PROBABILITY AND STATISTICS IN CIVIL ENGINEERING

An Introduction

G. N. Smith MSc, PhD, CEng, MICE

PROBABILITY AND STATISTICS IN CIVIL ENGINEERING

An Introduction

G. N. Smith MSc, PhD, CEng, MICE

Civil Engineering Department
Heriot-Watt University

NP

Nichols Publishing Company
New York

Nichols Publishing Company (W. G. Nichols Inc.)
PO Box 96, New York, New York 10024 and
155 West 72nd Street, New York, New York 10023.

Library of Congress Cataloging-in-Publication Data

Smith, G. N. (Geoffrey Nesbitt)
Probability and statistics in civil engineering.

1. Structural engineering—Statistical methods.
2. Soil mechanics—Statistical methods.
3. Probabilities. I. Title.
TA640.2.S55 1986 624'.01'5192 86-8344

First published in the United States of America
by Nichols Publishing Company 1986
First published in Great Britain by
Collins Professional and Technical Books 1986
8 Grafton Street, London W1X 3LA

Filmset by Eta Services (Typesetters) Ltd.,
Beccles, Suffolk
Printed and bound in Great Britain by
Mackays of Chatham, Kent

Preface

Until recently the ability of a structural element to withstand a particular loading, or to not deflect more than a prescribed limit, was expressed in terms of a single number, the factor of safety, F.

A drawback of this method is that the value of the factor of safety is simply a number obtained from a deterministic approach in which there is no allowance for any inherent variability within the design parameters and it is, perhaps, not surprising that there have been instances of structural failure where the calculated factor of safety was actually greater than 1.0.

Lumb (1970), speaking in the context of geotechnical engineering, summed up the situation:

> 'The traditional safety factor concept has the serious disadvantage that the actual variability of the soil strength is not directly taken into account and, consequently, a particular conventional safety factor value does not necessarily have the same meaning for all soils. Comparison of different designs with different soil types, or even different designs with the same soil type, is not easy, unless the conventional safety factors are so large as to preclude any practical risk of failure.'

The factor of safety, far from being of constant value, is really a random variable whose variability is due to the variability of the applied loads and the strength parameters of the structure.

If failure is defined as the event of F achieving a value equal to or less than 1.0 then the probability of this event is the probability of failure, P_f.

In Britain the first major step to allow for civil engineering uncertainties in design took place in 1972 when the Code of Practice CP110, *The Structural Use of Concrete*, was published by the British Standards Institution. This code adopted the policy of limit state design and probability theory was used, albeit indirectly, by the introduction of characteristic values.

Since 1972 the pressure for change has not diminished. Most of the proposed Eurocodes for structural design are now in draft form and advocate limit state design. There is a strong chance that these codes will either list values of partial safety factors, determined with the help of probability theory, or will make use of the reliability index which is becoming recognised as a powerful alternative to the use of partial safety factors in civil and structural engineering design. Along with this change there is a need for consultants, students, lecturers and research workers in civil engineering to at least become familiar with these developments and the new terms they involve.

This book is intended to present a straightforward summary of the most important aspects of statistics and probability theory that are relevant in civil engineering. Within these limitations the text is complete in that it should be possible for the reader to work through it without reference to other books.

The first three chapters deal with the fundamentals of statistics and its application to probability theory. At the end of these chapters there are exercises which, it is hoped, will be of assistance in the understanding of the subject matter. Those with knowledge of this material will be able to commence reading the book at chapter four, where the principles of reliability analysis are first discussed.

The author would like to take this opportunity of thanking those colleagues who gave helpful suggestions and encouragement during the preparation of this book. In particular he would like to thank Dr M. A. Paul of the Civil Engineering Department of Heriot-Watt University and Dr R. T. Murray of the Road Research Laboratory at Crowthorne.

G. N. Smith

Contents

Chapter One
Basic Probability Theory

Sets and events

The study of events and the probability of their happenings inevitably draws one towards the idea of the set.

In a test series of measurements the mean value obtained is an event resulting from the whole set of measured values.

A set is therefore a collection of items and, as with an event, is usually designated by a capital letter, A, B, C, etc. The individual elements that make up a set are generally denoted by lower case letters, a, b, c, …. For example, for set A:

$$A = a_1, a_2, a_3, a_4$$

Since the arrangement of elements does not affect a set, A is also given by:

$$A = a_3, a_4, a_1, a_2$$

The convention $a_1 \in A$ simply means that a_1 is an element of the set A.

In most civil engineering situations a set is defined by the listing of the elements within it, such as the measurements obtained for a particular test. However there are often occasions when it is not possible to determine the total elements of a set, although we know they exist, such as the infinite set of soil samples that could be collected from a particular stratum.

In such a situation, although the full set cannot be listed, the properties of the set can. For example a set B, consisting of all even numbers between 2 and 100, could be specified as:

$$B = [b\,;\, b \text{ is an even number between 2 and 100}]$$

where ';' means 'given that' or 'such that'.

Obviously set B could also have been listed as:

$$B = [2, 4, 6, 8, \ldots, 98, 100]$$

The universal set

The complete collection of all possible elements of a set is known as the universal set or the sample space and given the symbol Ω, the Greek letter omega, or the capital letter S.

Figure 1.1 shows the sample space, i.e. all the possible events, involved in the scores obtained from the throwing of two dice.

The sample space, such as the total 36 elements of Fig. 1.1, represents the certain event, in this case the event 'there will be some score'.

An impossible event is one which is outside the sample space, such as the event (7, 1) in Fig. 1.1.

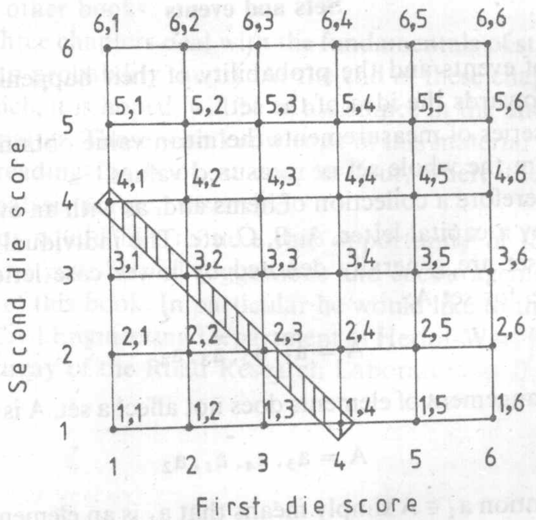


Fig. 1.1 Sample space for score of two dice

The subset

If B is a set of elements taken from a universal set, A, then B is referred to as a subset of A. This is expressed as:

$$B \subset A \quad \text{or} \quad A \supset B$$

meaning 'B is contained in A' or 'A contains B' respectively.

In Fig. 1.1 the subset [(4, 1), (3, 2), (2, 3), (1, 4)] represents the total number of ways of obtaining the score '5'. The event 'scoring 5' is by no means a single event as it can occur in 4 different ways.

Simple event

An event that can only occur once, such as the scoring of double 1, is referred to as a simple, or elementary, event.

Compound event

An event that can occur in more than one way is a compound event.

Union of sets (A ∪ B)

The union of two sets, A and B, is the set which contains all the elements that are in either A or B.

Intersection of sets (A ∩ B)

The intersection of two sets, A and B, is the set which contains all the elements that are in both A and B.

Example 1.1

The values of the blow count, *N*, measured during a series of penetration tests on a sand deposit were found to range between 1 and 15. The sample space representing all possible *N* values, therefore, is a set of figures, 1, 2, ..., 14, 15 and is shown in Fig. 1.2.

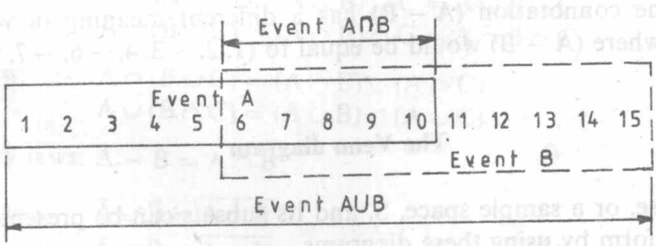


Fig. 1.2 Union and intersection of sets

If event A is that $1 \leqslant N \leqslant 10$ and the event B is that $6 \leqslant N \leqslant 15$ then the union of A and B, (A ∪ B), is the event that $1 \leqslant N \leqslant 15$ and the intersection of A and B, (A ∩ B) is the event that $6 \leqslant N \leqslant 10$.

The complement of a set

If B is a subset of A then $A > B$ and the set $(A - B)$ is called the complement of B relative to A and given the symbol $\bar{B}_A$.

If S is the total sample space then the set $(S - B)$ is known as the complement of B and given the symbol $\bar{B}$.

The complement of $A \cup B$ is written as $\overline{A \cup B}$.

Difference between sets

The set containing those elements of A that are not in B is the set $(A \cap \bar{B})$. Such a set is often referred to as the difference between A and B, as $(A \cap \bar{B})$ is numerically equal to $(A - B)$.

Example 1.2

A total sample space is the set (1, 2, 3, 4, 5, 6, 7, 8, 9, 10).

If A is (1, 2, 4, 5, 8, 9) and B is (3, 5, 6, 7, 8) show numerically that $(A \cap \bar{B})$ equals $(A - B)$.

Solution

$$\bar{B} = 1, 2, 4, 9, 10$$

Hence:

$$(A \cap \bar{B}) = (1, 2, 4, 9)$$

and:

$$(A - B) = (1, 2, 4, 9)$$

Note: The connotation $(A - B)$ has a different meaning in vectorial algebra where $(A - B)$ would be equal to $(1, 2, -3, 4, -6, -7, 9)$.

The Venn diagram

A universe, or a sample space, S, and its subsets can be presented in a pictorial form by using these diagrams.

The universal set, S, is represented as a rectangle with its subsets lying within it, as seen in Fig. 1.3a. The shaded area of Fig. 1.3b illustrates the difference set $(A - B)$ and the shaded area of Fig. 1.3c represents $\bar{A}$, the complement of set A.

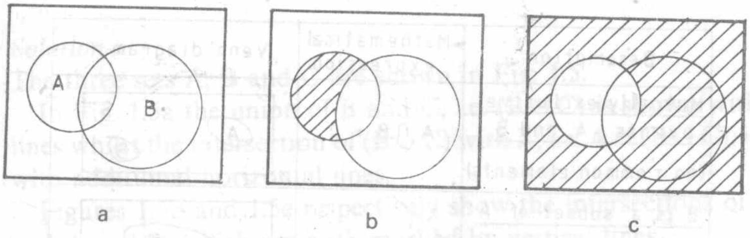


Fig. 1.3 The Venn diagram

The algebra of sets and events

Figure 1.1 illustrates that if an event can happen in several different ways then the event is a subset of the set of total possible events. It can be seen, therefore, that the algebra applicable to sets is identical to that for events. If A and B are events then, in set theory, the symbols mean:

(1) $A \cup B$ = the event 'the happening of either A or B or both'
(2) $A \cap B$ = the event 'the happening of both A and B'
(3) $\bar{A}$ = the event 'the non-happening of A'
(4) $A \cap \bar{B}$ = the event 'the happening of A but not B' = $(A - B)$

The Venn diagrams of Fig. 1.4 illustrate various set operations.

The most important theorems of set algebra are set out below and can be demonstrated by a study of the appropriate Venn diagrams.

Commutative law: $A \cup B = B \cup A$
$A \cap B = B \cap A$

Associative law: $A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$
$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$

Distributive law: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Complementary laws: $A - B = A \cap \bar{B}$

De Morgan's laws: $\bar{A} \cup \bar{B} = \overline{A \cap B}$
$\bar{A} \cap \bar{B} = \overline{A \cup B}$

Example 1.3

Illustrate the distributive law, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, by means of the Venn diagram.

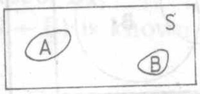
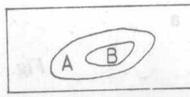
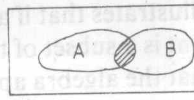
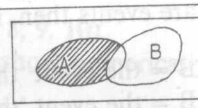
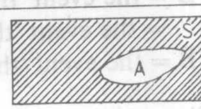
Description	Mathematical expression	Venn diagram
Mutually exclusive events A and B (no common elements)	A ∩ B = 0	
B is a subset of A (all elements of B are included in A)	B ⊂ A	
Union of A and B (all elements that are in either A or B)	A ∪ B	
Intersection of A and B (all elements in both A and B)	A ∩ B	
Difference between A and B (elements in A but not in B)	A − B	
Complementary set $\bar{A}$ (elements not in A)	$\bar{A}$ = S − A	

Fig. 1.4 Set operations

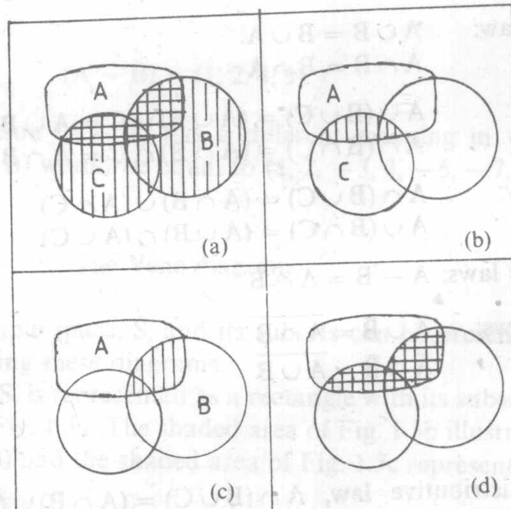


Fig. 1.5 Example 1.3

Solution

The three sets A, B and C are shown in Fig. 1.5.

In Fig. 1.5a the union of B and C, i.e. $(B \cup C)$ is shown with vertical lines whilst the intersection of $(B \cup C)$ with A, i.e. $A \cap (B \cup C)$, is marked with additional horizontal lines.

Figures 1.5b and 1.5c respectively show the intersections of A and C and A and B, which are both marked by vertical lines.

Figure 1.5d shows the union of $(A \cap B)$ and $(A \cap C)$, which is marked with both horizontal and vertical lines, and is seen to be identical to Fig. 1.5a, thus proving the theorem.

Note: The reader can prove the theorem numerically by assuming sets of values for each of the three sets A, B and C.

Example 1.4

By means of Venn diagrams prove the theorem $A - B = A \cap \bar{B}$.

Solution

When stated in words the theorem is, 'The elements contained in a set A, but not in a set B, are the same elements common to both set A and the complement of set B'. If set A and set B are as shown in Fig. 1.6a then the difference set, $A - B$, is represented by the hatched area shown.

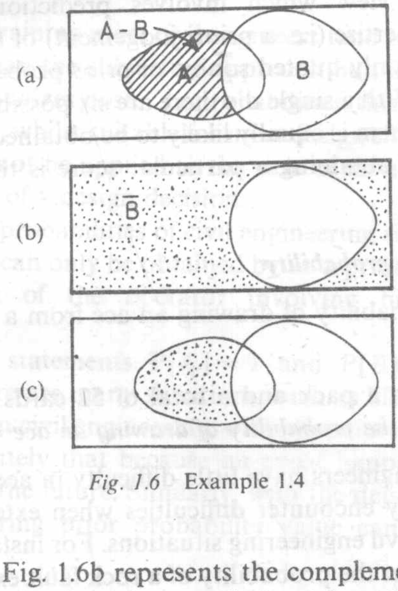


Fig. 1.6 Example 1.4

The dotted area of Fig. 1.6b represents the complement set, $\bar{B}$, and it is fairly obvious that the dotted area of Fig. 1.6c, which represents the elements common to A and $\bar{B}$, is the same as the hatched area of Fig. 1.6a. Hence:

$$A - B = A \cap \bar{B}$$

Note: The above axiom can be illustrated by considering the elements within the sets:

$$A - B = [x\,;\, x \in A \text{ and } x \notin B]$$

$$= [x\,;\, x \in A \text{ and } x \in \bar{B}] = A \cap \bar{B}$$

Probability

The probability that a particular event, A, will happen is expressed mathematically as P[A].

If the event A will never happen, e.g. pigs will fly, then the value of P[A] will be 0 whereas if event A will happen, e.g. the world will end sometime, then P[A] is 1.

Probability values are classified in one of two ways, depending upon how they are estimated, as follows.

PRIOR PROBABILITY VALUES

Prior probability values are obtained by the subjective, or degree of believe, point of view which involves predictions based on past experience and expertise (i.e. a priori judgement) of the decision maker.

The most commonly quoted source of prior probability values is the throwing of dice. With a single die there are six possible scores and, if the die is fair, each of them is equally likely to be obtained at any one throw. The probability of obtaining a particular score is therefore one in six, expressed as $\frac{1}{6}$.

Example 1.5: Prior probability

Determine the probability of drawing an ace from a full pack of cards.

Solution

There are 4 aces in a pack and a total of 52 cards. Hence $n = 4$ and $N = 52$. *Therefore, the probability of drawing an ace* $= n/N = \frac{4}{52} = \frac{1}{13}$.

Note: Most civil engineers have little difficulty in accepting this sort of reasoning but many encounter difficulties when extending the idea of degree of belief to civil engineering situations. For instance, few would be willing to accept that the probability of a rock fault existing at some site is 60%. Most would argue that, as the fault either exists or does not exist, then the probability is either 1 or 0.

It is in these situations that Bayes' theorem, described later in this chapter, can be of assistance in decision making.

POSTERIOR PROBABILITY VALUES

Posterior probabilities are estimated with hindsight, i.e. by the use of a frequenistic approach involving predictions based on a study of a series of repeatable events or tests.

Example 1.6: Posterior probability

Forty-five control tests were carried out on a long stretch of compacted subgrade. Five tests yielded results that were below specification.

If a further 10 tests had been carried out how many of these tests could have been expected to have given results below specification?

Solution

Probability of test results below specification is $\frac{5}{45} = \frac{1}{9}$.

For a further set of ten tests the expected number of results below specification is $\frac{10}{9} = 1.1$ i.e. one test.

Prior and posterior probabilities in civil engineering

Possibly because of their training most civil engineers tend to accept the frequenistic more easily than the subjective approach but, as most civil engineering design work involves posterior probabilities, this is generally no disadvantage. However, whilst suitable for most design situations, the frequenistic approach cannot be applied to the case of an unrepeatable event, such as the making of a design decision.

The estimation of prior probabilities of civil engineering situations, as opposed to dice throwing, can only be obtained by an a priori approach, i.e. subjective judgement of the operator involving his previous experience.

On the face of it the statements P[A] = 1 and P[B] = 0 imply absolute certainty and there are many situations, such as life and death, when this is so. However, in civil engineering, with posteriori judgement, one cannot assume absolutely that because an event happened in the past it will do so again in the future. Similarly, with the degree of belief approach, a civil engineering prior probability value can hardly be regarded as certain.

Generally speaking, when the statement $P[A] = 1$ occurs in this text it means that it is considered that A will most probably occur, not that it will occur.

Mutually exclusive events

If there is a set of events A, B, C, ... such that the happening of one excludes the happening of the others then we say that the events A, B, C, ... are mutually exclusive.

An example of mutual exclusion would be the acceptance of a tender from among several submitted. If contractor A is successful in his bid then there is no possibility of contractors B, C, etc. also being successful.

The summation law – union probability

This law applies to mutually exclusive events and states that for a series of mutually exclusive events, the union probability of at least one of these events occurring is equal to the sum of the separate probabilities of the events.

Consider three events, A, B and C. The probability that any one of these events will occur is:

$$P[A \cup B \cup C] = P[A] + P[B] + P[C]$$

(It may help readers if they consider the union symbol, $\cup$, to represent the word, 'or'.)

Example 1.7

Examples of the summation law are:

(i) The tossing of a fair coin:
The probability of a head = $P[A] = 0.5$ or 50%
The probability of a tail = $P[B] = 0.5$ or 50%
Probability of either a head or a tail = $P[A \cup B] = P[A] + P[B] = 1.0$ or 100%.

(ii) A set of strength measurements of a particular material:
$P[A]$ = the probability of the actual strength being equal to or less than the mean value = 0.5.
$P[B]$ = the probability of the actual strength being equal to or greater than the mean value = 0.5.
$P[A \cup B] = P[A] + P[B]$ = the probability that the actual strength