-3}$$

$$= \text{Int } (3141.592 \ldots + 0.5)\ 10^{-3}$$

$$= \text{Int } (3142.092 \ldots)\ 10^{-3}$$

$$= 3142 \cdot 10^{-3}$$

$$= \underline{3.142.}$$

It can be seen that the errors involved in the above two alternative processes differ. Denoting the error in "a" due to

---

* To normalize the number, say 3456.21, we write it in the form $3.45621 \cdot 10^3$.

truncation by δ_{a_T} and the error due to rounding by δ_{a_R}, we get:

$$\delta_{a_T} = a - a_T \in [0,\ 10^{-n})$$

$$\delta_{a_R} = a - a_R \in [-0.5\ 10^{-n},\ 0.5\ 10^{-n})$$

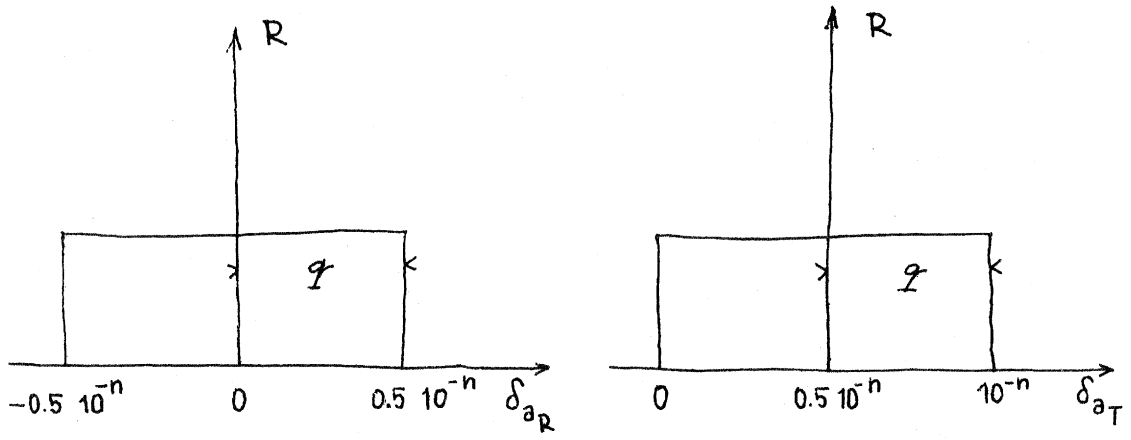
and we may postulate that δ_{a_T} has a parent random variable distributed according to the rectangular (uniform) PDF (see section 3.2.5):

$$R(0.5\ 10^{-n},\ \sigma;\ \delta_{a_T}) \qquad (6\text{-}26)$$

while δ_{a_R} has parent PDF:

$$R(0,\ \sigma;\ \delta_{a_R}) \qquad (6\text{-}27)$$

as shown in Figure 6.5.



PDF of rounding errors

PDF of truncation errors

Figure 6.5

From example 3.17, section 3.2.5 we know that $\sigma = q/\sqrt{3}$, where q equals half of the width of the R. In our case, obviously $q = 0.5\ 10^{-n}$ so that $\sigma \doteq 0.289\ 10^{-n}$.

Because of their different means, the error in truncation propagates according to the "total error law" and the errors in rounding propagates according to the "random error law". Hence, if we have a number x:

$$x = f(L) \ , \tag{6-28}$$

where

$$L = (\ell_i) \ , \quad i = 1, 2, \ldots, s$$

is a set of s numbers to be either truncated or rounded off individually, we can write the formulae for the errors in x due to truncation and rounding errors in the individual ℓ_i's as follows:

$$\delta_{x_T} = \sqrt{\left[\left(\sum_{i=1}^{s} \frac{\partial f}{\partial \ell_i} \quad 0.5 \cdot 10^{-n}\right)^2 + \sum_{i=1}^{s} \left(\frac{\partial f}{\partial \ell_i}\right)^2 \frac{1}{12}\ 10^{-2n}\right]}, \tag{6-29}$$

$$\delta_{x_R} = \sqrt{\left[\sum_{i=1}^{s} \left(\frac{\partial f}{\partial \ell_i}\right)^2 \quad \frac{1}{12}\ 10^{-2n}\right]} \ . \tag{6-30}$$

This indicates clearly that the error in x due to the rounding process is less than the corresponding error due to truncation; and this is why we always prefer to work with rounding rather than truncation.

Example 6.13: Let us determine the expected error in the sum x of a thousand numbers a_i, $x = \sum_{i=1}^{1000} a_i$, if

(i) the individual values a_i were truncated to five decimal places;

(ii) the individual values a_i were rounded-off to five decimal places.

Solution:

(i) The error δ_{x_T} due to the truncation of individual a_i is computed from equation (6-29) as follows:

$$\delta_{x_T} = \sqrt{\{[\sum_{i=1}^{1000} \frac{\partial x}{\partial a_i} (0.5 \cdot 10^{-5})]^2 + \sum_{i=1}^{1000} (\frac{\partial x}{\partial a_i})^2 (\frac{1}{12} \cdot 10^{-10})\}}$$

$$= \sqrt{\{[0.5 \cdot 10^{-5} \cdot 10^3]^2 + \frac{1}{12} \cdot 10^{-10} \cdot 10^3\}}$$

$$= \sqrt{\{\frac{1}{4} \cdot 10^{-4} + \frac{1}{12} \cdot 10^{-7}\}}$$

$$= \sqrt{\{10^{-8} (2500 + 0.833)\}}$$

$$\doteq 0.005001 \doteq \underline{0.005} .$$

(ii) The error δ_{x_R} due to the rounding of individual a_i is computed from equation (6-30) as follows:

$$\delta_{x_R} = \sqrt{\{\sum_{i=1}^{1000} (\frac{\partial x}{\partial a_i})^2 (\frac{1}{12} \cdot 10^{-10})\}}$$

$$= \sqrt{\{(1000) (\frac{1}{12} \cdot 10^{-10})\}}$$

$$= \sqrt{\{10^{-8} (0.833)\}}$$

$$\doteq \underline{0.000091} ,$$

which is much smaller than the corresponding δ_{x_T} .

6.3.5 Tolerance Limits, Specifications and Preanalysis

Another importantant application of the propagation laws for errors is the determination of specifications for a certain experiment when the maximum tolerable errors of the results, which are usually called tolerance limits, are known beforehand. Such process is known as pre-analysis. The set-up of the specifications should therefore result in the proper design of the experiment, i.e. the choice of observation techniques, instrumentation, etc., to meet the permissible tolerance limits.

The specifications for the elementary processes should account for both the random and the inevitable non-random (systematic) errors. This is, unfortunately, seldom the case in practice. It is usual to require that the specifications are prescribed in such a way as to meet the tolerance limits with the probability of approximately 0.99. If we hence expect the random errors to have the parent Gaussian PDF, the actual results should not have the total error, composed of the non-random error δ and 2.5 to 3 times the RMS , which corresponds to probability of 99% , larger than the prescribed tolerance limits, i.e.

$$\boxed{T \leq \sqrt{\{\delta^2 + (3\sigma)^2\}}.} \qquad (6\text{-}31)$$

Example 6.14: Assume that we want to measure a distance D = 1000 m, with a relative error (see 4.10) not worse than 10^{-4}, using a 20 m tape which had been compared to the "standard" with a precision not better than $3\sigma < 1$ mm , i.e. tolerance limits of the comparison were $\pm$ 1 mm . Assume also that the whole length D is divided into 50 segments d_i, i = 1, 2, ..., 50,

each of which is approximately 20 m. Providing that each segment d_i will be measured only twice: forward F_i and backward B_i, what differences can we tolerate (accept or permit) between the back and forth measurements of each segment?

Solution:

The tolerance limits in D, i.e. the permissible total error in D, is given by

$$T_D = 1000\,\text{m} \;\; 10^{-4} = 0.10\ \text{m} = \underline{10\ \text{cm.}}$$

This total error T_D is given by

$$T_D = \sqrt{\{\delta_D^2 + (3\sigma_D)^2\}} \quad ,$$

where δ_D is the non-random (systematic) error in D, σ_D is the random error in D and the factor 3 is used to get probability > 99% according to the assumed Gaussian PDF. Knowing that

$$D = \sum_{i=1}^{50} d_i, \quad \text{where } d_i = \frac{1}{2}(F_i + B_i) \ ,$$

we get:

$$\delta_D = \sum_{i=1}^{50} \frac{\partial D}{\partial d_i} \; \delta d_i \ ,$$

where

$$\delta d_i = \frac{\partial d_i}{\partial F_i} \delta F_i + \frac{\partial d_i}{\partial B_i} \delta B_i$$

$$= \frac{1}{2} \delta F_i + \frac{1}{2} \delta B_i = \frac{1}{2} (\delta_i + \delta_i) = \delta_i \ .$$

Hence,

$$\delta_D = \sum_{i=1}^{50} 1 \cdot \delta_i \leq 50\ \text{mm} = \underline{5\ \text{cm.}}$$

Thus, we must require that:

$$(3\sigma_D)^2 \leq T_D^2 - \delta_D^2 = (10)^2-(5)^2 = 75 \text{ cm}^2$$

or

$$\sigma_D^2 \leq \frac{75 \text{ cm}^2}{9} = \underline{8.33 \text{ cm}^2}$$

in order to meet the specifications.

Denoting the MSE in the individual segments d_i by $\sigma_{d_i} = \sigma_d$ (all assumed equal) we get

$$\sigma_D^2 = \sum_{i=1}^{50} \sigma_{d_i}^2 = 50\sigma_d^2 ,$$

from which we obtain

$$\sigma_d^2 \leq \frac{\sigma_D^2}{50} = \frac{8.33}{50} = \underline{0.16}\text{cm}^2 .$$

Remembering that each segment d_i is given by:

$$d_i = \frac{1}{2}(F_i + B_i) ,$$

and denoting the MSE in either F_i or B_i (both assumed equal) by σ we get:

$$\sigma_{d_i}^2 = \sigma_d^2 = \left(\frac{\partial d_i}{\partial F_i}\right)^2 \sigma_{F_i}^2 + \left(\frac{\partial d_i}{\partial B_i}\right)^2 \sigma_{B_i}^2$$

$$= \left(\frac{1}{2}\right)^2 \sigma^2 + \left(\frac{1}{2}\right)^2 \sigma^2$$

$$= \left(\frac{1}{4} + \frac{1}{4}\right) \sigma^2 = \frac{1}{2}\sigma^2$$

and

$$\sigma^2 \leq 2\sigma_d^2 = 2(0.1\bar{6}) = \underline{0.3\bar{3}} \text{ cm.}$$

Recalling that we want to know what differences between the forth and back measurements can we tolerate, and denoting such differences by Δ_i, we can write:

$$\Delta_i = F_i - B_i \ .$$

Then:

$$\sigma^2_{\Delta_i} = \sigma^2_{\Delta} = \left(\frac{\partial \Delta_i}{\partial F_i}\right)^2 \sigma^2_{F_i} + \left(\frac{\partial \Delta_i}{\partial B_i}\right)^2 \sigma^2_{B_i}$$

$$= (1)^2\sigma^2 + (-1)^2\sigma^2$$

$$= 2\sigma^2 \ .$$

Thus, we end up with the condition:

$$\sigma^2_{\Delta} \leq 2\sigma^2 = 2(0.3\bar{3}) = \underline{0.6\bar{6}} \text{ cm}^2$$

or

$$\sigma_{\Delta} \leq 0.816 \doteq \underline{0.8} \text{ cm.}$$

This means that if we postulate a parent Gaussian PDF for the differences Δ, the above σ_{Δ} is required to be smaller or equal than the RMS of the underlying PDF. Consequently, the specifications will be as follows: We should get 68% of the differences Δ's with $\pm \sigma_{\Delta}$, i.e. within $\pm$ 0.8 cm, and 95% of Δ within $\pm 2\sigma_{\Delta}$, i.e. within $\pm$ 1.6 cm. These specifications are looser than a man with an experience in practice would expect. It illustrates the fact that in practice the specifications are very often unnecessarily too stringent.

6.4 Problem of Adjustment

6.4.1 Formulation of the Problem

Let us resume now at the end of section 6.2 where we have defined the proper problem of adjustment as the transition

$$(\bar{L}, \Sigma_L) \rightarrow (\hat{X}, \Sigma_X) \tag{6.32}$$

for an overdetermined mathematical model

$$F(L, X) = 0 \, . \tag{6.33}$$

By "overdetermined" we mean that the known $\bar{L}$ contains too many components to generally fit the above model for whatever X we choose, i.e. yielding infinite number of solutions X . The only way to satisfy the model , i.e. the prescribed relations,is to allow some of or all the $\bar{L}$ to change slightly while solving for $\hat{X}$. In other words, we have to regard $\bar{L}$ as an approximate value of some other value $\hat{L}$ which yields a unique solution $\hat{X}$ and seek the final value $\hat{L}$ together with $\hat{X}$.

Denoting

$$\hat{L} - \bar{L} = V \tag{6.34}$$

we may reformulate our mathematical model (6.33) as:

$$F(\hat{L}, X) = F(\bar{L} + V, X) = 0 \tag{6.35}$$

where V is called the vector of discrepancies.

Note that V plays here very much the same role as the v's have played in section 4.8. From the mathematical point of view, there is not much difference between V and v. However, from the philosophical viewpoint, there is, because V represents a vector of discrepancies of

s different physical quantities (see also section 5.4) while v was a vector of discrepancies of n observations of the same physical quantity. To show the mathematical equivalence of these two we shall, in the next section, treat the computation of a sample mean as an instructive adjustment problem.

6.4.2 Mean of a Sample as an Instructive Adjustment Problem, Weights

Let us regard a random sample $L = (\ell_1, \ell_2, \ldots, \ell_n)$ of n observations representing one physical quantity ℓ as uncorrelated estimate of the mean . Further we shall denote the definition set of L by $\bar{L}$, where $\bar{L} = (\bar{\ell}_1, \bar{\ell}_2, \ldots, \bar{\ell}_m)$ consists of only m distinctly different values of ℓ's. Let us seek an estimate $\hat{x}$, satisfying the mathematical model

$$x = \ell \tag{6.36}$$

representing the identity transformation. Evidently, the model is overdetermined because the individual $\bar{\ell}_j$, $j = 1, 2, \ldots, m$, are different from each other and cannot therefore all satisfy the model.

So, we reformulate the model as:

$$x = \bar{\ell}_j + v_j \quad j = 1, 2, \ldots, m \tag{6.37}$$

where the v's are the discrepancies. We have to point out that, although we seek now the same result as we have sought in section 4.7, the formulation here will be slightly different to enable us to use analogies later on. While we have been taking all the n observations into account in section 4.7, we shall now work only with the m distinctly different values $\bar{\ell}_j$,

$j = 1, 2, \ldots, m$, that constitute the sample $\bar{L}$*).

Thus we shall have to compute the mean $\bar{\ell}$ from the second formula introduced in section 3.1.3 (equation (3.4)), i.e.:

$$\bar{\ell} = \sum_{j=1}^{m} \bar{\ell}_j \; P(\bar{\ell}_j) = \sum_{j=1}^{m} \bar{\ell}_j \, P_j \,, \tag{6.38}$$

rather than the first (equation (3.3)) as used in section 4.7. Here, according to section 3.1.3, $P_j = c_j/n$ with c_j, being the count of the same values ℓ_j' in the original sample L' containing all n observations. Hence P_j are the experimental (actual) probabilities. In other words, if we wish $\hat{x}$ to equal $\bar{\ell}$, the model (6.37) yields the following solution:

$$\hat{x} = \sum_{j=1}^{m} \bar{\ell}_j \, P_j \,, \tag{6.39}$$

or

$$\boxed{\hat{x} = P^T \bar{L}} \tag{6.40}$$

in vector notation, where $P^T = (P_1, P_2, \ldots, P_m)$.

The coefficients P_j are called <u>weight coefficients</u>, or simply <u>weights</u>, and $\hat{x}$ is called the weighted mean - analogy borrowed from mechanics (see section 3.1.3). Note that, with the weights being nothing else but the experimental probabilities, we put "more weight" on the values with which we are more "certain", i.e. which are repeated more often in the sample, which is intuitively pleasing.

---

* $\bar{L} = (\bar{\ell}_1, \bar{\ell}_2, \ldots, \bar{\ell}_m)$ can be regarded in this context as a sample of "grouped" observations, i.e. each constituent $\bar{\ell}_j$, $j = 1, 2, \ldots, m$, has a count (frequency) c_j associated with it in the original sample L'.

In our slightly different notation even the least-squares principle, as formulated in section 5.3, would sustain a minor change. While we were seeking such ℓ^o as to make

$$\frac{1}{n} \sum_{i=1}^{n} v_i^2 = \frac{1}{n} \sum_{i=1}^{n} (\ell_i - \ell^o)^2 \tag{6.41}$$

minimum, we would have to write now the condition of minimum variance as:

$$\min_{\ell^o \varepsilon R} \left[\sum_{j=1}^{m} P_j v_j^2 \right] , \tag{6.42}$$

where $v_j = \bar{\ell}_j - \ell^o$. In matrix notation, (6.42) becomes:

$$\boxed{\min_{\ell^o \varepsilon R} V^T P V} , \tag{6.43}$$

where P is a diagonal matrix, i.e.

$$P = \text{diag } (P_1, P_2, \ldots, P_m) . \tag{6.44}$$

The latter formulation, i.e. equations 6.42 and 6.43 is more general since we can regard the former formulation, i.e. euqation 6.41 as a special case of (6.42) and not vice-versa. We have

$$\frac{1}{n} \sum_{i=1}^{n} v_i^2 = \sum_{i=1}^{n} P_i v_i^2$$

which implies that $P_i = \frac{1}{n}$, for i = 1, 2, ..., n are equal weights for all the observations ℓ_i'. Hence we shall use (6.43) exclusively from now on. The same holds true even for the two formulae for $\bar{\ell}$ and we shall use equation (6.40).

Note that once we apply the condition 6.43, the discrepancies cease to be variable quantities and become residuals (see 4.8). We shall denote these residuals by $\hat{V}$.

Equation 3.7 can now be obviously written as

$$S_L^2 = \sum_{j=1}^{m} P_j \hat{v}_j^2 \tag{6.45}$$

or in matrix notation,

$$\boxed{S_L^2 = \hat{V}^T P \hat{V}.} \tag{6.46}$$

Consequently, we shall restate the least-squares principle as follows: the value $\hat{x}$ that makes the value of the quadratic form $V^T PV$ the least ensures automatically the minimum variance of the sample L. This property does not depend on any specific underlying PDF. If L has got normal parent PDF (or any symmetric distribution), $\hat{x}$ is the most probable estimate of x, which is sometimes called the maximum likelihood estimate of x.

6.4.3 Variance of the Sample Mean

We have shown that the simple problem of finding the mean of a sample can be regarded as a trivial adjustment problem. Hence we are entitled to ask the question: What will be the variance-covariance matrix of the result as derived from the variance-covariance matrix of the original sample? In other words, we may ask what value of variance can be associated with the result - the mean of the sample.

The question is easily answered using the covariance law (section 6.3.1). We have established that (equation 6.40):

$$\hat{x} = P^T L .$$

Hence, by applying the covariance law (equation 6.15)we obtain:

$$\Sigma_{\hat{x}} = B\Sigma_L B^T = P^T \Sigma_L P = S^2_{\hat{x}} \quad ,$$

i.e.

$$S^2_{\hat{x}} = P^T \Sigma_{\bar{L}} P \quad . \tag{6.47}$$

Here $\Sigma_{\bar{L}}$ is not yet defined. All we know is that $\bar{L} \equiv (\bar{\ell}_1, \bar{\ell}_2, \ldots \bar{\ell}_m)$ is a sample of "grouped" observations $\bar{\ell}_i$ with different weights (observed probabilities) P_i associated with them. Let us hence assume these observations uncorrelated and let us also assume that there can be some "variances" $S^2_{\bar{\ell}_i}$ attributed to these observations. In such a case, the variance covariance matrix of $\bar{L}$ can be expressed as

$$\Sigma_{\bar{L}} = \text{diag} \left(S^2_{\bar{\ell}_1}, S^2_{\bar{\ell}_2}, \ldots, S^2_{\bar{\ell}_m} \right) \quad . \tag{6.48}$$

Substituting (6.48) into (6.47), we get:

$$S^2_{\hat{x}} = \sum_{j=1}^{m} P_j^2 \, S^2_{\bar{\ell}_j} \quad *) \quad . \tag{6.49}$$

On the other hand the value of $\hat{x}$ (i.e. the sample mean) can be computed using the original sample of observations, $L = (\ell_1, \ell_2, \ldots, \ell_n)$, i.e. the ungrouped observations ℓ_i, $i = 1, 2, \ldots, n$, which all have equal experimental probabilities (equal weights) of $1/n$, yielding:

$$\hat{x} = \frac{1}{n} \sum_{i=1}^{n} \ell_i = \frac{1}{n} \left[\ell_1 + \ell_2 + \ldots + \ell_n \right] \quad . \tag{6.50}$$

Hence, we can compute the variance of the mean, i.e. $S^2_{\hat{x}}$, again by applying the law of propagation of errors on (6.50), and we get:

$$S^2_{\hat{x}} = \sum_{i=1}^{n} \left(\frac{\partial x}{\partial \ell_i} \right)^2 S^2_{\ell_1} = \left(\frac{1}{n} \right)^2 \sum_{i=1}^{n} S^2_{\ell_i} \quad , \tag{6.51}$$

---

*It should be noted here that since $\bar{L} \equiv (\bar{\ell}_1, \bar{\ell}_2, \ldots, \bar{\ell}_m)$ is a sample of group observations, for which a different weight P_j (experimental probability) is associated with each element $\bar{\ell}_j$, $j = 1, 2, \ldots, m$, the individual variances $S^2_{\bar{\ell}_j}$ assigned to the $\bar{\ell}_j$ are, in general, different from each other, i.e. they vary with the groups of observations.

in which all the 'variances' $S^2_{\ell_i}$ are again assumed to have the same value and equal to the sample variance S^2_L given by

$$S^2_L = \frac{1}{n} \sum_{i=1}^{n} (\ell_i - \hat{x})^2. \tag{6.52}$$

Equation (6.51) then gives:

$$\boxed{S^2_{\hat{x}} = \frac{1}{n^2} (n\, S^2_L) = S^2_L/n} \tag{6.53}$$

which indicates that the variance of the sample mean equals to the variance of the sample computed from equation (6.52) divided by the total number of elements of the sample*).

We thus ended up with two different formulae, (6.49) and (6.53), for the same value $S^2_{\hat{x}}$. In the first approach, we have regarded the individual observations (really groups of observations having the same value) as having different variances $S^2_{\bar{\ell}_j}$ associated with them. The second approach assumes that all the observations belong to the same sample with variance S^2_L. Numerically, we should get the same value of $S^2_{\hat{x}}$ from both formulae, hence

$$\sum_{j=1}^{m} (P_j^2 \; S^2_{\bar{\ell}_i}) = \frac{S^2_L}{n} \; . \tag{6.54}$$

Let us write the left hand side of (6.54) in the form:

$$\sum_{j=1}^{m} (P_j^2 \; S^2_{\bar{\ell}_j}) = \sum_{j=1}^{m} [P_j (P_j \; S^2_{\bar{\ell}_i})],$$

---

* In terms of our previous notation, we can write the variance of the sample mean as

$$S^2_{\bar{\ell}} = S^2_L / n \; .$$

and the right hand side in the form:

$$\frac{S_L^2}{n} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{S_L^2}{n}\right) .$$

Using the same manipulation as in section 6.4.2 when dealing with the V's and ℓ's, and also earlier, in section 3.1.4 when prooving equation (3.4), the right hand side can be rewritten as:

$$\frac{1}{n} \sum_{i=1}^{n} \left(\frac{S_L^2}{n}\right) = \sum_{j=1}^{m} \left[P_j \left(\frac{S_L^2}{n}\right)\right] ,$$

in which P_j has the same meaning as in (6.49).
Now, the condition (6.54) becomes:

$$\sum_{j=1}^{m} \left[P_j \left(P_j \, S^2_{\bar{\ell}_j}\right)\right] = \sum_{j=1}^{m} \left[P_j \left(\frac{S_L^2}{n}\right)\right], \tag{6.55}$$

which can be always satisfied if

$$P_j \, S^2_{\bar{\ell}_j} = \frac{S_L^2}{n} = K, \quad j = 1, 2, \ldots, m, \tag{6.56}$$

where K is a constant value for a specific sample that equals to the variance of the sample mean. From (6.56) we get:

$$S^2_{\bar{\ell}_j} = \frac{K}{P_j} , \qquad j = 1, 2, \ldots, m, \tag{6.57}$$

which shows that in order to get the correct result from (6.49) we have to assume that in the first approach the individual observations have variances inversely proportional to their weights.

This result is usually expressed in the form of the following principle: the weight of an observation is inversely proportional to its variance, i.e.

$$\boxed{P = \frac{K}{S^2} .} \tag{6.58}$$

We can also write using equation (6.57):

$$P_1 S_1^2 = P_2 S_2^2 = \ldots = 1 \cdot S_0^2 = K \qquad (6.59)$$

where S_0^2, constant for a specific sample, is known as the variance of unit weight. It can be interpreted as the variance of an imaginary observation whose weight equals to one. In the case of sample mean $\bar{\ell}$, S_o equals to $S_{\bar{\ell}}$.

From equations (6.46) and (6.53) we can write:

$$S_{\hat{x}}^2 = \frac{\hat{V}^T P \hat{V}}{n} \quad . \qquad (6.60)$$

This result will be often referred to in the subsequent development.

We have to point out that the whole argument in this section hinges on the acceptance of the "variances" $S_{\ell_i}^2$ and $S_{\bar{\ell}_i}^2$. They have been introduced solely for the purpose of deriving formulae (6.53),(6.58) that are consistent with the rest of the adjustment calculus. The more rigorous alternative is to accept the two formulae by definition.

6.4.4 Variance Covariance Matrix of the Mean of a Multisample

We have seen in section 6.4.3 that the mean $\bar{\ell}$ of a sample L has also a standard deviation $S_{\bar{\ell}}$ associated with it. This standard deviation is $\sqrt{(n)}$-times smaller than the standard deviation S_L of the sample itself and can be interpreted as a measure of confidence we have in the correctness of the mean $\bar{\ell}$. Evidently, our confidence increases with the number of observations.

We can now ask ourselves the following question: Does the mean $\bar{L}$ of a multisample L also have a variance-covariance matrix associated with it? The answer is - there is nothing to prevent us from defining it by generalising the discovery from the last section. We get

$$\Sigma_{\bar{L}} = \begin{bmatrix} S^2_{\bar{\ell}_1} & S_{\bar{\ell}_1\bar{\ell}_2} & \ldots & S_{\bar{\ell}_1\bar{\ell}_s} \\ S_{\bar{\ell}_2\bar{\ell}_1} & S^2_{\bar{\ell}_2} & \ldots & S_{\bar{\ell}_1\bar{\ell}_s} \\ S_{\bar{\ell}_s\bar{\ell}_1} & & \ldots & S^2_{\bar{\ell}_s} \end{bmatrix} \qquad (6.61)$$

where

$$s^2_{\bar{\ell}_i} = \frac{1}{n_i} s^2_{\ell_i}$$

and

$$s_{\bar{\ell}_i \bar{\ell}_j} = \frac{1}{n_i} s_{\ell_i \ell_j} .$$

Here we have to require again that $n_i = n_j$, i.e. that both components of the multisample have the same number of elements (see section 3.3.5). Obviously, if this requirement is satisfied for all the pairs of components we have

$$n_1 = n_2 = \ldots = n_s = n$$

and

$$\Sigma_{\bar{L}} = \frac{1}{n} \Sigma_L . \tag{6.62}$$

By analogy, the variance-covariance matrix obtained via the covariance law (see section 6.3.1) from the variance-covariance matrix of the mean of multisample is associated with the mean of the derived multisample, or statistical estimate $\hat{X}$. We say that

$$\boxed{\Sigma_{\hat{X}} = B \Sigma_{\bar{L}} B^T} \tag{6.53}$$

is the variance-covariance matrix of the statistical estimate $\hat{X}$, i.e. of the solution of uniquely determined mathematical model

$$\hat{X} = F(\bar{L}) .$$

Similar statements can be used for other laws of propagation of errors. Development of these is left to the student, who should also compare results of this sections with the solution of Example 6.14.

Example 6.15: Let us take again the experiment described in Examples 6.1, 6.3 and 6.4. This time we shall be interested in deriving the variance-covariance matrix $\Sigma_{\hat{X}}$ of the solution vector $\hat{X}$.

Solution: First we evaluate $\Sigma_{\bar{L}}$ from eq. (6.61). We obtain

$$S^2_{\bar{a}} = \frac{1}{5} S^2_a = \frac{0.004 \text{ cm}^2}{5} = 0.0008 \text{ cm}^2 ,$$

$$S^2_{\bar{b}} = \frac{1}{5} S^2_b = \frac{0.0056 \text{ cm}^2}{5} \doteq 0.0011 \text{ cm}^2 .$$

Since $S_{ab} = 0$ we get

$$\Sigma_{\bar{L}} = \begin{bmatrix} 0.0008 & 0 \\ 0 & 0.0011 \end{bmatrix} \text{ cm}^2 = \frac{1}{5} \Sigma_L .$$

Now, $\Sigma_{\hat{X}}$ can be evaluated from equation (6.63) and we have

$$\Sigma_{\hat{X}} = B(\tfrac{1}{5} \Sigma_L) B^T = \frac{1}{5} B \Sigma_L B^T = \frac{1}{5} \Sigma_X ,$$

or

$$\Sigma_{\hat{X}} \doteq \begin{bmatrix} 0.00081 \text{ cm}^2 & 0.1079 \text{ cm}^3 \\ 0.01079 \text{ cm}^3 & 21.51254 \text{ cm}^4 \end{bmatrix} .$$

Thus the standard deviations of the estimates $\hat{d}$ and $\hat{\alpha}$ are given by

$$S_{\hat{d}} = \sqrt{(0.00081 \text{ cm}^2)} \doteq \underline{0.028 \text{ cm}} ,$$

$$S_{\hat{\alpha}} = \sqrt{(21.51254 \text{ cm}^4)} \doteq \underline{4.64 \text{ cm}^2} .$$

6.4.5 The Method of Least-Squares, Weight Matrix

The least-squares principle as applied on the trivial identity transformation, i.e. the sample mean, can be generalized for other mathematical models. Taking the general formulation of the problem of adjustment as

described in section 6.4.1, i.e.

$$\boxed{F(\bar{L} + V, X) = 0,}$$

we can again ask for such X that would make the value of the quadratic form of the weighted discrepencies, $V^T PV$, minimum, i.e.

$$\min_{\hat{X}\varepsilon R^n} \quad V^T PV \ . \tag{6.64}$$

The condition (6.64) for the majority of mathematical models, is enough to specify such $X = \hat{X}$ uniquely. The approach to adjustment using this condition became known as the method of least-squares.

The question remains here as how to choose the matrix P. In the case of the sample mean we have used

$$P = \text{diag}\ (K/S^2_{\bar{\ell}_1}, K/S^2_{\bar{\ell}_2}, \ \ldots\ , K/S^2_{\bar{\ell}_m}),$$

that is

$$P = K\ \text{diag}\ (1/S^2_{\bar{\ell}_1}, 1/S^2_{\bar{\ell}_2}, \ \ldots\ , 1/S^2_{\bar{\ell}_m}).$$

Using the notation developed for the multisample, this can be rewritten as:

$$\boxed{P = K\ \Sigma^{-1}_{\bar{L}}} \tag{6.65}$$

which indicates that the matrix P is obtained by multiplying the constant K by the inverse of the variance-covariance matrix of the means of observations. This is in our case a diagonal matrix as we have postulated the sample L to be uncorrelated.

We again notice that, mathematically, there is not much difference between a sample and multisample - they can be hence treated in much the same way. Thus, there is not basic difference between the apparently

trivial adjustment of the sample mean and the general problem of adjustment. The only difference is that in the first case $\hat{X}$ is a vector of one component, while generally it may have many components.

This gives a rise to the question of what would be the role of K (K having been a scalar equal to S_x^2 in the adjustment of the mean of sample), in the least squares method, where $\hat{X}$ has several constituents. Let us just say at this time that we usually compute the weight matrix P, as it is called in the method of least-squares, as

$$P = \kappa \Sigma_{\underline{L}}^{-1} \tag{6.66}$$

where κ is an arbitrarily chosen constant, the meaning of which will be shown later. This can be done because, as will also be shown later, the solution $\hat{X}$ is independent of κ since it does not change the ratio between the weights or variances of the individual observations.

In this course we shall be dealing with only two particular mathematical models which are the most frequently encountered in practice. In these models, we shall use the following notation:

n for the number of constituents of the primary or original multisample L;

u for the number of constituents of the derived, or unknown(to be derived) multisample X;

r for the number of independent equations (relationships) that can be formulated between the constituents of L and X.

Moreover, we shall consider these models to be linear.

The first model is

$$A X = L , \tag{6.67}$$

in which A is an n b u matrix, X is a u b 1 vector and L is an n by 1 vector (n = r > u). The adjustment of this model is usually called parametric adjustment, adjustment of observation equations, or adjustment of indirect observations, etc.

The second model is

$$BL = C \tag{6.68}$$

in which B is an r by n matrix, L is n x 1 and C is r x 1 vectors (r < n). The adjustment of this model is known as conditional adjustment, adjustment of condition equations, etc.

The two mathematical models are evidently quite special since they are both linear. Fortunately many problems in practice, although non-linear by nature, can be linearized. This is the reason why the two treated models are important.

6.4.6 Parametric Adjustment

In this section, we are going to deal with the adjustment of the linear model (6.67), i.e.

$$AX + C = L \quad (n > u) \tag{6.69}$$

which, for the adjustment, will be reformulated as:

$$AX - (\bar{L} + V) = 0$$

or

$$V = AX - \bar{L}^{*}) \quad . \tag{6.70}$$

Here A is called the design matrix, X is the vector of unknown parameters, $\bar{L}$ is the vector of observations, ($\bar{L} = \bar{L}^{*} - C$ where $\bar{L}^{*}$ is the mean of the observed multisample), and V is the vector of discrepancies, which is also unknown. The formulation (6.70) is known as a set of observation equations.

We wish to get such $X = \hat{X}$ that would minimize the quadratic form $V^{T}PV$ in which P is the assumed weight matrix for the observations $\bar{L}$ (see the previous section). This quadratic form, which is sometimes called the quadratic form of weighted discrepancies, can be rewritten using the observation equations (6.70) as

$$\begin{aligned} V^{T}PV &= (AX - \bar{L})^{T} P(AX - \bar{L}) \\ &= ((AX)^{T} - \bar{L}^{T}) (PAX - P\bar{L} \\ &= X^{T}A^{T}PAX - \bar{L}^{T}PAX - X^{T}A^{T}L + \bar{L}^{T}P\bar{L} \end{aligned} \tag{6.71}$$

From equation (6.66) we have $P = \kappa \Sigma_{\bar{L}}^{-1}$, where κ is a constant scalar and $\Sigma_{\bar{L}}$ is the variance-covariance matrix of $\bar{L}$. Since $\Sigma_{\bar{L}}$ is symmetric, the weight matrix P is symmetric as well and $P^{T} = P$. We can thus write

$$\bar{L}^{T} = PAX = X^{T}A^{T}P\bar{L} \tag{6.72}$$

since it is a scalar quantity.

Substituting (6.72) into (6.71) we get

---

* If we have a non-linear model

$$L = F(X)$$

it can be easily linearized by Taylor's series expansion, i.e.

$$L = F(X^{o}) + \left.\frac{\partial F}{\partial X}\right|_{X=X^{o}} (X-X^{o}) + \ldots ,$$

in which we neglect the higher order terms. Putting ΔX for $X-X^{o}$, ΔL for $L-F(X^{o})$ and A (a matrix) for $\partial F/\partial X|_{X=X^{o}}$ we get

$$\Delta L \doteq A\Delta X \quad .$$

This is essentially the same form as equation (6.69). However, in this case we are solving for the corrections ΔX to the approximate value X^{o} of the vector X, instead of solving for X itself.

$$V^T PV = X^T A^T PAX - 2X^T A^T P\bar{L} + \bar{L}^T P\bar{L} \ . \qquad (6.73)$$

The quadratic function (6.73), called sometimes the variations function, is to be minimized with respect to X. This is accomplished by equating all the partial derivatives to zero, i.e.

$$\frac{\partial}{\partial X^i} V^T PV = 0 \quad i = 1, 2, \ldots, u, \qquad (6.74)$$

and we obtain, writing $\partial/\partial X$ for the whole vector of partial derivatives $\partial/\partial X^i$,

$$\frac{\partial}{\partial X} V^T PV = 2\hat{X}^T A^T PA - 2\bar{L}^T PA = 0 \ , \quad *)$$

which can be rewritten as:

$$\hat{X}^T A^T PA = \bar{L}^T PA$$

or by taking the transpose of both sides we get:

$$\boxed{(A^T PA)\hat{X} = A^T P\bar{L}} \ +) \qquad (6.75)$$

This system of linear equations is called the system of normal equations which can be written, as often used in the literature, in the following abbreviated form:

$$N\hat{X} = U \qquad (6.76)$$

where $N = (A^T PA)$ is known as the matrix of coefficients of the normal equations, or simply the normal equation matrix and $U = A^T P\bar{L}$ is the vector of absolute terms of the normal equation.

The system of normal equations (6.76) has a solution $\hat{X}$

---

* From matrix algebra we know that if A is a symmetric matrix and X is a vector we get:

$$\frac{\partial}{\partial X} AX = A \text{ and } \frac{\partial}{\partial X} (X^T AX) = 2X^T A \ .$$

\+ Note that the normal equations can be obtained directly from the mathematical model by pre-multiplying it by $A^T P$.

given by

$$\hat{X} = N^{-1}U = (A^T PA)^{-1} (A^T P\bar{L}) \tag{6.77}$$

if the normal equation matrix, $N = A^T PA$, has an inverse. Note that N is a symmetric positive definite matrix.*)

To discuss the influence of the weight matrix P on the solution vector $\hat{X}$, let us use a different weight matrix, say P', such that

$$P' = \gamma P \tag{6.78}$$

where γ is an arbitrary constant. Substituting (6.78) into (6.77) we get:

$$\begin{aligned} \hat{X}' &= (A^T P'A)^{-1} (A^T P'\bar{L}) \\ &= (A^T \gamma PA)^{-1} (A^T \gamma P\bar{L}) \\ &= \frac{1}{\gamma} (A^T PA)^{-1} \gamma (A^T P\bar{L}) \\ &= \hat{X} \ . \end{aligned} \tag{6.79}$$

This result indicates that the factor κ in equation (6.66) for computing the weight matrix P from $\Sigma_{\bar{L}}$, can be chosen arbitrarily without any influence on $\hat{X}$, which really verifies the statement we have made earlier, in section 6.4.4.

It should be noted that the vector of discrepancies V as defined in (6.70), becomes after minimization of the vector of residuals (see 4.8) of the observed quantities. As such, it should be again denoted by a different symbol, say R, to show that it is no longer a vector of variables (function of X) but a vector of fixed quantities. Some authors use $\hat{V}$ for this purpose and this is the convention we are going to use (see also 6.4.2). The values $\hat{v}_i$ are computed directly from equation (6.70) in the same units as these of the vector $\bar{L}$. Then the adjusted observations will be given by $\hat{L} = \bar{L} + \hat{V}$.

---

* A^Tmatrix say N, is positive definite if the value of the quadratic form $Y^T NY$ is positive for any vector Y (of the appropriate dimension).

We should keep in mind that one of the main features of the parametric method of adjustment is that the estimate of the vector of unknown parameters, i.e. $\hat{X}$, is a direct result of this adjustment as given by equation (6.77).

At this stage, it is worthwhile going back to the trivial problem of adjustment - the sample mean. According to the equation (6.79), we can choose the weights of the individual observations to be inversely proportional to their respective variances with an arbitrary constant κ of proportionality. This indicates that the weights do not have to equal to the experimental probabilities for which $\sum_{i=1}^{n} P_i = 1$, as we required in sections 6.4.2 and 6.4.3. In this case, the observation equations will be

$$\hat{x} = \bar{\ell}_1 + \hat{v}_1, \text{ with weight } p_1,$$

$$\hat{x} = \bar{\ell}_2 + \hat{v}_2, \text{ with weight } p_2,$$

$$\cdots\cdots\cdots$$

$$\hat{x} = \bar{\ell}_n + \hat{v}_n, \text{ with weight } P_n.$$

Or, in matrix form

$$A\hat{X} = \bar{L} + \hat{V} ,$$

where

$$A = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} , \qquad \bar{L} = \begin{bmatrix} \bar{\ell}_1 \\ \bar{\ell}_2 \\ \cdot \\ \cdot \\ \cdot \\ \bar{\ell}_n \end{bmatrix} ,$$

with weight matrix, $P = \text{diag}\,(p_1, p_2, \ldots, p_n)$.

Substituting in equation (6.77) we get the solution, i.e. the weighted mean of the sample, as

$$\hat{x} = \frac{\sum_{i=1}^{n} p_i \bar{\ell}_i}{\sum_{i=1}^{n} p_i} \qquad (6.80)$$

which agrees with the result in section 6.4.2., when $\sum_{i=1}^{n} p_i$ equals to one.

Formula (6.80) is the general formula used to compute the weighted mean of a sample of weighted observations.

Example 6.16: Let us have a levelling line connecting two junction points, G and J, the elevations of which, H_G, H_J, are known. The levelling line is divided into three sections d_1, d_2 and d_3 long. Each level difference h_1, h_2 and h_3 was observed, with results $\bar{h}_1$, $\bar{h}_2$ and $\bar{h}_3$. The observations $\bar{h}_i$ are considered uncorrelated with variances proportional to the corresponding lengths d_i, $i = 1, 2, 3$.

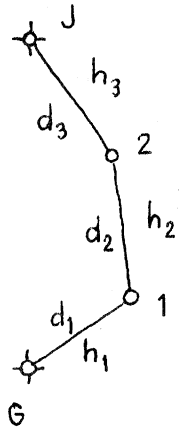


Figure 6.6

It is required to determine the adjusted values of the elevations of points 1 and 2, i.e. H_1 and H_2 respectively, using the parametric adjustment.

Solution

From the given data we have: number of observations n=3 ; number of unknowns u = 2. Therefore, we have one redundant observation. The independent relationships between the observations and the unknowns are written as follows (each relation corresponds to one observation):

$$h_1 = H_1 - H_G \ ,$$

$$h_2 = H_2 - H_1 \ ,$$

$$h_3 = H_J - H_2 \ .$$

The above relations can be rewritten in the general form used in the previous development:

$$\underset{3,2}{A}\ \underset{2,1}{X} = \underset{3,1}{L}$$

where $X = (H_1, H_2)$ and

$$H_1 = h_1 + H_G = L_1 \ ,$$

$$-H_1 + H_2 = h_2 = L_2 \ ,$$

$$-H_2 = h_3 - H_J = L_3 \ .$$

Putting this in matrix form, we get

$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} = \begin{bmatrix} (h_1 + H_G) \\ h_2 \\ (h_3 - H_J) \end{bmatrix} \ .$$

The corresponding set of observation equations are:

$$H_1 = H_G + (\bar{h}_1 + v_1) \ ,$$

$$-H_1 + \hat{H}_2 = (\bar{h}_2 + v_2) \ ,$$

$$H_2 = -H_J + (\bar{h}_3 + v_3) \ .$$

These observation equations can be written in matrix form as:

$$\underset{3,1}{V} = \underset{3,2}{A}\ \underset{2,1}{X} - \underset{3,1}{\bar{L}} \ ,$$

where:

$$\underset{3,1}{V} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \quad \underset{2,1}{X} = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}, \quad \bar{L} = \begin{bmatrix} (\bar{h}_1 + H_G) \\ \bar{h}_2 \\ (\bar{h}_3 - H_J) \end{bmatrix}$$

and the design matrix A is given by

$$\underset{3,2}{A} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

We assumed that the observed values $\bar{h}_1$, $\bar{h}_2$ and $\bar{h}_3$ are uncorrelated. We will also assume that H_G and H_J are errorless, Hence:

$$\Sigma_{\bar{L}} = \text{diag} \left(S_{\bar{h}_i}^2 , S_{\bar{h}_2}^2 , S_{\bar{h}_3}^2 \right) .$$

But $S_{\bar{h}_i}^2$ is proportional to d_i, i = 1, 2, 3;

thus

$$\Sigma_{\bar{L}} = \kappa \text{ diag } (d_1, d_2, d_3).$$

Further, we choose $\kappa = 1$ and we get

$$P = \kappa \, \Sigma_{\bar{L}}^{-1} = \text{diag} \left(\frac{1}{d_1}, \frac{1}{d_2}, \frac{1}{d_3} \right) .$$

Applying the method of least-squares the normal equations are

$$\underset{2,2}{N} \quad \underset{2,1}{\hat{X}} = \underset{2,1}{U},$$

where

$$N = A^T P A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{d_1} & 0 & 0 \\ 0 & \frac{1}{d_2} & 0 \\ 0 & 0 & \frac{1}{d_3} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

This gives

$$N_{2,2} = \begin{bmatrix} (\frac{1}{d_1} + \frac{1}{d_2}), & -\frac{1}{d_2} \\ -\frac{1}{d_2}, & (\frac{1}{d_2} + \frac{1}{d_3}) \end{bmatrix}$$

and

$$U = A^T P \bar{L} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{d_1} & 0 & 0 \\ 0 & \frac{1}{d_2} & 0 \\ 0 & 0 & \frac{1}{d_3} \end{bmatrix} \begin{bmatrix} (\bar{h}_1 + H_G) \\ \bar{h}_2 \\ (\bar{h}_3 - H_J) \end{bmatrix}$$

Hence

$$U = \begin{bmatrix} (\frac{\bar{h}_1 + H_G}{d_1} - \frac{\bar{h}_2}{d_2}) \\ (\frac{\bar{h}_2}{d_2} - \frac{\bar{h}_3 - H_J}{d_3}) \end{bmatrix} .$$

The solution $\hat{X}$ is given by

$$\underset{2,1}{\hat{X}} = \underset{2,2}{N^{-1}} \quad \underset{2,1}{U} ,$$

where

$$N^{-1} = \frac{d_1 d_2 d_3}{(d_1 + d_2 + d_3)} \begin{bmatrix} (\frac{1}{d_2} + \frac{1}{d_3}) , & \frac{1}{d_2} \\ \frac{1}{d_2} , & (\frac{1}{d_1} + \frac{1}{d_2}) \end{bmatrix} .$$

Performing the multiplication $N^{-1}U$ and realizing that $\hat{X} = (\hat{H}_1, \hat{H}_2)$, we obtain:

$$\hat{H}_1 = H_G + \bar{h}_1 + \frac{d_1}{\Sigma_i d_i} (H_J - H_G - \Sigma_i \bar{h}_i),$$

$$\hat{H}_2 = H_J - \bar{h}_3 - \frac{d_3}{\Sigma_i d_i} (H_J - H_G - \Sigma_i \bar{h}_i) .$$

Now, we compute the residuals $\hat{v}_i$ from the equation $\hat{V} = A\hat{X} - \bar{L}$ and find

$$\hat{v}_i = \frac{H_J - H_G - \Sigma_i \bar{h}_i}{\Sigma_i d_i} d_i , \quad i = 1, 2, 3.$$

Finally, we compute the adjusted observations from

$$\hat{L} = \bar{L} + \hat{V} .$$

Remembering that H_G and H_J are assumed errorless we get:

$$\hat{h}_i = \bar{h}_i + \hat{v}_i, \qquad i = 1, 2, 3.$$

Example 6.17: A local levelling network composed of 6 sections, shown in Figure 6.7, was observed. Note that the arrow heads indicate the direction of increasing elevation . The following table summarizes the observed differences in heights $\bar{h}_i$ along with the corresponding length of each section.

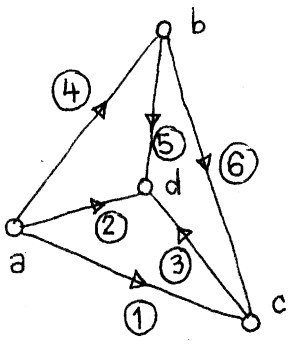


Figure 6.7

Section No.	Stations		$-\bar{h}_i$ (m)	length ℓ_i (km)
	from	to		
1	a	c	6.16	4
2	a	d	12.57	2
3	c	d	6.41	2
4	a	b	1.09	4
5	b	d	11.58	2
6	b	c	5.07	4

Assume that the variances $S^2_{\bar{h}_i}$, $i = 1, 2, \ldots, 6$, are proportional to the corresponding lengths ℓ_i. The elevation H_a of station a is considered to be 0 metres. It is required to adjust this levelling net by the parametric method of adjustment and deduce the least-squares estimates $\hat{H}_b$, $\hat{H}_c$, and $\hat{H}_d$ for the elevations H_b, H_c and H_d of the points b, c, and d.

Solution:

From the given data we have - number of independent observations: n = 6, number of unknowns: u = 3. Hence we have 3 redundant observations, i.e. 3 degrees of freedom .

Our mathematical model in this case is linear, i.e.

$$\underset{6,3}{A}\ \underset{3,1}{X} = \underset{6,1}{L} ,$$

where

$$\underset{3,1}{X} = (H_b, H_c, H_d) .$$

The 6 independent observation equations will be (one equation for each observed quantity):

$$\bar{h}_1 + v_1 = H_c - H_a = H_c - 0.0 = H_c,$$

$$\bar{h}_2 + v_2 = H_d - H_a = H_d - 0.0 = H_d,$$

$$\bar{h}_3 + v_3 = H_d - H_c ,$$

$$\bar{h}_4 + v_4 = H_b - H_a = H_b - 0.0 = H_b ,$$

$$\bar{h}_5 + v_5 = H_d - H_b,$$

$$\bar{h}_6 + v_6 = H_c - H_b .$$

The above set of equations can be rewritten in the following form, after substituting the values of $\bar{h}_i$:

$$\begin{aligned}
v_1 &= \hat{H}_c - 6.16 , \\
v_2 &= \hat{H}_d - 12.57 , \\
v_3 &= -\hat{H}_c + \hat{H}_d - 6.41 , \\
v_4 &= \hat{H}_b - 1.09 , \\
v_5 &= -\hat{H}_b + \hat{H}_d - 11.58 , \\
v_6 &= -\hat{H}_b + \hat{H}_c - 5.07 .
\end{aligned}$$

In matrix form we can write

$$\underset{6,1}{V} = \underset{6,3}{A}\ \underset{3,1}{X} - \underset{6,1}{\bar{L}} \quad ,$$

where

$$\underset{6,1}{V} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix}, \qquad \underset{3,1}{X} = \begin{bmatrix} H_b \\ H_c \\ H_d \end{bmatrix}, \quad \underset{6,1}{\bar{L}} = \begin{bmatrix} 6.16 \\ 12.57 \\ 6.41 \\ 1.09 \\ 11.58 \\ 5.07 \end{bmatrix}$$

and the design matrix, A, is

$$\underset{6,3}{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 0 & 0 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} .$$

Since we have no information about the correlation between $\bar{h}_i$, we will treat them as uncorrelated. Hence, the variance-covariance matrix $\Sigma_{\bar{L}}$ of the observed quantities will be:

$$\underset{6,6}{\Sigma_{\bar{L}}} = \text{diag } (4,\ 2,\ 2,\ 4,\ 2,\ 4)$$

understanding that the constant factor κ is assumed one.

The corresponding weight matrix is given as:

$$\underset{6,6}{P} = \text{diag } (0.25,\ 0.5,\ 0.5,\ 0.25,\ 0.5,\ 0.25).$$

The normal equations are

$$\underset{3,3}{N}\ \underset{3,1}{\hat{X}} = \underset{3,1}{U} \quad ,$$

yielding the solution

$$\underset{3,1}{\hat{X}} = \underset{3,3}{N^{-1}}\ \underset{3,1}{U} \quad ,$$

where

$$\underset{3,3}{N} = \underset{3,6}{A^T}\ \underset{6,6}{P}\ \underset{6,3}{A} \ .$$

Thus:

$$N = \begin{bmatrix} 0 & 0 & 0 & 1 & -1 & -1 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.25 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.25 \end{bmatrix} \times$$

$$\times \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 0 & 0 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix}$$

and

$$N = \begin{bmatrix} 0 & 0 & 0 & 0.25 & -0.5 & -0.25 \\ 0.25 & 0 & -0.5 & 0 & 0 & 0.25 \\ 0 & 0.5 & 0.5 & 0 & 0.5 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 0 & 0 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} .$$

Finally:

$$\underset{3,3}{N} = \begin{bmatrix} 1.00 & -0.25 & -0.5 \\ -0.25 & 1.00 & -0.5 \\ -0.5 & -0.50 & 1.50 \end{bmatrix} .$$

Note that N is a symmetric, positive-definite matrix.

Hence:

$$N^{-1}_{3,3} = \begin{bmatrix} 1.6 & 0.8 & 0.8 \\ 0.8 & 1.6 & 0.8 \\ 0.8 & 0.8 & 1.2 \end{bmatrix} .$$

Computing $U = A^T P \bar{L}$, we get.

$$U_{3,1} = \begin{bmatrix} 0 & 0 & 0 & 0.25 & -0.5 & -0.25 \\ 0.25 & 0 & -0.5 & 0 & 0 & 0.25 \\ 0 & 0.5 & 0.5 & 0 & 0.5 & 0 \end{bmatrix} \begin{bmatrix} 6.16 \\ 12.57 \\ 6.41 \\ 1.09 \\ 11.58 \\ 5.07 \end{bmatrix}$$

and

$$U_{3,1} = \begin{bmatrix} -6.7850 \\ -0.3975 \\ 15.2800 \end{bmatrix} .$$

Performing the multiplication $N^{-1} U$, we get $\hat{X}$ as:

$$\hat{X}_{3,1} = \begin{bmatrix} 1.6 & 0.8 & 0.8 \\ 0.8 & 1.6 & 0.8 \\ 0.8 & 0.8 & 1.2 \end{bmatrix} \begin{bmatrix} -6.7850 \\ -0.3975 \\ 15,2800 \end{bmatrix} = \begin{bmatrix} 1.05 \\ 6.16 \\ 12.59 \end{bmatrix} .$$

Therefore, we have obtained the following estimates

$\hat{H}_b = 1.05$ m,

$\hat{H}_c = 6.16$ m,

$\hat{H}_d = 12.59$ m.

By substituting the values of $\hat{X}$ we get the residual vector $\hat{V}$ for the observed $\bar{h}_i$ from the equation

$$\hat{V} = A \hat{X} - \bar{L} .$$

Namely:

$$\underset{6,1}{V} = \begin{bmatrix} 0.00\ m \\ 0.02\ m \\ 0.02\ m \\ -0.04\ m \\ -0.04\ m \\ 0.04\ m \end{bmatrix} .$$

The adjusted observations $\hat{h}$ are computed from:

$$\hat{h}_i = \bar{h}_i + \hat{v}_i \quad , \quad i = 1, 2, \ldots, 6$$

and we get:

$$\hat{h} = \begin{bmatrix} 6.16 \\ 12.57 \\ 6.41 \\ 1.09 \\ 11.58 \\ 5.07 \end{bmatrix} + \begin{bmatrix} 0.00 \\ 0.02 \\ 0.02 \\ -0.04 \\ -0.04 \\ 0.04 \end{bmatrix} = \begin{bmatrix} 6.16 \\ 12.59 \\ 6.43 \\ 1.05 \\ 11.54 \\ 5.11 \end{bmatrix}$$

in metres.

The computations can be checked by deriving the heights of points b, c and d from H_a using the adjusted $\hat{h}_i$. The resulting values must not differ from the adjusted values $\hat{H}_b$, $\hat{H}_c$ and $\hat{H}_d$.

6.4.7 Variance-Covariance Matrix of the Parametric Adjustment Solution Vector, Variance Factor and Weight Coefficient Matrix

The parametric adjustment solution vector $\hat{X}$ is given by equation (6.77), i.e.

$$\hat{X} = (A^T P A)^{-1} A^T P \bar{L} = (N^{-1} A^T P) \bar{L} .$$

This can be written as:

$$\hat{X} = B\,\bar{L} \tag{6.81}$$

where

$$B = N^{-1}\,A^T\,P\ . \tag{6.82}$$

The variance-covariance matrix $\Sigma_{\hat{X}}$ of the solution vector $\hat{X}$ can be easily deduced by applying the covariance law (equation (6.15)) on (6.81); we get:

$$\Sigma_{\hat{X}} = B\Sigma_{\bar{L}}\,B^T. \tag{6.83}$$

From equation (6.66), we have

$$P = \kappa\Sigma_{\bar{L}}^{-1}$$

and inverting both sides we obtain

$$\Sigma_{\bar{L}} = \kappa\,P^{-1}\ . \tag{6.84}$$

Substituting (6.82) and (6.84) into (6.83) we get

$$\Sigma_{\hat{X}} = (N^{-1}A^TP)\,\kappa\,P^{-1}\,(N^{-1}A^TP)^T. \tag{6.85}$$

Both P and N are symmetric matrices, so that we can write:

$$P^T = P,\quad N^T = N \text{ and } (N^{-1})^T = N^{-1}\ .$$

Substituting this into (6.85) we get

$$\begin{aligned}\Sigma_{\hat{X}} &= \kappa\ N^{-1}\,A^T\,P\,P^{-1}\,P\,A\,N^{-1}\\ &= \kappa\ N^{-1}\,(A^TPA)\,N^{-1} = \kappa\,N^{-1}\,N\,N^{-1} = \kappa\,N^{-1}\ ,\end{aligned}$$

that is

$$\boxed{\Sigma_{\hat{X}} = \kappa\ N^{-1} = \kappa\,(A^TPA)^{-1}\ .} \tag{6.86}$$

On the other hand, by putting $P = \kappa\Sigma_{\bar{L}}^{-1}$ in (6.86) we get

$$\Sigma_{\hat{X}} = \frac{1}{\kappa}\,\kappa\,(A^T\,\Sigma_{\bar{L}}^{-1}A)^{-1} = (A^T\Sigma_{\bar{L}}^{-1}A)^{-1}\ , \tag{6.87}$$

which shows that $\Sigma_{\hat{X}}$ does not depend on the choice of the factor κ. In fact, this statement is valid only if we know the correct values of the elements

of $\Sigma_{\underline{L}}$. Unfortunately, however, $\Sigma_{\underline{L}}$ is often known only up to a scale factor, i.e. we know the relative variances and covariances of the observations only. This means that we have to work with the weight matrix $\kappa\Sigma_{\underline{L}}^{-1}$ without knowing the actual value of the factor κ. Therefore $\Sigma_{\hat{X}}$ cannot be computed from equation (6.87).

If we develop the quadratic form $\hat{V}^T P \hat{V}$ *) considering the observations L to be influenced by random errors only, we get an estimate $\hat{\kappa}$ for the assumed factor κ given by

$$\hat{V}^T P \hat{V} = (n - u)\,\hat{\kappa} \quad . \tag{6.88}$$

The multiplier in the right-hand side is nothing else but the difference between the number of independent observations and the number of unknown parameters, i.e. the number of redundant observations, which is sometimes denoted by "df" and called the number of degrees of freedom, i.e.

$$df = n - u \quad . \tag{6.89}$$

df must be greater than zero in order to be able to perform a least-squares adjustment. Hence equation (6.88) becomes

$$\boxed{\hat{\kappa} = \frac{\hat{V}^T P \hat{V}}{df} \quad .} \tag{6.90}$$

Usually, in the literature, κ is known as the a priori variance factor and $\hat{\kappa}$ is called the least-squares estimate of the variance factor, or, simply, estimated or a posteriori variance factor. The estimated variance factor can be now used instead of the a priori one, yielding an estimate of $\Sigma_{\hat{X}}$

---

* Here, the vector $\hat{V}$ is the vector of residuals from the least squares adjustment .

$$\boxed{\begin{aligned}\hat{\Sigma}_{\hat{X}\hat{X}} &= \hat{\kappa}\, N^{-1} = \hat{\kappa}(A^T PA)^{-1} \\ &= \frac{\hat{V}^T P \hat{V}}{df}\,(A^T PA)^{-1}\end{aligned}}\ , \qquad (6.91)$$

which is known as the estimated variance-covariance matrix of $\hat{X}$.

To discuss the influence of the chosen variance factor κ in the weight matrix $P = \kappa\, \Sigma_{\bar{L}\bar{L}}^{-1}$ on $\hat{\Sigma}_{\hat{X}\hat{X}}$, as defined by (6.91), we take another factor, say κ'. We obtian $P' = \kappa' \Sigma_{\bar{L}\bar{L}}^{-1} = \gamma P$. Substituting in equation (6.91) we get:

$$\hat{\Sigma}'_{\hat{X}\hat{X}} = \frac{\gamma\,(\hat{V}^T P \hat{V})}{df}\,\frac{1}{\gamma}\,(A^T PA)^{-1} = \hat{\Sigma}_{\hat{X}\hat{X}}\ .$$

The above result indicates that $\hat{\Sigma}_{\hat{X}\hat{X}}$ given by equation (6.91) is independent of the choice of the a priori variance factor κ. We recall that the same holds true for the estimated solution vector $\hat{X}$ (equation 6.79).

It often happens in the adjustment calculus, that we have to use the estimated parameters $\hat{X}$ in subsequent adjustments as "observations". Then we have to take into account their respective weights. We know that the weight matrix of an observation vector must be proportional to the inverse of its variance-covariance matrix (equation 6.66). Thus, we can see that the matrix of normal equations, N, can be immediately used as the weight matrix of the vector $\hat{X}$, since the inverse N^{-1} is proportional to the variance-covariance matrix $\Sigma_{\hat{X}\hat{X}}$. Accordingly, the matrix N^{-1} is known also as the weight coefficient matrix and the square roots of its diagonal elements are called (Hansen's) weight coefficients.

Note that $\hat{X}$ is called uncorrelated when N^{-1} is diagonal, i.e. when N is diagonal. In such a case, we can solve the normal equations separately for each component of $\hat{X}$ which satisfies our intuition. The correlation of $\hat{X}$ is only remotely related to the correlation of $\bar{L}$. $\hat{X}$

will be uncorrelated if $\bar{L}$ is uncorrelated, i.e. P is diagonal, and if the design matrix A is orthogonal. On the other hand, N may be diagonal even for some other general matrices P, A.

Let us now turn once more back to the "adjustment" of the sample mean (see 6.4.3). It is left to the student to show that the normal equations degenerate into a single equation, namely equation (6.40) On the other hand, using eq.(6.91), we obtain the estimated variance of the mean $\hat{x}$ as

$$\hat{S}_{\hat{x}}^2 = \frac{\hat{V}^T P \hat{V}}{df} = \frac{\hat{V}^T P \hat{V}}{n-1} \tag{6.92}$$

Evidently the estimated variance of $\hat{x}$ differs from the variance $S_{\hat{x}}^2$ (see eq. 6.60) in the denominator. By analogy we define a new statistical quantity, the estimated variance $\hat{S}_L^2$ of a sample L

$$\boxed{\hat{S}_L^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\ell_i - \bar{\ell})^2 = \frac{1}{n-1} \sum_{i=1}^{n} \hat{v}_i^2 ,} \tag{6.93}$$

(compare with eq. 3.6) which is used in statistics wherever the mean $\bar{\ell}$ of the sample L is also being determined. It is again left to the student to show that using the estimated variances for the grouped observations (see 6.4.2) the formula (6.92) (instead of 6.60) can be derived using the argumentation of 6.4.2 and 6.4.3.

The estimated variances of the sample L and its mean $\bar{\ell}$ can be also computed using non-normalized weights, i.e. weights p_i for which $\sum_{i=1}^{n} p_i \neq 1$ (see 6.4.6). It can be shown that the appropriate formulae are

$$\hat{S}_L^2 = \frac{1}{n-1} \sum_{i=1}^{n} p_i \hat{v}_i^2 \tag{6.94}$$

and

$$\hat{S}_{\ell}^{2} = \frac{1}{(n-1)\sum_{i=1}^{n} p_i} \sum_{i=1}^{n} p_i \hat{v}_i^{2} . \qquad (6.95)$$

To conclude this section, let us try to interpret the meaning of the variance factor κ, introduced for the first time in 6.4.5. Let us take, for simplicity, an experiment yielding a unit matrix of normal equations, i.e. $N = I$. What would be the variance-covariance matrix of the solution vector $\hat{X}$? It will be a diagonal matrix

$$\Sigma_{\hat{X}} = \text{diag}\,(\kappa, \kappa, \ldots, \kappa). \qquad (6.96)$$

This implies that all the variances $S_{\hat{x}_i}^{2}$ of the components of $\hat{X}$ equal to κ. Since the square roots of the diagonal elements of N (all equal to 1) can be considered as the weights P_i of the components $\hat{x}_i$ of $\hat{X}$ we can also write:

$$P_1 S_1^{2} = P_2 S_2^{2} = \ldots = P_n S_n^{2} = \kappa \quad . \qquad (6.97)$$

Comparison with equation (6.59) gives some insight into the role the variance factor κ plays. It can be regarded as the variance of unit weight (see 6.4.3) and is accordingly usually denoted by either S_o^{2} or σ_o^{2} (in case of postulated variances). This is again intuitively pleasing since it ties together formulae (6.66) and (6.65), where κ can be also equated to S_o^{2}. Analogically, we denote $\hat{\kappa}$ by either $\hat{S}_o^{2}$ or $\hat{\sigma}_o^{2}$.

By adopting the notation $\hat{\sigma}_o^{2}$ for $\hat{\kappa}$, and further by denoting the weight coefficient matrix of the estimated parameters $\hat{X}$, i.e. N^{-1}, by Q, the equations (6.90) and (6.91) become:

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{df} \quad , \tag{6.98}$$

$$\hat{\Sigma}_{\hat{X}} = \hat{\sigma}_o^2 Q \quad . \tag{6.99}$$

Example 6.18: Let us compute the estimated variance-covariance matrix $\hat{\Sigma}_{\hat{x}}$ of the adjusted parameters $\hat{X}$ in example 6.16. The $\hat{\Sigma}_{\hat{x}}$ matrix is computed from equation (6.99). First, from the above mentioned example we have:

$$\underset{1,3}{\hat{V}^T} = \frac{H_J - H_G - \sum_i \bar{h}_i}{\sum_i d_i} [d_1, d_2, d_3],$$

$$\underset{3,3}{P} = \text{diag}\left[\frac{1}{d_1}, \frac{1}{d_2}, \frac{1}{d_3}\right]$$

and $df = n - u = 3 - 2 = 1$.

Hence,

$$\underset{1,3}{\hat{V}^T P} = \frac{H_J - H_G - \sum_i \bar{h}_i}{\sum_i d_i} [1, 1, 1],$$

$$\hat{V}^T P \hat{V} = (H_J - H_G - \sum_i \bar{h}_i)^2 / \sum_i d_i$$

and

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{df} = (H_J - H_G - \sum_i \bar{h}_i)^2 / d_i \quad .$$

As we have seen, $N^{-1} = Q$ is given by

$$\underset{2,2}{Q} = N^{-1} = \frac{d_1 d_2 d_3}{\sum_i d_i} \begin{bmatrix} \frac{d_2 + d_3}{d_2 d_3} & , & \frac{1}{d_2} \\ \frac{1}{d_2} & , & \frac{d_1 + d_2}{d_1 d_2} \end{bmatrix} .$$

We thus obtain finally

$$\hat{\Sigma}_{\hat{X}} = \hat{\sigma}_o^2 Q = \left(\frac{H_J - H_G - \sum_i \bar{h}_i}{\sum_i d_i}\right)^2 \begin{bmatrix} d_1(d_2 + d_3) & , & d_1 d_3 \\ d_1 d_3 & , & d_3(d_2 + d_1) \end{bmatrix} .$$

Example 6.19: Let us compute the estimated variance-covariance matrix $\hat{\Sigma}_{\hat{X}\hat{X}}$ of the adjusted parameters $\hat{X}$ in example 6.17. We are going to use equations (6.98) and (6.99). First, from the above mentioned example we have

$$\underset{1,6}{\hat{V}^T} = [0.00,\ 0.02,\ 0.02,\ -0.04,\ -0.04,\ 0.04]$$

in metres,

$$\underset{6,6}{P} = \text{diag}\,[0.25,\ 0.5,\ 0.5,\ 0.25,\ 0.5,\ 0.25]$$

in m^{-2} and

$$df = n - u = 6 - 3 = 3.$$

Hence

$$\hat{V}^T P \hat{V} = 0.002 \text{ (unitless)},$$

and

$$\hat{\sigma}_o^{\,2} = \frac{V^T P V}{df} = \frac{0.002}{3} \doteq 0.00067 \text{ (unitless)}.$$

Also, from example 6.17, we have

$$\underset{3,3}{Q} = N^{-1} = \begin{bmatrix} 1.6 & 0.8 & 0.8 \\ 0.8 & 1.6 & 0.8 \\ 0.8 & 0.8 & 1.2 \end{bmatrix} \text{ in } m^2.$$

Finally,

$$\underset{3,3}{\hat{\Sigma}_{\hat{X}\hat{X}}} = \hat{\sigma}_o^{\,2} Q = 10^{-4} \begin{bmatrix} 10.67 & 5.33 & 5.33 \\ 5.33 & 10.67 & 5.33 \\ 5.33 & 5.33 & 8.0 \end{bmatrix} \text{ in } m^2,$$

or

$$\hat{\Sigma}_{\hat{X}\hat{X}} = \begin{bmatrix} 10.67 & 5.33 & 5.33 \\ 5.33 & 10.67 & 5.33 \\ 5.33 & 5.33 & 8.0 \end{bmatrix} \text{ in } cm^2.$$

6.4.8 Some Properties of the Parametric Adjustment Solution Vector

It can be shown that the choice of the weight matrix P of the observations $\bar{L}$ (proportional to the inverse of variance-covariance matrix $\Sigma_{\bar{L}}$) and the choice of the least-squares method (minimization of $V^T PV$) to get the solution $X = \hat{X}$ ensures that the resulting estimate $\hat{X}$ has got the smallest possible trace of its variance-covariance matrix $\Sigma_{\hat{X}}$. In other words: taking $P = \hat{\sigma}_o^2 \, \Sigma_L^{-1}$ and seeking $\min_{X \in R^u} V^T PV$, provides such a solution $\hat{X}$ that satisfies at the same time the condition

$$\min_{X \in R^u} \text{trace } \Sigma_{\hat{X}} \, . \tag{6.100}$$

This is a result similar to the consequence of the least squares principle applied to random multivariate (section 5.4) and we are not going to prove it here.

Similarly, it can be shown that for uncorrelated multisample of observations $L = (L_1, L_2, \ldots, L_n)$ which are assumed to be normally distributed with PDF given by:

$$\phi(L_o, S; L) = \prod_{i=1}^{n} \frac{1}{S_i \sqrt{(2\pi)}} \exp \left[- \frac{(L_i - L_i^o)^2}{2 S_i^2} \right], \tag{6.101}$$

we get the most probable estimate of L_o if the condition $\min_{X \in R^u} V^T PV$ is satisfied. This can be verified by writing

$$\phi(L_o, S; L) = \frac{1}{(2\pi)^{n/2} \prod_{i=1}^{n} S_i} \exp \left[-\frac{1}{2} \sum_{i=1}^{n} \frac{V_i^2}{S_i^2} \right]$$

$$= \frac{1}{(2\pi)^{n/2} \prod_{i=1}^{n} S_i} \exp \left[- \frac{1}{2\kappa} V^T PV \right],$$

which is maximum if both $V^T PV$ and trace $(\Sigma_{\hat{X}})$ are minimum. This is valid for any fixed κ.

6.4.9 Relative Weights, Statistical Significance of A Priori and A Posteriori Variance Factors

We have seen in section 6.4.6 that the choice of the a priori variance factor σ_o^2, or κ, does not influence the estimated solution vector $\hat{X}$. Also, in section 6.4.7 we have seen that the same holds true even for the estimated variance-covariance matrix $\hat{\Sigma}_{\hat{X}}$. Hence, for the purpose of getting the solution vector $\hat{X}$ along with its $\hat{\Sigma}_{\hat{X}}$, we can assume any relative weights, i.e. $P = \sigma_o^2 \Sigma_{\underline{L}}^{-1}$, with σ_o^2 chosen arbitrarily. On the other hand, the matrix of normal equations, i.e. $N = A^T PA$, and the estimated variance factor, i.e., $\hat{\sigma}_o^2 = \hat{V}^T P \hat{V}/df$, are influenced by the selection of σ_o^2.

These features of σ_o^2 are used in practice for two different purposes. First, is to render the magnitude of the elements of the normal equation matrix N such as to make the numerical process of its inversion the most precise. This is accomplished by choosing the value of σ_o^2 such as to make the average of the elements of N close to one.

The second purpose is to test the consistency of the mathematical model with the observations and to test the correctness of the assumed variance-covariance matrix $\Sigma_{\underline{L}}$. Usually, if we do not have any idea about the value of the variance factor σ_o^2, we assume $\sigma_o^2 = 1$. Then, after performing the least-squares adjustment, we get $\hat{\sigma}_o^2 = \hat{V}^T P \hat{V}/df$ as an estimate of the assumed σ_o^2. The ratio $\hat{\sigma}_o^2/\sigma_o^2$, provides some

testimony about the correctness of $\Sigma_{\bar{L}}$ and the consistency of the model. This ratio should be approaching 1. By assuming in particular, $\sigma_o^2 = 1$, we should end with $\hat{\sigma}_o^2 = 1$ as well. If this is not satisfied, we start looking into the assumed $\Sigma_{\bar{L}}$ and use the obtained $\hat{\sigma}_o^2$ from the adjustment instead of σ_o^2 in computing the weights. If the resulting new variances and covariances of the observations are beyond the expected range known from experience, we have to start examining the consistency of the mathematical model with the observations, i.e. if it really represents the correct relationship between the observed and the unknown quantities.

This approach is also used to help detecting the existing "systematic errors" in the observations L, that manifest themselves as deviations from the mathematical model. These deviations cause an "overflow" into the value of the quadratic form $V^T PV$ and consequently, into $\hat{\sigma}_o^2$.

The theoretical relation between the a priori and a posterior variance factors allows us to test statistically the validity of our hypothesis. However, this particular topic is going to be dealt with elsewhere. Let us just comment here on the results of the adjustment of the levelling network discussed in Examples 6.17 and 6.19. In computing the weight matrix P, we assumed $\sigma_o^2 = 1$. After the adjustment we obtained $\hat{\sigma}_o^2 \doteq 0.00067$. Thus the ratio $\sigma_o^2/\hat{\sigma}_o^2$ equals to 1500 which is considerably different from 1. This suggests that the variance-covariance matrix $\Sigma_{\bar{L}}$ was postulated too "pessimistically" and that the actual variances of the observations are much lower.

6.4.10 Conditional Adjustment

In this section we are going to deal with the adjustment of the linear model (6.68), i.e.

$$B \quad L \quad = \quad C \quad , \ (r < n), \tag{6.102}$$

which represents a set of "r" independent linear conditions between n observations L. Note that C is an r by 1 vector of constant values arising from the conditions.

For the adjustment, the above model is reformulated as:

$$B\ (\bar{L} + V) - C = 0$$

or, as we usually write it

$$B \quad V \quad + \quad W \quad = \quad 0\ ^{*)}, \tag{6.103}$$

where:

$$\boxed{W = B\ \bar{L} - C\ .} \tag{6.104}$$

The system of equations (6.103) is known as the condition equations, in which B is the coefficient matrix, V is the vector of discrepencies and W is the vector of constant values. We recall that "n" is the number of observations and "r" is the number of independent conditions. It should also be noted that no unknown parameters, i.e. vector X, appear in the condition equations. The discrepencies V are the only unknowns

---

* If we have a non-linear model $F(L) = 0$, it can be again linearized by Taylor's series expansion, yielding:

$$F(L) = F(L^\circ) + \frac{\partial F}{\partial L}\bigg|_{L=L^\circ} (L-L^\circ) + \ldots \ ,$$

in which we again neglect the higher order terms. Putting $V = (L-L^\circ)$, B for $\partial F/\partial \bar{L}$ and $W = F(L^\circ)$, we end up with the linearized condition equations of the form: $BV + W = 0$, which is the same as (6.103).

We wish again to get such estimate $\hat{V}$ of V that would minimize the quadratic form $V^T PV$, where $P = \sigma_o^2 \Sigma_{\bar{L}}^{-1}$ is the assumed weight matrix of the observations $\bar{L}$. The formualtion of this condition, i.e. $\min_{V \epsilon R^n} V^T PV$, is not as straightforward, as it is in the parametric case (section 6.4.6). This is due to the fact that V in equation (6.103) can not be easily expressed as an explicit function of B and W. However, the problem can be solved by introducing the vector K of r unknowns, called Lagrange's multipliers or correlates*). We can write:

$$\min_{V \epsilon R^n} V^T PV = \min_{V \epsilon R^n} [V^T PV + 2K^T (BV + W)] , \qquad (6.105)$$

since the second term on the right hand side equals to zero. Let us denote

$$\phi = V^T PV + 2K^T (BV + W) .$$

To minimize the above function, we differentiate with respect to V and equate the derivatives to zero. We get:

$$\frac{\partial \phi}{\partial} = 2\hat{V}^T P + 2K^T B = 0$$

which, when transposed, gives $P\hat{V} + B^T K = 0$.

The last equation can be solved for $\hat{V}$ and we obtain:

$$\boxed{\hat{V} = -P^{-1} B^T K .} \qquad (6.106)$$

This system of equations is known as the correlate equations.

Substituting equation (6.106) into (6.103), we eliminate V:

$$B(-P^{-1} B^T K) + W = 0 ,$$

or

$$\boxed{(BP^{-1}B^T) K = W .} \qquad (6.107)$$

This is the system of normal equations for conditional adjustment. It is usually written in the following abbreviated form:

---

* This is why the conditional adjustment is sometimes called: adjustment by correlates.

$$M\,K = W, \tag{6.108}$$

where

$$M = B\,P^{-1}\,B^T. \tag{6.109}$$

The solution of the above normal equations system for K yields:

$$K = M^{-1}\,W = (BP^{-1}B^T)^{-1}\,W. \tag{6.110}$$

Once we get the correlates K, we can compute the estimated residual vector $\hat{V}$ from the correlate equations (6.106). finally the adjusted observations $\hat{L}$ are computed from

$$\hat{L} = \bar{L} + \hat{V}. \tag{6.111}$$

In fact, if we are not interested in the intermediate steps, the formula for the adjusted observations $\hat{L}$ can be written in terms of the original matrices B and P and the vectors $\bar{L}$ and C. We get

$$\begin{aligned}\hat{L} &= \bar{L} + \hat{V} \\ &= \bar{L} - P^{-1}\,B^T\,K \\ &= \bar{L} - P^{-1}B^T(BP^{-1}B^T)^{-1}\,(B\,\bar{L} - C)\,.\end{aligned} \tag{6.112}$$

It can also be written in the following form:

$$\boxed{\hat{L} = (I - T)\,\bar{L} + HC\,,} \tag{6.113}$$

where:

$$\begin{aligned} I &\quad \text{is the identity matrix,} \\ T &= P^{-1}\,[B^T(BP^{-1}B^T)^{-1}\,B], \\ H &= P^{-1}\,B^T(BP^{-1}\,B^T)^{-1}\,. \end{aligned} \tag{6.114}$$

Example 6.20: Let us solve example 6.16 again, using this time the conditional method of adjustment. We have only one condition equation between the observed height differences $\bar{h}_i$, i = 1,2,3, and we thus note that the number of degrees of freedom

is the same as in example 6.16. Denoting $H_J - H_G$ by ΔH, the existing condition can be written as:

$$\Sigma \bar{h}_i = \Delta H.$$

After reformulation we get:

$$v_1 + v_2 + v_3 + (\bar{h}_1 + \bar{h}_2 + \bar{h}_3 - \Delta H) = 0,$$

which can be easily written in the matrix form as

$$\underset{1,3}{B}\ \underset{3,1}{V} + \underset{1,1}{W} = 0$$

where

$$B = [1, 1, 1], \quad V = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix},$$

and

$$W = \bar{h}_1 + \bar{h}_2 + \bar{h}_3 - \Delta H$$

$$= \sum_i \bar{h}_i - \Delta H .$$

The weight matrix of the observations is given by (see example 6.16):

$$\underset{3,3}{P} = \text{diag}\left(\frac{1}{d_1}, \frac{1}{d_2}, \frac{1}{d_3}\right)$$

and

$$\underset{3,3}{P^{-1}} = \text{diag}\,(d_1, d_2, d_3).$$

The system of normal equations for the correlates K is given by equation (6.108) as

$$\underset{1,1}{M}\ \underset{1,1}{K} = \underset{1,1}{W},$$

where

$$M = BP^{-1}B^T = (d_1 + d_2 + d_3) = \sum_i d_i.$$

The solution for K is

$$K = M^{-1} W = \frac{1}{\Sigma_i d_i} (\Sigma_i \bar{h}_i - \Delta H).$$

The estimated residuals are then computed from equation (6.106) as:

$$\hat{V} = -P^{-1} B^T K$$

$$= -\begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \left(\frac{\Sigma_i \bar{h}_i - \Delta H}{\Sigma_i d_i}\right)$$

$$= -\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \frac{\Sigma_i h_i - H}{\Sigma_i d_i}$$

and we get:

$$\hat{v}_i = \frac{\Delta H - \Sigma_i \bar{h}_i}{\Sigma_i d_i} d_i, \quad i = 1, 2, 3.$$

This is the same result as obtained in example 6.16 when using the parametric adjustment (note that $\Delta H = H_J - H_G$). In this particular problem, we notice that the adjustment divide the misclosure, i.e. $(\Delta H - \Sigma_i \bar{h}_i)$, to the individual observed height differences proportionally to the corresponding lengths of levelling sections, i.e. inversely proportionally to the individual MSE's. The adjusted observations are given by equation (6.111) i.e.

$$\hat{L} = \bar{L} + \hat{V},$$

or

$$\hat{h}_i = \bar{h}_i + \hat{v}_i, \quad i = 1, 2, 3.$$

This yields:

$$\hat{h}_i = \bar{h}_i + \frac{\Delta H - \sum_i \bar{h}_i}{\sum_i d_i} \, d_i .$$

Finally, the estimates of the unknown parameters, i.e. $\hat{X} = (\hat{H}_1, \hat{H}_2)$, are computed from the known elevations H_G, H_J and the adjusted observations $\hat{h}_i$, as follows:

$$\hat{H}_1 = H_G + \hat{h}_1$$

$$= H_G + \bar{h}_1 + \frac{d_1}{\sum_i d_i} \, (\Delta H - \sum_i \bar{h}_i)$$

and

$$\hat{H}_2 = H_J - \hat{h}_3$$

$$= H_J - \bar{h}_3 - \frac{d_3}{\sum_i d_i} \, (\Delta H - \sum_i \bar{h}_i) .$$

The results are again identical to the ones obtained from the parametric adjustment.

Example 6.21: Let us solve example 6.17 again, but this time using the conditional adjustment. The configuration of the levelling network in question is illustrated again in Figure 6.8, for convenience.

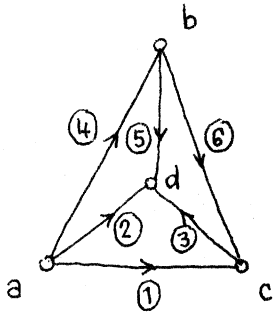


Figure 6.8

From the above mentioned example we have:

No. of observations, n = 6,

No. of unknown parameters, u = 3.

Then df = 6 - 3 = 3, and we shall see that we can again formulate only 3 independent condition equations between the given observations.

By examining Figure 6.8, we see that there are 4 closed loops, namely: (a - c - d - a), (a - d - b - a), (b - c - d - b) and (a - c - b - a).

This means that we can write 4 condition equations, one for each closed loop. However, one can be deduced from the other 3, e.g. the last mentioned loop is the summation of the other three loops.

Let us,for instance, choose the following three loops:

loop I = a - c - b - a ,

loop II = a - c - d - a ,

loop III = a - d - b - a .

These loops give the condition equations as follows

$$(\bar{h}_1 + v_1) - (\bar{h}_6 + v_6) - (\bar{h}_4 + v_4) = 0,$$

$$(\bar{h}_1 + v_1) + (\bar{h}_3 + v_3) - (\bar{h}_2 + v_2) = 0,$$

$$(\bar{h}_2 + v_2) - (\bar{h}_5 + v_5) - (\bar{h}_4 + v_4) = 0.$$

Then we get

$$v_1 - v_4 - v_6 + (\bar{h}_1 - \bar{h}_4 - \bar{h}_6) = 0 ,$$

$$v_1 - v_2 + v_3 + (\bar{h}_1 - \bar{h}_2 + \bar{h}_3) = 0 ,$$

$$v_2 - v_4 - v_5 + (\bar{h}_2 - \bar{h}_4 - \bar{h}_5) = 0 .$$

The above set of condition equations can be written in the matrix notation as

$$\underset{3,6}{B}\ \underset{6,1}{V} + \underset{3,1}{W} = 0 ,$$

where:

$$B_{3,6} = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & -1 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 0 \end{bmatrix} ,$$

$$y^T_{1,6} = (v_1, v_2, v_3, v_4, v_5, v_6)$$

and

$$W_{3,1} = \begin{bmatrix} (\bar{h}_1 - \bar{h}_4 - \bar{h}_6) \\ (\bar{h}_1 - \bar{h}_2 + \bar{h}_3) \\ (\bar{h}_2 - \bar{h}_4 - \bar{h}_5) \end{bmatrix} .$$

Substituting the observed quantities $\bar{h}_i$, i = 1, 2, ..., 6, into the above vector we get

$$W_{3,1} = \begin{bmatrix} 0.0 \\ 0.0 \\ -0.1 \end{bmatrix} \quad \text{in metres .}$$

The weight matrix P of the observations is formulated as: (see example 6.17):

$$P_{6,6} = \text{diag } (0.25, 0.5, 0.5, 0.25, 0.5, 0.25)$$

and

$$P^{-1}_{6,6} = \text{diag } (4, 2, 2, 4, 2, 4).$$

The normal equations for the correlates K are

$$M_{3,3} \; K_{3,1} = W_{3,1} ,$$

where

$$\underset{3,3}{M} = B\,P^{-1}\,B^T = \begin{bmatrix} 12 & 4 & 4 \\ 4 & 8 & -2 \\ 4 & -2 & 8 \end{bmatrix} .$$

By inverting M we get:

$$\underset{3,3}{M^{-1}} = \begin{bmatrix} 0.15 & -0.1 & -0.1 \\ -0.1 & 0.2 & 0.1 \\ -0.1 & 0.1 & 0.2 \end{bmatrix} .$$

The solution for K is given by

$$\underset{3,1}{K} = M^{-1}W = \begin{bmatrix} 0.15 & -0.1 & -0.1 \\ -0.1 & 0.2 & 0.1 \\ -0.1 & 0.1 & 0.2 \end{bmatrix} \begin{bmatrix} 0.0 \\ 0.0 \\ -0.1 \end{bmatrix} = \begin{bmatrix} 0.01 \\ -0.01 \\ -0.02 \end{bmatrix} .$$

The estimated residuals are computed from equation (6.106):

$$\underset{6,1}{\hat{V}} = - \; P^{-1}\,B^T\,K = \begin{bmatrix} 0.00 \\ 0.02 \\ 0.02 \\ -0.04 \\ -0.04 \\ 0.04 \end{bmatrix} \quad \text{m}$$

and are again identical with the results of example 6.17.

The adjusted observations will be.

$$\hat{L} = \bar{L} + \hat{V} \;, \quad \text{i.e.}$$

$$\begin{bmatrix} \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{h}_4 \\ \hat{h}_5 \\ \hat{h}_6 \end{bmatrix} = \begin{bmatrix} 6.16 \\ 12.57 \\ 6.41 \\ 1.09 \\ 11.58 \\ 5.07 \end{bmatrix} + \begin{bmatrix} 0.00 \\ 0.02 \\ 0.02 \\ -0.04 \\ -0.04 \\ 0.04 \end{bmatrix} = \begin{bmatrix} 6.16 \\ 12.59 \\ 6.43 \\ 1.05 \\ 11.54 \\ 5.11 \end{bmatrix}$$

in metres.

Finally, to compute the estimated elevations of points b, c, and d, i.e. $\hat{H}_b$, $\hat{H}_c$ and $\hat{H}_d$, we will use the given elevation H_a and the adjusted observations $\hat{h}_i$.

For instance;

$$\hat{H}_b = H_a + \hat{h}_4 = 0.0 + 1.05 = 1.05 \text{ m},$$

$$\hat{H}_c = H_a + \hat{h}_1 = 0.0 + 6.16 = 6.16 \text{ m},$$

$$\hat{H}_d = H_a + \hat{h}_2 = 0.0 + 12.59 = 12.59 \text{ m}.$$

These are obviously identical with the corresponding results of the parametric adjustment.

Note again that when computing the estimates of the unknown parameters from the adjusted observations we can follow any route in computing them. They all lead to the same answer.

6.4.11 Variance-Covariance Matrix of the Conditional Adjustment Solution

The formula for the variance-covariance matrix $\Sigma_{\hat{L}}$ of the adjusted observations - the "result" of the conditional least squares adjustment - can be developed by applying the law of propagation of variance-covariance matrix (equation 6.15) on equation (6.113). In this equation, the matrices I,T,H are, obviously, fixed. Similarly, the vector C is considered as a vector of theoretically deduced, and therefore errorless, values, then Σ_C will be zero. Hence, we get:

$$\Sigma_{\hat{L}} = \left(\frac{\partial \hat{L}}{\partial \bar{L}}\right) \Sigma_{\bar{L}} \left(\frac{\partial \hat{L}}{\partial \bar{L}}\right)^T$$

$$= (I - T)\, \Sigma_{\bar{L}}\, (I - T)^T$$

$$= \Sigma_{\underline{L}} - I\Sigma_{\underline{L}} T^T - T\Sigma_{\underline{L}} I + T\Sigma_{\underline{L}} T^T. \qquad (6.115)$$

It is not difficult to see that both $(I\Sigma_{\underline{L}} T^T)$ and $(T\Sigma_{\underline{L}} I)$ are square symmetric matrices, hence we can write:

$$\Sigma_{\hat{L}} = \Sigma_{\underline{L}} - 2T\Sigma_{\underline{L}} I + T\Sigma_{\underline{L}} T^T$$

$$= \Sigma_{\underline{L}} - T\Sigma_{\underline{L}} (2I - T^T). \qquad (6.116)$$

Recalling, from equation 6.114, that $T = P^{-1} [B^T(BP^{-1}B^T)^{-1}B]$ and knowing that $P = \sigma_o^2 \Sigma_{\underline{L}}^{-1}$, i.e.

$\Sigma_{\underline{L}} = \sigma_o^2 P^{-1}$, then by substituting these quantities into equation (6.116) we get:

$$\Sigma_{\hat{L}} = \sigma_o^2 P^{-1} - \sigma_o^2 P^{-1} [B^T(BP^{-1}B^T)^{-1}B] P^{-1} \times$$

$$\times \{2I - (P^{-1}[B^T(BP^{-1}B^T)^{-1}B]^T\} \qquad (6.117)$$

$$= \sigma_o^2 P^{-1} \{ I - 2[B^T(BP^{-1}B^T)^{-1}B] P^{-1} +$$

$$+ B^T \underbrace{(BP^{-1}B^T)^{-1}}_{M^{-1}} \underbrace{(BP^{-1}B^T)}_{M} \underbrace{(BP^{-1}B^T)^{-1}}_{M^{-1}} BP^{-1}\}.$$

Noting that

$$((BP^{-1}B^T)^{-1})^T = (BP^{-1}B^T)^{-1}$$

we get finally

$$\boxed{\Sigma_{\hat{L}} = \sigma_o^2 P^{-1} (I - B^T (BP^{-1}B^T)^{-1} BP^{-1}).} \qquad (6.118)$$

Here, similar to the parametric adjustment, to obtain the estimated variance-covariance matrix we use $\hat{\sigma}_o^2$ instead of σ_o^2, where:

$$\boxed{\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{df} = \frac{\hat{V}^T P \hat{V}}{r}} \qquad (6.119)$$

and we end up with

$$\hat{\Sigma}_{\hat{L}} = \hat{\sigma}_o^2 \, (P^{-1} - P^{-1} B^T (BP^{-1}B^T)^{-1} BP^{-1}), \qquad (6.120)$$

or, in an abbreviated form:

$$\hat{\Sigma}_{\hat{L}} = \hat{\sigma}_o^2 \, (I - T)P^{-1} \quad *) . \qquad (6.121)$$

Analogous to the parametric adjustment, it also can be shown that the estimate $\hat{L}$ assures the minimum trace of its variance-covariance matrix $\Sigma_{\hat{L}}$. Under the same assumptions as stated in section 6.4.8, the estimate $\hat{L}$ is also the most probable estimate of L.

Regarding the correlation between the adjusted observations $\hat{L}$, we can see that $\hat{L}$ will be uncorrelated if: (i) L is uncorrelated and (ii) the coefficient matrix B is orthogonal. If these two conditions are satisfied then T and P^{-1} will be diagonal matrices. On the other hand, we can experience uncorrelated $\hat{L}$ even for some other general T and P.

Finally, we note that again the choice of the a priori variance factor σ_o^2 does not influence the estimated $\Sigma_{\hat{L}}$ defined by equation (6.121).

Example 6.22: Let us determine the variance-covariance matrix $\hat{\Sigma}_{\hat{L}}$ for the conditional adjustment formulated in example 6.20.

We have

$$\hat{v}_i = \frac{\Delta H - \sum \bar{h}_i}{\sum_i d_i} \, d_i \quad , \; i = 1, 2, 3,$$

and $P_{3,3} = \text{diag} \left(\frac{1}{d_1}, \frac{1}{d_2}, \frac{1}{d_3}\right)$.

Thus we get

---

* It can be shown that similarly $\hat{\Sigma}_{\hat{V}} = \hat{\sigma}_o^2 \, TP^{-1}$.

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{r} = \frac{(\Delta H - \sum_i \bar{h}_i)^2}{\sum_i d_i} .$$

The required variance-covariance matrix is given by equation 6.121, i.e.

$$\hat{\Sigma}_{\hat{L}\hat{L}} = \hat{\sigma}_o^2 (I - T) P^{-1}.$$

First we compute $T = P^{-1} B^T M^{-1} B$. We recall from example 6.20 that $M^{-1}_{1,1} = 1/\sum_i d_i$, $B_{1,3} = [1,1,1]$ and $P^{-1}_{3,3} = \text{diag}(d_1, d_2, d_3)$.

Hence,

$$T = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \left[\frac{1}{\sum_i d_i} \right] [1, 1, 1]$$

and we obtain

$$T_{3,3} = \frac{1}{\sum_i d_i} \begin{bmatrix} d_1 & d_1 & d_1 \\ d_2 & d_2 & d_2 \\ d_3 & d_3 & d_3 \end{bmatrix}$$

Further we get

$$(I - T) P^{-1} = \begin{bmatrix} \alpha_1 & \alpha_1 & \alpha_1 \\ \alpha_2 & \alpha_2 & \alpha_2 \\ \alpha_3 & \alpha_3 & \alpha_3 \end{bmatrix}$$

where $\alpha_i = d_i - \frac{d_i^2}{\sum_i d_i}$, $i = 1, 2, 3$.

Finally we get:

$$\underset{3,3}{\hat{\Sigma}_{\hat{L}}} = \hat{\sigma}_o^2 (I - T) P^{-1}$$

$$= \frac{(\Delta H - \sum_i \bar{h}_i)^2}{\sum_i d_i} \begin{bmatrix} \alpha_1 & \alpha_1 & \alpha_1 \\ \alpha_2 & \alpha_2 & \alpha_2 \\ \alpha_3 & \alpha_3 & \alpha_3 \end{bmatrix} .$$

Example 6.23: Let us determine the variance-covariance matrix $\hat{\Sigma}_{\hat{L}}$ for the conditionally adjusted levelling net of example 6.21.

We have

$$\underset{1,6}{\hat{V}^T} = [0.00, 0.02, 0.02, -0.04, -0.04, 0.04]$$

in metres,

$$\underset{6,6}{P} = \text{diag } [0.25, 0.5, 0.5, 0.25, 0.5, 0.25]$$

in m^{-2}, and

$$r = df = n - u = 6 - 3 = 3.$$

Hence,

$$\hat{V}^T P \hat{V} = 0.002 \quad \text{(unitless)},$$

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{r} = \frac{0.002}{3} = 0.00067 \quad \text{(unitless)}.$$

The required $\hat{\Sigma}_{\hat{L}}$ matrix is computed again from equation 6.121 as

$$\underset{6,6}{\hat{\Sigma}_{\hat{L}}} = \hat{\sigma}_o^2 (I - T) P^{-1}$$

where

$$\underset{6,6}{P^{-1}} = \text{diag } [4, 2, 2, 4, 2, 4]$$

and T is computed from

$$T_{6,6} = P^{-1} (B^T M^{-1} B).$$

From example 6.21, we have:

$$M^{-1} = (BP^{-1}B^T)^{-1} = \begin{bmatrix} 0.15 & -0.1 & -0.1 \\ -0.1 & 0.2 & 0.1 \\ -0.1 & 0.1 & 0.2 \end{bmatrix}$$

and

$$B_{3,6} = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & -1 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 0 \end{bmatrix} .$$

Hence,

$$(B^T M^{-1} B) =$$

$$= \begin{bmatrix} 0.15 & -0.1 & 0.1 & -0.05 & 0 & -0.05 \\ -0.1 & 0.2 & -0.1 & -0.1 & -0.1 & 0 \\ 0.1 & -0.1 & 0.2 & 0 & -0.1 & 0.1 \\ -0.05 & -0.1 & 0 & 0.15 & 0.1 & 0.05 \\ 0 & -0.1 & -0.1 & 0.1 & 0.2 & -0.1 \\ -0.05 & 0 & 0.1 & 0.05 & -0.1 & 0.15 \end{bmatrix}$$

and: $T_{6,6} = P^{-1} (B^T M^{-1} B) =$

$$= \begin{bmatrix} 0.6 & -0.4 & 0.4 & -0.2 & 0 & -0.2 \\ -0.2 & 0.4 & -0.2 & -0.2 & -0.2 & 0 \\ 0.2 & -0.2 & 0.4 & 0 & -0.2 & 0.2 \\ -0.2 & -0.4 & 0 & 0.6 & 0.4 & 0.2 \\ 0 & -0.2 & -0.2 & 0.2 & 0.4 & -0.2 \\ -0.2 & 0 & 0.4 & 0.2 & -0.4 & 0.6 \end{bmatrix} .$$

Hence

$$(I - T) P^{-1} =$$

$$\begin{bmatrix} 1.6 & -0.8 & 0.8 & -0.8 & 0 & -0.8 \\ -0.8 & 1.2 & -0.4 & -0.8 & -0.4 & 0 \\ 0.8 & -0.4 & 1.2 & 0 & -0.4 & 0.8 \\ -0.8 & -0.8 & 0 & 1.6 & 0.8 & 0.8 \\ 0 & -0.4 & -0.4 & 0.8 & 1.2 & -0.8 \\ -0.8 & 0 & 0.8 & 0.8 & -0.8 & 2.4 \end{bmatrix}.$$

Finally we get

$$\hat{\Sigma}_{\hat{L}\,6,6} = \hat{\sigma}_o^2 \; (I - T) \; P^{-1} \text{ as}$$

$$= 10^{-4} \begin{bmatrix} 10.72 & -5.36 & 5.36 & -5.36 & 0 & -5.36 \\ -5.36 & 8.04 & -2.68 & -5.36 & -2.68 & 0 \\ 5.36 & -2.68 & 8.04 & 0 & -2.68 & 5.36 \\ -5.36 & -5.36 & 0 & 10.72 & 5.36 & 5.36 \\ 0 & -2.68 & -2.68 & 5.36 & 8.04 & -5.36 \\ -5.36 & 0 & 5.36 & 5.36 & -5.36 & 16.08 \end{bmatrix}$$

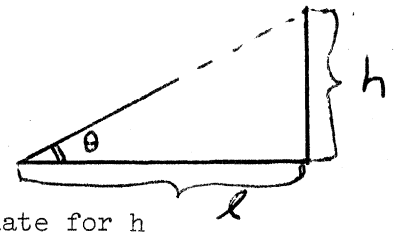
in metres squared.

By dropping the scalar 10^{-4} we get the results in cm^2.

The comments stated at the end of section 6.4.9 regarding the value of $\hat{\sigma}_o^2$ versus the assumed value of 1.0 for σ_o^2 hold true here as well.

6.5 Exercise 6

1. To determine the height h of a wall shown in the Figure, the horizontal distance ℓ and the vertical angle θ were observed and found to be:

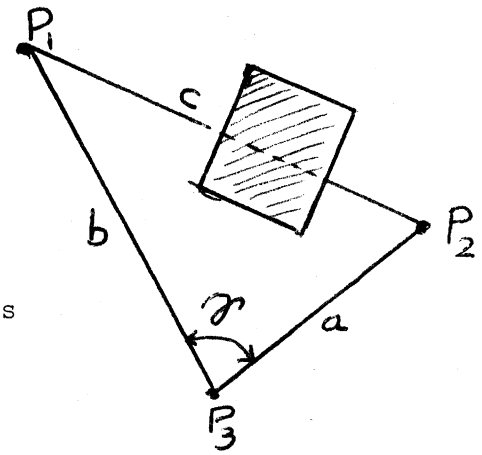


$\bar{\ell} = 85.34$ m, with $S_{\bar{\ell}} = 2$ cm.

$\bar{\theta} = 12°37'30''$, with $S_{\bar{\theta}} = 10''$.

Required: Compute the statistical estimate for h along with its RMS.

2. To determine the distance $\overline{P_1P_2} = c$, which cannot be measured directly due to the existence of some obstacle as shown in the Figure, the following measurements were taken:



$\overline{P_2P_3} = \bar{a} = 30$ m, with $S_{\bar{a}} = 3$ cm,

$\overline{P_1P_3} = \bar{b} = 40$ m, with $S_{\bar{b}} = 4$ cm,

$\bar{\gamma} = 60°$, with $S_{\bar{\gamma}} = 25''$.

Required: Compute the distance $\overline{P_1P_2}$ and its standard error to the nearest mm.

3. Determine the standard error of the estimated height h of the tower given in Problem number 9, Exercise 4, section 4.11. Consider all the measured quantities, namely ℓ, α, β and θ to be uncorrelated.

4. From a point P_o in the x-y coordinate system shown below, a distance $\bar{d} = 5637.8$ m and an azimuth $\bar{T} = 49.9873$ grads (100 grads = 90 degrees) to a second point P_1 were measured. The relative error of $\bar{d}$ is $1.2 \cdot 10^{-4}$. The RMS of $\bar{T}$ is 0.08 centigrads (1 grad = 100 centigrads).

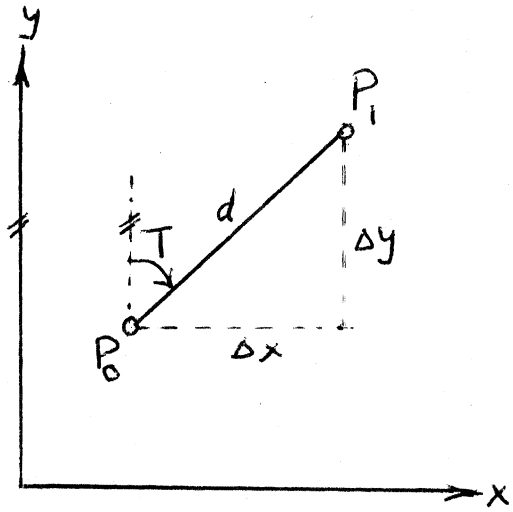
Required: Compute the following:

(i) The coordinate differences $(\hat{\Delta x}, \hat{\Delta y})$ between points P_o and P_1.

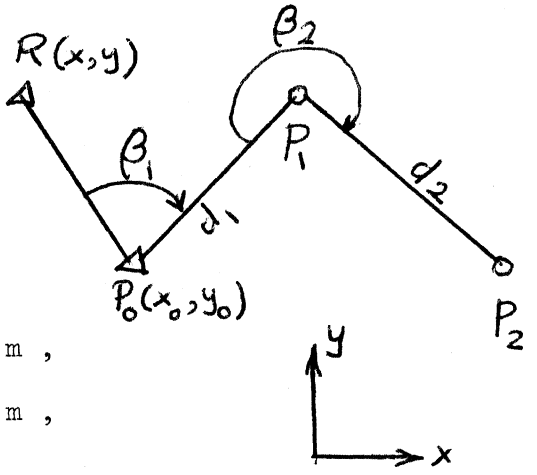
(ii) The variance-covariance matrix $\Sigma_{\hat{X}}$, where $\hat{X} \equiv (\hat{\Delta x}, \hat{\Delta y})$, in m^2.

(iii) The RMS of $\hat{\Delta x}$ and $\hat{\Delta y}$ respectively.

(iv) The correlation coefficient between Δx and Δy.



5. The shown traverse consists of two legs $\overline{P_oP_1}$ and $\overline{P_1P_2}$. The coordinates (x_o, y_o) of the initial point P_o as well as the (x, y) coordinates of the reference mark R are considered to be error-free (errorless), i.e. fixed quantities. The measured quantities are the horizontal angles β_1 and β_2 and the horizontal distances d_1 and d_2 respectively. The available data are:



$x = 100.0$ m , $y = 200.0$ m ,

$x_o = 150.0$ m , $y_o = 150.0$ m ,

$\bar{\beta}_1 = 75°$, with $S_{\bar{\beta}_1} = 3''$,

$\bar{\beta}_2 = 270°$, with $S_{\bar{\beta}_2} = 2''$,

$\bar{d}_1 = 100$ m and $\bar{d}_2 = 200$ m.

The standard error of the observed distance is to be calculated according to the formula :

$$S_{\bar{d}}(cm) = 1.0\ (cm) + d(m)\ .\ 10^{-2}.$$

Required: Compute the following:

(i) The estimated coordinates $(\hat{x}_1, \hat{y}_1)$ of point P_1, along with their associated variance-covariance matrix $\Sigma_{(\hat{x}_1, \hat{y}_1)}$.

(ii) The estimated coordinates $(\hat{x}_2, \hat{y}_2)$ of point P_2 and their variance-covariance matrix $\Sigma_{(\hat{x}_2, \hat{y}_2)}$.

Note that the coordinates are required to the nearest mm and the variances and covariances are required in cm^2.

(iii) Discuss the correlation among the estimated coordinates $\hat{x}_1$, $\hat{y}_1$, $\hat{x}_2$ and $\hat{y}_2$.

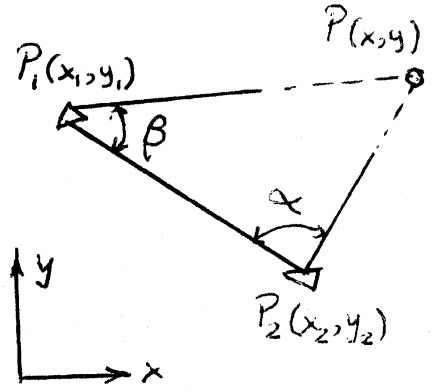
6. Having an intersection problem, as shown in the Figure, i.e. observing the two horizontal angles β and α from the two known stations P_1 (x_1, y_1) and P_2 (x_2, y_2) in order to determine the (x, y) coordinates of an unknown station P.

Given: the following data:

$\bar{x}_1 = 200.0$ m
$\bar{y}_1 = 500.0$ m , $\Sigma_{(\bar{x}_1, \bar{y}_1)} = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}$ cm^2 ,

$\bar{x}_2 = 546.4$ m
$\bar{y}_2 = 300.0$ m , $\Sigma_{(\bar{x}_2, \bar{y}_2)} = \begin{bmatrix} 2 & -0.5 \\ -0.5 & 3 \end{bmatrix}$ cm^2 ,

$\bar{\alpha} = 90°$, $S_{\bar{\alpha}} = 3''$
$\bar{\beta} = 60°$, $S_{\bar{\beta}} = 2''$ and $S_{\bar{\alpha}\bar{\beta}} = 0$.



Required: Compute the estimated $(\hat{x}, \hat{y})$ coordinates of the unknown station P, in metres, and their associated variance-covariance matrix $\Sigma_{(\hat{x}, \hat{y})}$ in cm^2.

7. Consider problem number 1 of this exercise. Assume that the observed quantities ℓ and θ have got also non-random (systematic) errors of -1 cm and $5''$ respectively. Compute the expected total error in the derived height h in centimetres.

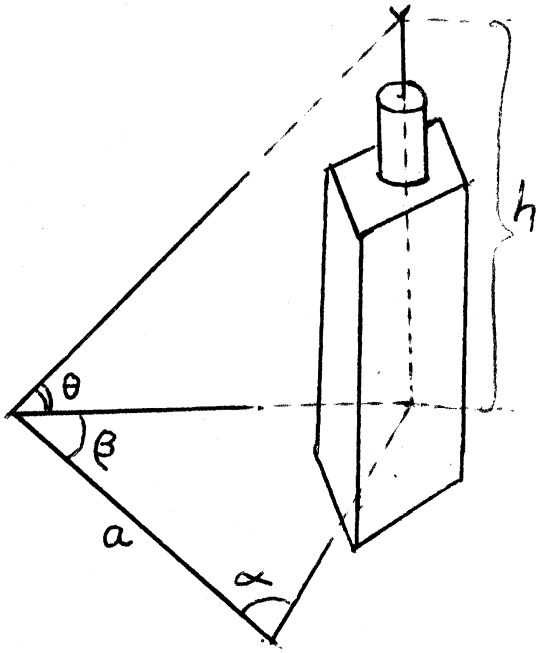
8. Determine the expected error in the sum of a hundred numbers in the following two cases:

   (i) each individual number is to be rounded-off to three decimal places.

   (ii) each individual number is to be truncated to three decimal places.

   Then compare and comment on the results.

9. To determine the height h of a tower, the technique shown in the Figure was proposed, in which α, β, θ and a are the quantities to be measured. The approximate values of these quantities were obtained from a preliminary investigation and found to be:



   $\alpha = 60°$, $\beta = 30°$,

   $\theta = 45°$ and $a = 100$ m.

   Providing that the horizontal angles α and β are to be measured with a precision of $2''$ (i.e. $S_{\bar{\alpha}} \doteq S_{\bar{\beta}} \doteq 2''$), what are the required precisions in measuring both the horizontal distance a and the vertical angle θ (i.e. $S_{\bar{a}}$ and $S_{\bar{\theta}}$) such that their contributions to the standard error $S_{\hat{h}}$ of the derived height $\hat{h}$ - which is specified not to be worse than 2.45 cm - will be the same.

10. Assume that the horizontal angles in a triangulation network are to be measured using two theodolites, "I" and"II", of different quality. These two theodolites were tested by measuring one particular angle several times, from which it was found that the standard deviation of one observation, i.e. the standard deviation of the sample, observed with theodolite "I" was $S_{L_1} = 1''.5$ and with theodolite "II" was $S_{L_2} = 2''.5$. If it is specified that all the angles of the network are required to have a standard deviation of the mean, i.e. $S_{\bar{\ell}}$, not worse than $0''.5$, how many times should we measure each angle when using theodolite number"I" and when using number "II"?

11. The following observations of the length of an iron bar in metres are made on a comparator:

    3.284, 3.302, 3.253, 3.273, 3.310, 3.321, 3.304,

    3.295, 3.263, 3.270.

    Required: (i) the length of the bar (i.e. the mean);

    (ii) the RMS of one observation;

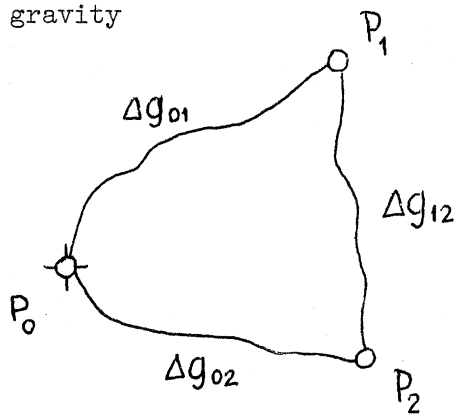
    (iii) the standard deviation of the mean.

12. The following table shows the means $\bar{\ell}_i$ of the daily measurements of the same distance ℓ during a five day period, along with their respective standard errors $S_{\bar{\ell}_i}$.

Day	Mon.	Tues.	Wed.	Thurs.	Fri.
$\bar{\ell}_i$ (m)	101.01	100.00	99.97	99.96	100.02
$S_{\bar{\ell}_i}$ (cm)	2	1	4	5	3

Required: Compute the weighted mean of the distance ℓ, say $\hat{\ell}$, along with its associated RMS, i.e. $\hat{S}_{\hat{\ell}}$.

13. Given a gravimetric network, as shown in the figure below, determine the gravity values g_1 and g_2 at points P_1 and P_2 respectively, with their variance-covariance matrix. The gravity g_o = 979832.12 mgal at the initial point P_o is known.

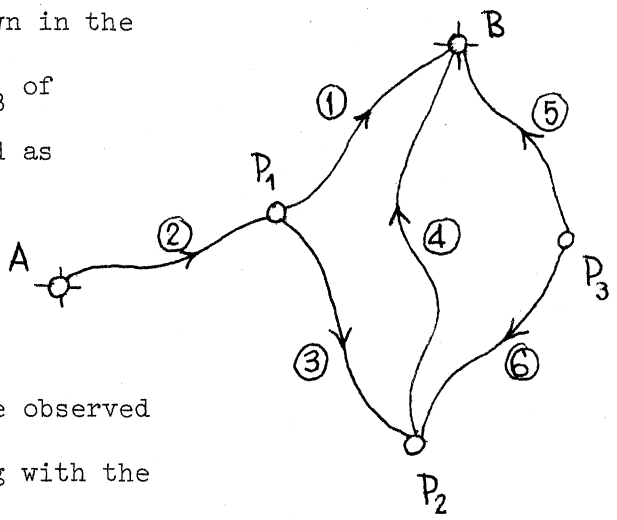


The following table gives the observed gravity differences with their signs, along with the time needed for each observation.

Station		Δg (mgal)	ΔT (hr)
From	To		
P_o	P_1	- 9.82	2.5
P_2	P_0	-27.78	1.5
P_1	P_2	+38.42	2.0

Assume that the observed differences Δg's are uncorrelated, and their variances are proportional to the corresponding time intervals ΔT.

14. Given a levelling net as shown in the Figure, the elevations H_A, H_B of points A and B are considered as known and errorless:



H_A = 300.000 m,

H_B = 302.245 m.

The following table gives the observed height differences $\bar{h}_i$'s along with the length ℓ_i of each section.

Section No.	Station		$\bar{h}_i$ (m)	ℓ_i (km)
	From	To		
1	P_1	B	1.245	1.0
2	A	P_1	0.990	0.5
3	P_1	P_2	0.500	1.0
4	P_2	B	0.751	1.0
5	P_3	B	1.486	0.5
6	P_3	P_2	0.740	1.5

Note that the arrows in the given figure indicate the directions of increasing elevations. The above observations are considered uncorrelated with $S^2_{\bar{h}_i}$ proportional to ℓ_i.

Required: Perform a parametric least squares adjustment of the above levelling net and find out the following:

(i) The estimated elevations $\hat{H}_1$, $\hat{H}_2$ and $\hat{H}_3$ of points P_1, P_2 and P_3.

(ii) The adjusted values of the given six height differences.

(iii) The estimated variance-factor $\hat{S}_o^2$ and compare it with the assumed, apriori variance factor S_o^2; comment on the results.

(iv) The estimated variance-covariance matrix $\hat{\Sigma}_{\hat{X}}$ of $\hat{X} = (\hat{H}_1, \hat{H}_2, \hat{H}_3)$.

15. Adjust the levelling net given in problem no. 14 again by using the conditional method of adjustment. Replace the requirement no. iv by computing the estimated variance-covariance matrix $\hat{\Sigma}_{\hat{L}}$ of the adjusted height differences. Compare the results of the other three requirements with the corresponding results from the parametric adjustment.

16. Two physical quantities Z and Y are assumed to satisfy the following linear model

$$Z = \alpha Y + \beta \ ,$$

where α and β are constants to be determined. The observations $\bar{Y}_i$ and $\bar{Z}_i$ obtained from an experiment are given in the following table.

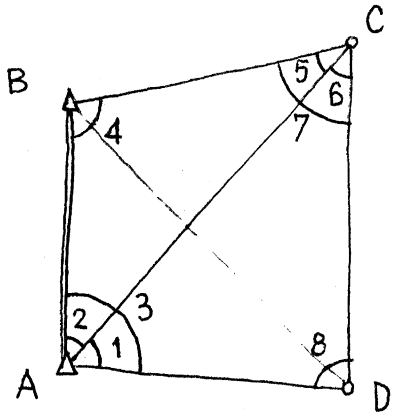
$\bar{Y}$	1	3	4	6	8	9	11	14
$\bar{Z}$	1	2	4	4	5	7	8	9

Assume that the $\bar{Y}$'s are errorless, and the $\bar{Z}$'s were observed with equal precision.

Required: Determine $\hat{\alpha}$ and $\hat{\beta}$ which provide the best fitting line between Z and Y, in the least squares sense.

17. Solve problem No. 16 again, but this time consider the $\bar{Z}$'s errorless, and the $\bar{Y}$'s with equal variances. Compare the results for $\hat{\alpha}$ and $\hat{\beta}$ with the corresponding results from problem No. 16.

18. The given figure shows a triangulation network with fixed base AB = 2 km. The eight numbered angles in the Figure are all measured, each with a different number n_i of observations as shown in the following table:

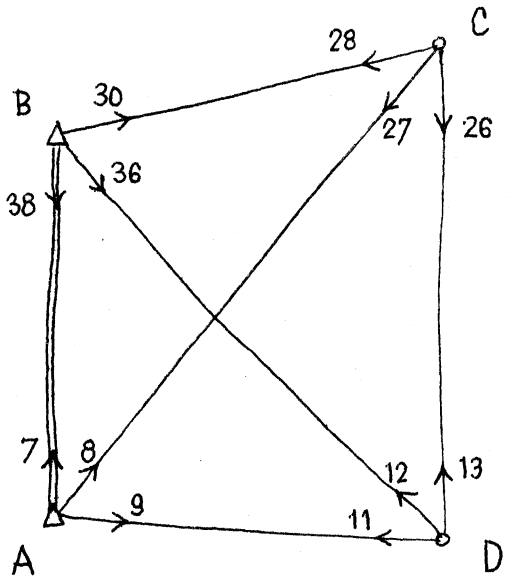


Angle No.	n_i	Mean value of the angle
1	2	82° 07' 09".50
2	2	28 22 17.70
3	5	110 29 25.02
4	3	125 53 33.67
5	2	25 44 09.30
6	2	29 19 17.50
7	5	55 03 29.32
8	3	68 33 32.33

Assume that all the measurements were done with the same instrument and under similar circumstances. (Hint: the weight of each angle will be proportional to the corresponding number of repetitions n_i). Required: (i) Neglecting the spherical excess in this network, compute the distance $\widehat{\overline{CD}}$ using the adjusted values of the observed angles.

(ii) Considering the fixed base $\overline{AB}$ to be errorless, find the estimated relative error of the estimated length $\hat{\overline{CD}}$.

19. The given figure shows a braced quadrilateral ABCD in a triangulation network, in which all the directions marked by arrowheads were measured with the same precision. The base AB = 25 km is assumed errorless. The spherical excess ε in the four formed triangles is computed approximately and given by:



ΔABC, $\varepsilon = 3''.126$,

ΔABD, $\varepsilon = 1''.556$,

ΔCBD, $\varepsilon = 3''.085$,

ΔACD, $\varepsilon = 1''.515$.

The results of the direction observations are summarized in the following table.

Occupied Station	Target Station	Direction No.	Observed Direction
A	B	7	00° 00' 00".00
	C	8	91 30 30.35
	D	9	125 53 33.91
B	C	30	00 00 00.00
	D	36	28 22 17.26
	A	38	110 29 27.13
C	D	26	00 00 00.00
	A	27	29 19 17.52
	B	28	35 03 26.80
D	A	11	00 00 00.00
	B	12	35 07 29.00
	C	13	68 33 32.60

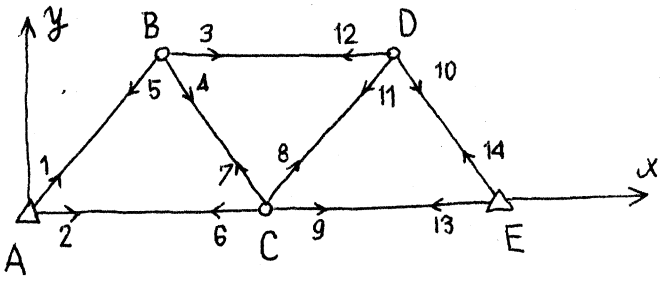
Required: (i) Perform a conditional adjustment and find out the adjusted values of the observed directions, along with their estimated variance-covariance matrix $\hat{\Sigma}_{\hat{L}}$.

(ii) Using Legendre's theorem for the spherical triangle, i.e. by subtracting one third of the spherical excess from each adjusted angle and then solving the triangle as if it was a plane triangle compute the side $\widehat{\overline{CD}}$ from the known base $\overline{AB}$ and the adjusted directions. Then check the computed $\widehat{\overline{CD}}$ by following another route in its computation.

(iii) Compute the estimated relative error of the estimated length $\widehat{\overline{CD}}$.

20. The given figure shows a triangulation network with two fixed (errorless) stations A and E whose x and y coordinates are:

	x	y
A:	0.0 m,	0.0 m,
E:	200.0 m,	0.0 m.



The 14 marked directions (with arrowheads) were observed with the same precision. The results of the observations are tabulated below.

Occupied Station	Target Station	Direction No.	Observed Directions
A	B	1	00° 00' 00".0
	C	2	60 00 10.0
B	D	3	00 00 00.0
	C	4	60 00 05.0
	A	5	119 59 50.0
C	A	6	00 00 00.0
	B	7	59 59 55.0
	D	8	120 00 00.0
	E	9	180 00 05.0
D	E	10	00 00 00.0
	C	11	59 59 45.0
	B	12	119 59 55.0
E	A	13	00 00 00.0
	D	14	60 00 15.0

Required: Prepare the input for a computer program performing a parametric least squares adjustment using the directions (not the angles) to estimate the unknown coordinates of points B, C and D by providing the following:

(i) The number of unknown parameters and the number of degrees of freedom.

(ii) The non-linear mathematical model.

(iii) The approximate values for the x, y coordinates of points B,C,D.

(iv) The linearized form of the mathematical model, i.e. $V = A\Delta X - \Delta\bar{L}$, giving the symbolic elements of the vectors V and ΔX and the numerical values of the elements of the design matrix A and the vector $\Delta\bar{L}$.

(v) Construct the variance-covariance matrix $\Sigma_{\Delta\bar{L}}$ of the observations assuming the standard errors in directions to equal 2".

APPENDIX I

Assumptions for and Derivation of the Gaussian PDF

The derivation of the Gaussian PDF presented here is due to G.H.L. Hagen (1837). The first formulation of the normal law, however, originates with De Moivre (1733).

(i) Let us assume that m independent physical causes are influencing the measurement. Let each cause contribute an elementary error either $+\Delta$ or $-\Delta$ towards the overall error ε. Any value of ε can thus be expressed as a combination (there are $n = 2^m$ such combinations) of m elementary errors $\pm \Delta$. We note first that the span of ε is $\langle -m\Delta, m\Delta \rangle$. Further, we realize that ε can attain only a value of an integral multiple of Δ. It is not difficult to see that any two adjacent values of ε differ by 2Δ since one is obtained from the other by replacing $-\Delta$ by $+\Delta$ and vice-versa. Dividing the range of ε

$$Ra\ (\varepsilon) = m\Delta - (-m\Delta) = 2m\Delta$$

by the step of ε, i.e. 2Δ, we discover that ε can attain any of the following $m + 1$ values

$$\varepsilon_i = (2_i - m)\Delta \quad , \quad i = 0, 1, \ldots m, \qquad (I\text{-}1)$$

corresponding to particular distinguishable combinations of the m elementary errors.

(ii) Let us regard now the set D of all permissible values of ε

$$D \equiv \{\varepsilon_o, \varepsilon_1, \cdots, \varepsilon_i, \cdots, \varepsilon_m\}$$

the probability space of the random sample consisting of all the 2^m combinations ε. Obviously, many of the 2^m combinations have the same values, because there are only m + 1 different values available. The counts, c_i, of the individual values ε_i (see section 3.1.1) can be computed using the combined probability (see section 2.3):

$$c_i = \binom{m}{i} = \frac{m\ (m-1)\ (m-2)\ \ldots\ (m-i+1)}{i\ (i-1)\ (i-2)\ \ldots\ 1} = \prod_{j=m-i+1}^{m} j \Big/ \prod_{j=1}^{i} j. \qquad \text{(I-2)}$$

The actual probability of any value ε_i is then given by

$$P(\varepsilon_i) = \frac{c_i}{n} = \binom{m}{i}/2^m\ , \qquad \text{(I-3)}$$

(see section 3.1.2).

(iii) The above formula describes the actual PDF of our sample ε in the discrete probability space D. Since our ultimate aim is to derive the analytic expression for the "corresponding" (we shall see later what is meant by corresponding) continuous PDF, we want to be able to express P as a function of ε_i rather than i. The easiest way to do it is to use the finite differences.

Let us define

$$\delta P(\varepsilon_i) = P(\varepsilon_i) - P(\varepsilon_{i-1})$$

and we get from equation (I-3)

$$\delta P(\varepsilon_i) = \binom{m}{i}/2^m - \binom{m}{i-1}/2^m$$

$$= \left[\binom{m}{i} - \binom{m}{i}\frac{i}{m-i+1}\right]/2^m$$

$$= \binom{m}{i}/2^m\ \left(1 - \frac{i}{m-i+1}\right).$$

Obviously, the ratio $\delta P\ (\varepsilon_i)/P(\varepsilon_i)$ is then given by

$$\delta P(\varepsilon_i)/P(\varepsilon_i) = 1 - i/(m-i+1). \qquad (I-4)$$

On the other hand, i can be expressed as a function of ε_i from equation (I-1)

$$2i - m = \varepsilon_i/\Delta$$

or

$$i = \frac{1}{2}(\varepsilon_i/\Delta + m).$$

Renoting $\delta\varepsilon = 2\Delta$ and substituting for i in equation (I-4) we obtain

$$\frac{\delta P(\varepsilon)}{P(\varepsilon)} = 1 - \frac{\varepsilon/\delta\varepsilon + m/2}{m - \varepsilon/\delta\varepsilon - m/2 + 1}$$

$$= \frac{1 + m/2 - \varepsilon/\delta\varepsilon - \varepsilon/\delta\varepsilon - m/2}{1 + m/2 - \varepsilon/\delta\varepsilon}$$

$$= \frac{1 - 2\varepsilon/\delta\varepsilon}{1 + m/2 - \varepsilon/\delta\varepsilon} = \frac{2\varepsilon - \delta\varepsilon}{(1 + m/2)\,\delta\varepsilon - \varepsilon}.$$

(iv) The next step in the development is to convert the discrete PDF, $P(\varepsilon)$, to a continuous PDF, i.e. to derive the "corresponding" continuous PDF. The "corresponding" PDF is assumed to be the PDF of such a variable ε which is defined the same way as the discrete ε in (i) with the exception that m is let to grow beyond all limits, i.e. $m \to \infty$. By letting m grow we would obtain infinitely large values of ε (see equation (I-1)). This would contradict our experience teaching us that the errors are always finite in value. Hence, we have to adopt another assumption and that is that as m grows to infinity, the absolute value of elementary error Δ grows to zero, making the product $m\Delta$ in equation (I-1) always finite.

Accepting these two assumptions we can write the finite difference equation as

$$\lim_{\substack{m \to \infty \\ \delta\varepsilon \to 0}} \frac{\delta P(\varepsilon)}{P(\varepsilon)} = - \lim_{\substack{m \to \infty \\ \delta\varepsilon \to 0}} \frac{2\varepsilon - \delta\varepsilon}{(1+m/2)\,\delta\varepsilon - \varepsilon} \tag{I-5}$$

which is nothing else but an ordinary differential equation for the continuous PDF $P(\varepsilon)$. It can thus be rewritten as

$$\frac{d\,P(\varepsilon)}{P(\varepsilon)} = - \frac{2\varepsilon - d\varepsilon}{(1 + m/2)\,d\varepsilon - \varepsilon} \quad .$$

To simplify the solution of this differential equation let us multiply both the numerator and denominator of the right hand side by $d\varepsilon$ and assume that $m d\varepsilon^2$ is **constant** We further assume

$$d\varepsilon^2 << \varepsilon d\varepsilon << m d\varepsilon^2 = C \ .$$

Then we can write

$$\frac{dP}{P} \doteq - \frac{2\,\varepsilon d\varepsilon}{C/2} = - \frac{4}{C}\,\varepsilon d\varepsilon \ . \tag{I-6}$$

(v) We can now finally solve the differential equation. It is solvable by direct integration and we get

$$\int \frac{dP}{P} \doteq - \int \frac{4}{C}\,\varepsilon d\varepsilon + \text{const.}$$

$$\ell n\ P \doteq - \frac{4}{C}\,\frac{\varepsilon^2}{2} + \text{const.}$$

Denoting the integration constant by $\ell n\ K$ we finally obtain

$$P \doteq K \exp\,(-2\varepsilon^2/C). \tag{I-7}$$

The question now arises whether we are free to regard both K and C as two independent parameters of the above PDF. We know that the basic equation for a PDF, i.e

$$\int_{-\infty}^{\infty} P(\varepsilon)\ d\varepsilon = 1 \qquad \text{(I-8)}$$

has to be satisfied. Substituting for P into the basic equation we get

$$\int_{-\infty}^{\infty} P(\varepsilon)\ d\varepsilon = \int_{-\infty}^{\infty} K \exp\ (-2\varepsilon^2/C)\ d\varepsilon =$$

$$= K \int_{-\infty}^{\infty} \exp\ (-2\varepsilon^2/C)\ d\varepsilon = 1$$

and $$K = 1 \Big/ \int_{-\infty}^{\infty} \exp\ (-2\varepsilon^2/C)\ d\varepsilon\ .$$

Hence the answer is that K must not be regarded as an independent parameter. It is a function of C and can be evaluated by solving the integral above. We obtain

$$\int_{-\infty}^{\infty} \exp\ (-2\varepsilon^2/C)\ d\varepsilon = 2 \int_{0}^{\infty} \exp\ (-2\varepsilon^2/C)\ d\varepsilon = \sqrt{\frac{C\pi}{2}} \qquad \text{(I-9)}$$

and $$K = \sqrt{\frac{2}{C\pi}}\ . \qquad \text{(I-10)}$$

The Gaussian PDF can then be written as

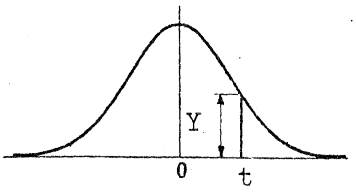
$$P(\varepsilon) = G\ (C;\varepsilon) = \sqrt{\frac{2}{C\pi}}\ \exp\ (-2\varepsilon^2/C) \qquad \text{(I-11)}$$

and we can see that it is a one-parametric PDF.

APPENDIX II - A

ORDINATES OF THE STANDARD NORMAL CURVE

$$Y = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$$



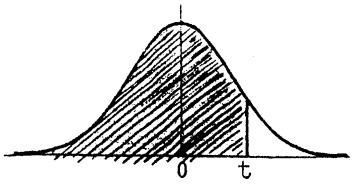
t	0	1	2	3	4	5	6	7	8	9
0.0	.3989	.3989	.3989	.3988	.3986	.3984	.3982	.3980	.3977	.3973
0.1	.3970	.3965	.3961	.3956	.3951	.3945	.3939	.3932	.3925	.3918
0.2	.3910	.3902	.3894	.3885	.3876	.3867	.3857	.3847	.3836	.3825
0.3	.3814	.3802	.3790	.3778	.3765	.3752	.3739	.3725	.3712	.3697
0.4	.3683	.3668	.3653	.3637	.3621	.3605	.3589	.3572	.3555	.3538
0.5	.3521	.3503	.3485	.3467	.3448	.3429	.3410	.3391	.3372	.3352
0.6	.3332	.3312	.3292	.3271	.3251	.3230	.3209	.3187	.3166	.3144
0.7	.3123	.3101	.3079	.3056	.3034	.3011	.2989	.2966	.2943	.2920
0.8	.2897	.2874	.2850	.2827	.2803	.2780	.2756	.2732	.2709	.2685
0.9	.2661	.2637	.2613	.2589	.2565	.2541	.2516	.2492	.2468	.2444
1.0	.2420	.2396	.2371	.2347	.2323	.2299	.2275	.2251	.2227	.2203
1.1	.2179	.2155	.2131	.2107	.2083	.2059	.2036	.2012	.1989	.1965
1.2	.1942	.1919	.1895	.1872	.1849	.1826	.1804	.1781	.1758	.1736
1.3	.1714	.1691	.1669	.1647	.1626	.1604	.1582	.1561	.1539	.1518
1.4	.1497	.1476	.1456	.1435	.1415	.1394	.1374	.1354	.1334	.1315
1.5	.1295	.1276	.1257	.1238	.1219	.1200	.1182	.1163	.1145	.1127
1.6	.1109	.1092	.1074	.1057	.1040	.1023	.1006	.0989	.0973	.0957
1.7	.0940	.0925	.0909	.0893	.0878	.0863	.0848	.0833	.0818	.0804
1.8	.0790	.0775	.0761	.0748	.0734	.0721	.0707	.0694	.0681	.0669
1.9	.0656	.0644	.0632	.0620	.0608	.0596	.0584	.0573	.0562	.0551
2.0	.0540	.0529	.0519	.0508	.0498	.0488	.0478	.0468	.0459	.0449
2.1	.0440	.0431	.0422	.0413	.0404	.0396	.0387	.0379	.0371	.0363
2.2	.0355	.0347	.0339	.0332	.0325	.0317	.0310	.0303	.0297	.0290
2.3	.0283	.0277	.0270	.0264	.0258	.0252	.0246	.0241	.0235	.0229
2.4	.0224	.0219	.0213	.0208	.0203	.0198	.0194	.0189	.0184	.0180
2.5	.0175	.0171	.0167	.0163	.0158	.0154	.0151	.0147	.0143	.0139
2.6	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110	.0107
2.7	.0104	.0101	.0099	.0096	.0093	.0091	.0088	.0086	.0084	.0081
2.8	.0079	.0077	.0075	.0073	.0071	.0069	.0067	.0065	.0063	.0061
2.9	.0060	.0058	.0056	.0055	.0053	.0051	.0050	.0048	.0047	.0046
3.0	.0044	.0043	.0042	.0040	.0039	.0038	.0037	.0036	.0035	.0034
3.1	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026	.0025	.0025
3.2	.0024	.0023	.0022	.0022	.0021	.0020	.0020	.0019	.0018	.0018
3.3	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014	.0013	.0013
3.4	.0012	.0012	.0012	.0011	.0011	.0010	.0010	.0010	.0009	.0009
3.5	.0009	.0008	.0008	.0008	.0008	.0007	.0007	.0007	.0007	.0006
3.6	.0006	.0006	.0006	.0005	.0005	.0005	.0005	.0005	.0005	.0004
3.7	.0004	.0004	.0004	.0004	.0004	.0004	.0003	.0003	.0003	.0003
3.8	.0003	.0003	.0003	.0003	.0003	.0002	.0002	.0002	.0002	.0002
3.9	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001

APPENDIX II - B

AREAS UNDER THE STANDARD NORMAL CURVE

from $-\infty$ to t

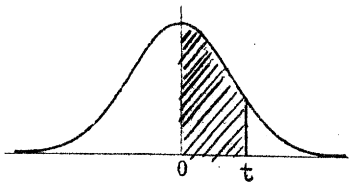
$$\Psi_n(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{x^2}{2}\right) dx$$



t	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5754
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7258	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7518	.7549
0.7	.7580	.7612	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7996	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

APPENDIX II - C

AREAS under the STANDARD NORMAL CURVE from 0 to t



t	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

BIBLIOGRAPHY

Anderson, T.W. 1966 (7th printing): An Introduction to Multivariate Statistical Analysis, Wiley & Sons.

Dorrer, E., 1966: Adjustment Computations, Department of Surveying Engineering, U.N.B., Lecture Notes No. 12.

Fraser, D.A.S., 1967 (4th printing): Statistics - An Introduction, Wiley & Sons.

Hamilton, A.C., 1964: Statistics in Physical Science, Ronald.

Hirvonen, R.A., 1971: Adjustment by Least Squares in Geodesy and Photogrammetry, Ungar.

Hogg, R.V., Craig, A.T., 1966 (5th printing): Introduction to Mathematical Statistics, MacMillan.

Lipschutz, S., 1968: Probability, Schaum's Outline Series, MacGraw-Hill.

Speigel, M.R., 1961: Statistics, Schaum's Outline Series, McGraw-Hill.

Van der Waerden, B.L., 1969: Mathematical Statistics, Springer-Verlag.

Wells, D.E. and Krakiwsky, E.J., 1971: The Method of Least Squares, Department of Surveying Engineering, U.N.B., Lecture Notes No. 18.

Wilks, S.S., 1963 (2nd printing): Mathematical Statistics, Wiley and Sons.

Wonnacott, T.H. and Wonnacott, R.J., 1972 (2nd edition): Introductory Statistics, Wiley & Sons.