Comparative Statistical Inference

Third Edition

Vic Barnett

Professor of Environmental Statistics
University of Nottingham, UK

JOHN WILEY & SONS, LTD
Chichester • New York • Weinheim • Brisbane • Singapore • Toronto

This Page intentionally left blank

Comparative Statistical Inference

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

A complete list of the titles in this series appears at the end of this volume

Comparative Statistical Inference

Third Edition

Vic Barnett

Professor of Environmental Statistics
University of Nottingham, UK

JOHN WILEY & SONS, LTD
Chichester • New York • Weinheim • Brisbane • Singapore • Toronto

Baffins Lane, Chichester,
West Sussex, PO19 IUD, England

National 01243 779777
International (+44) 1243 779777
e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on http://www.wiley.co.uk
or http://www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Wiley-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 129809

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

Library in Congress Cataloging-in-Publication Data

Barnett, Vic.
Comparative statistical inference / Vic Barnett. — 3rd ed.
p. cm.—(Wiley series in probability and statistics)
Includes bibliographical references and index.
ISBN 0-471-97643-1 (cased : alk. paper)
1. Mathematical statistics. 2. Statistical decision. I. Title
II. Series.
QA276.B2848 1999
519.5′—dc21 98-44623
CIP

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-471-97643-1

Typeset in 10/12pt Times by Laser Words, Madras, India

To Audrey, Kate and Emma

This Page intentionally left blank

Contents

Preface

This century has seen a rapid development of a great variety of different approaches to statistical inference and decision-making. These may be divided broadly into three categories: the estimation and hypothesis testing theory of Fisher, Neyman, Pearson and others; Bayesian inferential procedures; and the decision theory approach originated by Wald.

Each approach is well documented, individually. Textbooks are available at various levels of mathematical sophistication or practical application, concerned in the main with one particular approach, but with often only a passing regard for the basic philosophical or conceptual aspects of that approach. From the mathematical and methodological viewpoint the different approaches are comprehensively and ably described. The vast amount of material in the professional journals augments this and also presents a detailed discussion of fundamental attitudes to the subject. But inevitably this discussion is expressed in a sophisticated form with few concessions to the uninitiated, is directed towards a professional audience aware of the basic ideas and acquainted with the relevant terminology, and again is often oriented to one particular approach. As such, the professional literature cannot (nor is it intended to) meet the needs of the student or practising statistician who may wish to study, at a fairly elementary level, the basic conceptual and interpretative distinctions between the different approaches, how they interrelate, what assumptions they are based on, and the practical implications of such distinctions. There appears to be no elementary treatment which surveys and contrasts the different approaches to statistical inference from this conceptual or philosophical viewpoint. This book on comparative statistical inference has been written in an attempt to fill this gap.

The aim of the book is modest; by providing a general cross-sectional view of the subject, it attempts to dispel some of the 'air of mystery' that must appear to the inexperienced statistician to surround the study of basic concepts in inference. In recognizing the inevitable arbitrary and personal elements that must be reflected in any attempt to construct a 'rational' theory for the way individuals react, or should react, in the face of uncertainty, he may be better able to understand why factional groupings have developed, why their members attach 'labels' to themselves and others, and why discussion so easily reaches a somewhat 'emotional' level. By stressing the interrelationships as well as the conceptual conflicts it is hoped that the different approaches may be viewed as a composite theory of inference, the different methods having separate relevance in different situations, depending on local circumstances. The book achieves its object substantially if it does no more than persuade the reader that he is not required to 'stand up and be counted' with those who have committed themselves to one particular attitude towards inference to the exclusion of all others.

The idea of the book originated from my experience, over several years, of running a lecture course on comparative statistical inference. The course was attended by final-year undergraduate, and first-year postgraduate, students in Mathematical Statistics at the University of Birmingham; it was introduced to augment their knowledge of the mathematics and techniques of different approaches to statistical inference and decision theory by presenting them with the spectrum of philosophical attitudes inherent in a comparison of the different approaches. Other universities offer similar courses and this book should provide useful collateral reading for such courses, as well as being a general treatment of comparative statistical inference for a wider audience.

This book is not intended as a comprehensive discussion of the mathematics or methodology of any particular approach, nor as an authoritative history of the development of statistical consciousness. Some historical comment is included to add interest, and the mathematics and methodology are developed to the stage where cogent comparison is possible. This comment and development, however, remains subservient to the prime objective of a comparison of different philosophical and conceptual attitudes in statistical inference.

No detailed mathematical proofs are given, and the treatment assumes only a knowledge of elementary calculus and algebra (perhaps to first-year university level) and an acquaintance with the elements of the theory of probability and random variables. Some familiarity with specific methods of inference and decision theory would be an advantage.

The first two chapters of the book are introductory. Preliminary ideas of inference and decision-making are presented in Chapter 1, and applied in Chapter 2 to the informal construction of various inferential techniques in the context of a practical example. Chapter 3 traces the range of different definitions and interpretations of the probability concept that underlie the different approaches to statistical inference and decision-making; Chapter 4 outlines utility theory and its implications for general decision-making. In Chapters 5 to 7 specific approaches are introduced and developed with a general distinction drawn between *classical* inference on the Neyman—Pearson approach, *Bayesian* methods and *Decision Theory*. Particular attention is given to the nature and importance of the basic concepts (probability, utility, likelihood, sufficiency, conjugacy, admissibility, etc.) both within and between the different approaches. The final chapters (8 and 9) present a sketch of some alternative attitudes, and some brief concluding remarks, respectively.

A subject and author index and bibliography are included, and textual references to *books* relate, by author's name and date of publication, to the bibliography. References to *papers* in the professional journals bear the author's name and an index number, e.g. Savage[7], and relate to the list of references at the end of the *current* chapter. It is hoped that readers may be stimulated to delve deeper into the various topics that are presented. To assist them to do so, particularly when the book is used as collateral reading for a lecture course on comparative inference, certain references in the list at the end of each chapter are marked with a dagger, e.g. †3. Cox, D. R.... This indicates material that seems particularly suitable as the basis for extended study or discussion of the subject matter of the current chapter. The marked references have been chosen, in the main, on the basis of providing a broad, non-detailed, extension or reappraisal of relevant material—often surveying, interrelating or comparing the different approaches. The dagger (†) is in no sense intended as a mark of distinction in terms of the merits of different authors' contributions.

It is necessary to explain one or two particular points of notation at the outset. Frequently we will be concerned with data arising from an assumed parametric model. The data are denoted x, the parameter, θ. Usually no indication is given (or is needed within the general treatment) of dimensionality. Thus, x may be a single observation of a univariate random variable, or of a multivariate random variable, or may be a sample of independent (univariate or multivariate) observations. Similarly, θ may have one, or many, components. In the same spirit, X denotes the general random variable of which x is a realisation. The sample space and parameter space are denoted by $\mathscr{X}$ and Ω, respectively. The probability mechanism governing the occurrence of x is represented by the function $p_\theta(x)$, with a correspondingly broad interpretation as a probability or probability density, or where the emphasis demands it as a likelihood. To avoid the unnecessary complication of distinguishing between discrete and continuous variables, and to maintain the presence of $p_\theta(x)$ as a central component in mathematical expressions, an individual style is adopted to denote integration or summation. The expressions

$$\int_{\mathscr{X}} h(x, \theta) \text{ or } \int_{\Omega} h(x, \theta)$$

are used to represent the appropriate single, or multiple, integrals or sums of some function $h(x, \theta)$ over the range of variation of x or θ, respectively. The subscript $\mathscr{X}$, or Ω, will also be attached to the expectation operator $\mathrm{E}(\cdot)$, to indicate the space over which the expectation is calculated. For example,

$$\mathrm{E}_{\mathscr{X}}[g(X)] = \int_{\mathscr{X}} g(x) p_\theta(x)$$

is the expected value of the function $g(X)$ of the random variable, X. (The subscripts $\mathscr{X}$, or Ω, will be omitted if the appropriate space is obvious in the context of the discussion.)

The use of the usual integral sign for this purpose may offend the purist. However, it seems more appropriate to take such a liberty for the sake of the intuitive simplicity of the notation, than to introduce some new symbol and require the reader constantly to remind himself of its meaning.

In particular examples where it is important to distinguish the structure of x or θ the more conventional notation for integrals, sums, density functions, and so on, will be explained and used.

It is a pleasure to acknowledge my indebtedness to all those who have contributed, in various ways, to the production of this book. Its existence owes much to the example and stimulus of colleagues. My debt to the vast literature on the subject is self-evident, from the extent to which I have referred to other writers who have so often expressed ideas far better than I can. I am grateful also to my students, whose comments and enquiries have directed my thoughts to what I feel is a reasonable mode of presentation of the material. My thanks are especially due to a few friends who have given their time to read and comment on sections of the book; in particular David Kendall, Toby Lewis, Dennis Lindley and Robin Plackett. Toby Lewis has been a constant source of help and encouragement; I am very grateful to him. Every effort has been made to ensure that factual details are correct and that historical and interpretative attribution is fair and just; also to avoid any implicit bias towards a particular approach to the subject. It is not easy, however, to assess these matters objectively, and any errors, omissions or bias are unintentional and my responsibility alone.

October, 1972

Vic Barnett

Preface to Second Edition

Much has happened in the field of inference and decision-making over the decade since the first edition of this book was published. The preparation of a second edition presents a valuable opportunity to provide more detailed treatment of some topics, to offer some discussion of new emphases, techniques and even whole approaches to inference, and to reflect changes of basic attitude to the subject.

In classical inference specific attention is given to multi-parameter problems, and to notions of ancillarity and conditional inference. The revitalisation of the distinction between hypothesis tests and 'pure significance tests' is discussed and interpreted. The treatment of the role of likelihood is broadened to encompass comment on modified forms of likelihood (marginal, conditional, etc.), and to expand on the significance of the likelihood principle in the various approaches (particularly its relationship to the concept of coherency in Bayesian inference). Greater attention is given to practical ways of representing and assessing subjective probabilities and utilities, and to work on the application of Bayesian methods. The method of Bayesian prediction is outlined. Two new approaches are briefly described: pivotal inference and plausibility inference.

The above topics represent some of the additions in this second edition. The book has been thoroughly revised throughout to reflect changes of substance and attitude in the intervening period. In particular the reference lists at the end of each chapter, and the bibliography, are much more extensive and contain relevant contributions to the literature up to the time of the revision.

It must be stressed, however, that the overall aim of the book is unchanged. It aims to present and develop the various principles and methods of inference and decision-making to such a level that the reader can appreciate the basic characteristics, interrelationships and distinctions of the different approaches. Accordingly, detailed mathematical development or proof and comprehensive coverage of applications are eschewed (in text, and in references), in order not to cloud the objective of presenting in managable proportions a basic understanding of essential principle, philosophy, method, interpretation and interrelationship.

Sheffield, May 1981

Vic Barnett

This Page intentionally left blank

Preface to Third Edition

This third edition of *Comparative Statistical Inference* incorporates a range of new emphases and topics that are having a major influence on inference and decision-making, as well as an expanded treatment of the material of earlier editions. It reflects the changing relative appeal of certain techniques and principles, advances in methodology and the ever-increasing benefits that flow from the power of modern computers and computing methods.

Some of the changing emphases, advances and new topics described in this third edition relate to causal inference, chaos theory, developments of modified likelihood forms, fuzzy sets, the generalised linear model, the Gibbs sampler and Markov chain Monte Carlo methods, meta-analysis and combining information, prediction, prequential inference, sample reuse procedures, and so on.

A revised edition also offers the opportunity to improve the communication of ideas through modified descriptions of some of the material, with alternative styles of presentation and with alternative and additional examples. A specific difference relates to the reference material, which has, of course, been fully up-dated to cover the latest developments and is much expanded in coverage and level. Additionally, all references, whether in the form of books or papers in journals are now described in the text in standard form—e.g. as Barnett (1996)—and gathered together at the rear of the book in a single consolidated set of reference and bibliographic material. This is more accessible than the previous mix of chapter-end lists and end-of-book bibliography, particularly when the coverage has been so substantially increased in the amount and levels of treatment with many references having relevance at various points throughout the book.

In spite of the many changes in this new edition, the essential aim remains the same: to explain and compare the many approaches to inference and decision-making in their various forms and emphases in sufficient detail for an understanding of the historical development of the theme and of the latest advances in the expression of concepts, principles and methods.

Nottingham, June 1998

Vic Barnett

This Page intentionally left blank

Acknowledgements

The author gratefully acknowledges the assistance of the following publishers who have granted permission for the reproduction of extracted material quoted in the text, sources being given in the reference and bibliographic section at the end of the book:

Academic Press, Inc., London.
George Allen & Unwin Ltd., London.
The Annals of Eugenics (Cambridge University Press, London).
Biometrika Trust, London.
Cambridge University Press, London.
Chelsea Publishing Co. Inc., New York.
The Clarendon Press, Oxford.
Gauthier-Villars, Paris.
Charles Griffin & Co. Ltd., London and High Wycombe.
Hafner Publishing Company, New York.
Holt, Rinehart and Winston of Canada Ltd., Toronto.
Institute of Actuaries, London.
The Institute of Mathematical Statistics, California (*The Annals of Mathematical Statistics*).
International Statistical Institute, The Hague.
The Macmillan Company, New York.
Massachusetts Institute of Technology, Cambridge, Mass.
Methuen & Co. Ltd., London.
The Regents of the University of California.
The Royal Society, London.
Royal Statistical Society, London.
Statistica Neerlandica, Utrecht.
The University of Adelaide.
University of Western Australia.

This Page intentionally left blank

CHAPTER 1

Introduction: Statistical Inference and Decision-making

1.1 WHAT IS STATISTICS?

It is usual for a book to commence by defining and illustrating its subject matter. Books on statistics are no exception to this rule, and we frequently find two particular characteristics in such preliminary definitions. The definitions are often brief and superficial with the aim of merely 'setting the ball rolling', relying largely on the reader's intuitive understanding and motivation. They also tend to be directed implicitly towards the particular level and emphasis of the treatment being presented. It is interesting to consider some examples.

The early use of the word 'statistics' was to describe 'tabulated numerical facts relating to the state'. Much later the term began to be employed as a label for a particular scientific discipline. This distinction is still apparent in introductory remarks in books at various levels. The stress is on the former or latter aspect depending on whether the book is more concerned with the collection and presentation of data ('descriptive statistics') or with statistical methods for analysis and interpretation of the data.

For example, Mason, Lind and Marchal (1983) define *descriptive statistics* as:

> Methods used to describe the data that have been collected (p. 3)

distinguishing this from *statistics*:

> The body of techniques used to facilitate the collection, organisation, presentation, analysis, and interpretation of data for the purpose of making better decisions. (p. 3)

Apart from the reference to 'decisions', Neter, Wasserman and Whitmore (1978) in an introductory book on 'applied statistics' adopt a similar stance:

> . . . statistics refers to the methodology for the collection, presentation, and analysis of data, and for the uses of such data. (p. 1).

A further refinement appears in the definition of *Inferential statistics* by Scheaffer and McClave (1982) as:

> ... use of data to make intelligent, rigorous, statements (inferences) about a much larger phenomenon from which the data were selected. (p. 1)

Signalling the distinction between the sample and population, and chiming with the early definition of statistics by Egon Pearson (Bartholomew, 1995);

> ... the study of the collective characters of populations.

When interest focuses on the formal (mathematical) derivation and detail of the statistical methods, definitions become more specific on the nature of the data to be analysed and on the presence of a chance mechanism operating in the situations that yield the data. Stuart and Ord (1994) remark:

> Statistics is the branch of scientific method that deals with the data obtained by counting or measuring the properties of populations of natural phenomena. In this definition 'natural phenomena' includes all the happenings of the external world, whether human or not. (p. 2)

Cox and Hinkley (1974) introduce at the outset the idea of indeterminateness in saying:

> Statistical methods of analysis are intended to aid the interpretation of data that are subject to appreciable haphazard variability. (p. 1)

as does Stevens (1968) in defining statistics as:

> ... a straightforward discipline designed to amplify the power of common sense in the discernment of order amid complexity. (p. 854)

Fraser (1976) is more formal:

> *Statistical theory* ... builds on the use of probability to describe variation .... (p. 2)

Most of the above definitions are general enough to place little constraint on the subsequent development of the statistical methods or theory being presented. However, many examples may be found where the preliminary definition of the subject matter reflects a particular philosophical or conceptual emphasis in the later material. They may imply a concentration on a single interpretation of the probability concept, or a particular attitude to what constitutes relevant information for statistical study and to how it should be processed. The *frequency* concept of probability is explicit in the comments by Hoel (1971):

> ... statistics is the study of how to deal with data by means of probability models. It grew out of methods for treating data that were obtained by some repetitive operation .... The statistician looks on probability as an idealisation of the proportion of times that a certain result will occur in repeated trials of an experiment .... (p. 1)

The notion of 'information' is made explicit in Miller and Miller (1994):

> The object of statistics is information. The objective of statistics is the understanding of information contained in data.

Decision-making features in some definitions as we have noted above. Sometimes it is informally described and not clearly differentiated from inference:

> ... statistics has been concerned with drawing judgments from observations. (Hinkelmann and Kempthorne, 1994, p. 27)

Chernoff and Moses (1959), in an introductory text on *decision theory*, are dissatisfied with a definition that places emphasis on 'data handling':

> Years ago a statistician might have claimed that statistics deals with the processing of data. ... to-day's statistician will be more likely to say that statistics is concerned with decision making in the face of uncertainty. (p. 1)

In an elementary treatment of *Bayesian statistical inference*, Lindley (1965b) sees statistics as the study of 'how degrees of belief are altered by data'. Savage et al. (1962) stresses a personalistic function of the subject:

> By [statistical] inference I mean how we find things out—whether with a view to using the new knowledge as a basis for explicit action or not—and how it comes to pass that we often acquire practically identical opinions in the light of evidence. (p. 11)

These few illustrations are not intended to be comprehensive or even representative. They merely serve to demonstrate possible purposes behind an introductory definition of statistics. Definitions of the general type, often deliberately cursory or incomplete, serve ably to motivate a mathematical or methodological treatment of the subject at any level and from (almost) any viewpoint. In cases where the definitions are more specific, more personal, they provide an indication of the emphasis and viewpoint adopted in the subsequent development. For our present needs they underline a feature of the study of statistics that is basic to the purpose of this book: that there are a variety of aspects of the subject in which there is room for discussion and individual viewpoints. Different attitudes to

(i) what is meant by *probability*,

(ii) what constitutes *relevant information* for the application of statistical methods,

(iii) whether or not any limitations need to be placed on the areas of human activity amenable to statistical analysis, and so on,

will all inevitably colour and influence the approach to statistics. The object of this book is to examine the fundamental nature of statistical theory and practice by a comparative study of the different philosophical, conceptual and attitudinal (sometimes personal) 'approaches' to the subject. To achieve this it will be necessary to consider basic concepts in detail, and to develop the mathematics and methods associated with the different approaches to the stage where detailed comparison is possible.

For these needs, however, we cannot be content with either a superficial or an idiosyncratic definition of statistics. To commence with an 'emotionally charged' definition is to defeat at the outset the aim of presenting a fair cross-section of views and attitudes. In a sense no definition is needed, since the book as a whole may be regarded as an attempt to provide such a definition. But we must start somewhere, and the best course is to construct a preliminary description of the purpose and function of statistical study that is, on the one hand, general enough to accommodate the widely differing philosophical and conceptual views that exist, and, at the same time, specific enough to mark out the basic components of *any* theory of statistics.

This is no easy task; in attempting to be 'impartially specific' there is the risk that we end up by saying nothing of value. But better to say too little at this stage than too much. The gaps can be filled in as the book progresses.

With this attitude we provisionally define *statistics* as **the study of how information should be employed to reflect on, and give guidance for action in, a practical situation involving uncertainty.**

This definition contains a variety of ingredients that require fuller description. What is meant by 'a practical situation involving uncertainty'? What constitutes 'information'? What is the implied distinction between the 'reflection' and 'action guidance' function of statistics? The following sections consider these points in more detail.

1.2 PROBABILITY MODELS

In amplifying this definition the natural starting point is to consider what is meant by 'a practical situation involving uncertainty'.

We have in mind that circumstances exist, or have been contrived through experimentation, in which different possible outcomes may arise that are of concern to us as the observer, or experimenter. The essential feature of the situation is that *there is more than one possible outcome* and that *the actual outcome is unknown to us in advance*: it is *indeterminate*. Our interest may be in knowing what that outcome will be, or in deciding on a course of action relevant to, and affected by, that outcome. A doctor prescribes a drug for a patient—will it be successful in curing his patient? Should we decide to spend tomorrow at the beach—when the weather may or may not be fine, and the state of the weather will seriously affect our enjoyment of the exercise?

Any attempt to construct a theory to guide behaviour in such 'situations involving uncertainty' must depend on the construction of a formal (logical or mathematical) **model** of such situations. This requires the formulation of a concept of **probability**, and associated ideas of independence, randomness, etc. as a mechanism for distinguishing between the different outcomes in terms of their degree of uncertainty. We shall see later that, in response to the nature of the situation and depending on individual attitudes, *a variety of philosophical interpretations of the probability concept can exist.* These interpretations may colour the theory of statistics developed to deal with the situation, and it is therefore useful at this stage to indicate some of the broad distinctions that exist.

A hint of the dilemma is provided in the example of a doctor prescribing a drug. A simple model for this situation is one that specifies two possible outcomes, that the patient is cured or is not cured, and assigns a probability, p, to a cure, $1-p$ to no-cure. But how are we to interpret the probability that the patient is cured by the drug? We might adopt the attitude that the patient is 'typical' or 'representative' of a larger population of patients with a similar physical condition who have been, or will be, treated with this drug. This leads to a *frequency-based view of probability*, in which p is related to the proportion of cures, or potential cures, in the larger population.

Alternatively, we may prefer to regard the patient as an individual, whose personal physiological and psychological make-up will determine the success or failure of the drug in *his* or *her* case: we cannot now usefully regard the patient as a representative member of some larger population. After all, even if we know that 80 per cent of the larger population are cured we still do not know whether or not this particular patient will be. If we reject representativeness in a larger population, the frequency concept of probability no longer has relevance, and some other interpretation is needed. One possibility is now to regard the probability of a cure as a measure of the doctor's (and patient's) '*degree-of-belief*' in the success of the treatment. Here we are likely to need a *subjective* interpretation of the probability concept. In practice the doctor's decision to prescribe the drug is likely to stem from an informal combination of *frequency* and *subjective* considerations, aimed at an assessment, or 'estimate', of the value of p as a guide for action. We shall return to this question of alternative interpretations of probability, and their implications for the construction of statistical theories, in Chapter 3, where the contrasting ideas of '*classical*', *frequency, logical* and *subjective* or *personal* concepts of probability are discussed and illustrated.

The model of the practical situation consists, essentially, of *a statement of the set of possible outcomes and specification of the probabilistic mechanism governing the pattern of outcomes that might arise.* Inevitably, the model is an idealisation of the real situation. Its adequacy depends on how valid and appropriate are the (often simple) assumptions on which it is based. The fundamental concern of the statistician is to construct an *adequate* model, either as a description of the real situation (and there the interest rests) or to suggest a reasonable course of action relevant to that situation. The real-world situation becomes replaced by the model and any description, or action, relates to the model as a substitute for the situation. Some simple examples may clarify this:

(i) A radioactive substance emits α-particles. The substance is characterised by the rate at which it emits these particles. A Geiger counter is placed nearby and records the α-particles that bombard it. The usual assumptions of randomness and independence of occurrence of the α-particles lead to a probability model for this situation—in which the number of α-particles recorded in a fixed time interval, $(0, T)$, has a particular probability distribution; namely, the Poisson distribution. This model is fully specified by a single *parameter*, its mean, which is proportional to the rate of emission, λ, of the α-particles.

Thus, a fairly complex physical situation has been replaced by a simple probability model, in which the extent of our uncertainty is reduced to a single unknown quantity related directly to our prime interest in the physical situation itself. Methods of probability theory make it possible to deduce the pattern of results we would expect to observe *if a fully specified form of the model is appropriate*. Thus we can calculate, say, the probability of recording no α-particles in a five-second period. *In reverse*, by comparing the actual recordings on the Geiger counter with the deductions from the probability model we can attempt to both *validate the model* and *estimate the unknown parameter* (the mean).

(ii) In an agricultural experiment conducted by a seed supplier, four varieties of Winter wheat are being compared; for example, in terms of their relative yields (or resistance to disease). Large variations of yield will arise due to inhomogeneity, of soil characteristics, the geographical aspect of the plot and the cumulative random effects of many other factors (including non-constancy of care and measurement errors). Several plots are planted with each variety and efforts are made to reduce the systematic effects of soil, aspect, etc. and to encourage optimal growth for each variety. The yields are measured as an indication of the relative merits of the four different varieties with a view to marketing a new improved variety for use by farmers.

Any statistical analysis of the data again rests on the construction of an appropriate model and on an assumption of its adequacy. Here, the assumptions may be that the observed values of yield (etc.), possibly after a suitable transformation, arise independently from normal distributions with constant variance, differing from one variety to another in at most the values of their means. Thus, the model embodies a great deal of structure, being unspecified only to the extent of the unknown values of a few parameters. Again, the real situation has been replaced by the model, and the ideas of probability theory may be applied to deduce the characteristic behaviour of data arising from the model, and hence by *assumption* from the real situation. The use of a statistical procedure, e.g. analysis of variance, again attempts to *reverse* this process, by using the observed data to reflect back on the model (both for validation and estimation purposes).

What is to be learnt from these examples? First, the relationship between the practical situation, its probability model and the information it generates. Secondly, the roles played by a formal *theory of probability* and by *statistical procedures and methods*, in linking these components.

It is the practical situation that is of prime interest, but this is both inaccessible and intangible. An idealisation of it is provided by the *probability model*, with the hope that it constitutes a valid representation. Being logically and mathematically expressed, this model is amenable to formal development. Logical deduction through the ideas of mathematical probability theory leads to a description of the probabilistic properties of data that might arise from the real situation—*on the assumption that the model is appropriate*. In this way probability acts as the communication channel or 'language' that links the situation (model) with the data: it provides a *deductive* link.

The *theory of statistics* is designed to reverse this process. It takes the real data *that have arisen* from the practical situation (perhaps fortuitously or through a designed experiment) and uses the data to validate a specific model, to make 'rational guesses' or 'estimates' of the numerical values of relevant parameters, or even to suggest a model in the first place. This reverse, *inductive*, process is possible only because the 'language' of probability theory is available to form the *deductive* link. Its aim is to *enable inferences to be drawn* about the model from the information provided by the sample data (or perhaps by other information) or to construct procedures to aid *the taking of decisions* relevant to the practical situation. What constitutes 'relevant' information, and the implied differences in the *descriptive* and *decision-making* functions of statistical theory, will be taken up in the next two sections.

The different components (*practical situation, model, information*) and the links between them (*deductive* or *inductive*) are represented diagrammatically in Figure 1.2.1.

1.3 RELEVANT INFORMATION

Returning to the definition given in Section 1.1, a second component that needs fuller discussion is the 'information' that is to be 'employed'.

In examples (i) and (ii) of the previous section, information took the specific form of 'realisations' of the practical situation: that is, observed outcomes from that situation arising from what are assumed to be independent repetitions of the situation under identical or similar circumstances. This type of information will be termed **sample data**. Some writers would claim that it is only for information of this type, obtained in a repetitive situation (potentially at least, if not practically) that a concept of probability can be adequately defined, and a statistical theory developed. In von Mises' formalisation of the frequency concept of probability

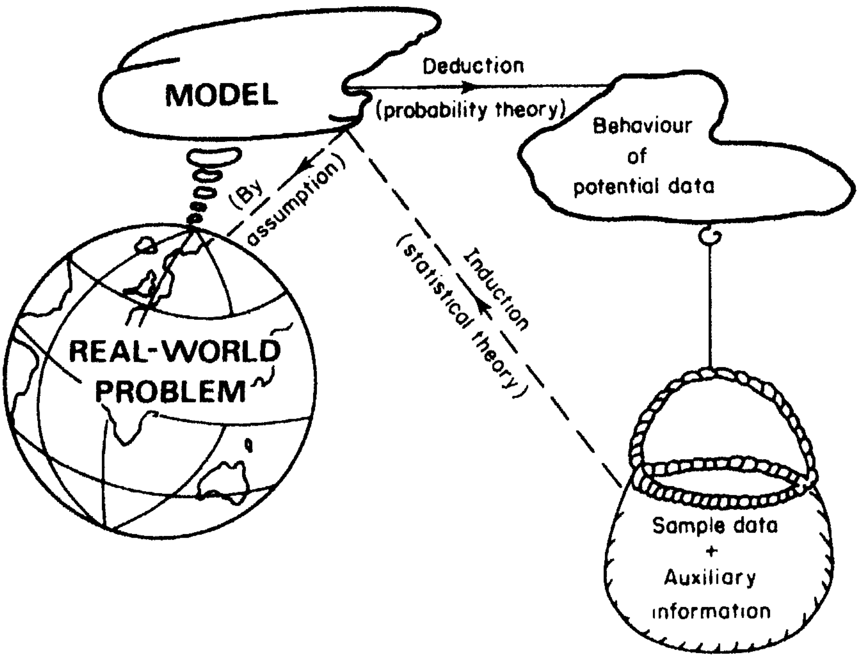

Figure 1.2.1

(1957; first published 1928, in German) and later in its application to statistics (1964) he says (quoting from his 1964 book, published posthumously):

> Probability ... theory is the mathematical theory of ... repetitive events. Certain classes of probability problems which deal with the analysis and interpretation of statistical enquiries are customarily designated as 'theory of statistics'...
>
> ... we limit our scope, roughly speaking, to *a mathematical theory of repetitive events*. (p. 1)

and later

> ... if one talks of the probability that the two poems known as the *Iliad* and the *Odyssey* have the same author, no reference to a prolonged sequence of cases is possible and it hardly makes sense to assign a *numerical* value to such a conjecture. (pp. 13, 14)

Such restriction of interest to repetitive situations was for the purpose of constructing a theory of probability. But since probability is the 'language' of statistical analysis, there is clearly the implication that only sample data, derived from repetitive situations, may be the subject of such analysis. If the 'language' contains no word for other forms of information how can we 'talk about' these other forms?

A similar attitude is expressed by Bartlett (1962) who, in an essay on the theory of statistics, says that statistics

> *is concerned with things we can count.* In so far as things, persons, are unique or ill defined, statistics are meaningless and statisticians silenced; in so far as things are similar and definite—so many male workers over 25, so many nuts and bolts made during December—they can be counted and new statistical facts are born. (p. 11)

Again, the emphasis is on the repetitive nature of the situation, and information is restricted to sample data.

Bartlett and von Mises are not alone in this attitude that statistical analysis must be based on the use of sample data evaluated through a *frequency* concept of probability. It is a view widely held and expressed, and acts as the corner-stone of a whole approach to statistics—what we shall call the **classical approach** stemming from the work of Fisher, Neyman, E. S. Pearson and others.

But is sample data really the only form of information relevant to a statistical study? Other information certainly exists. The engineer who is considering using a component offered by a particular supplier, for assembly in a new design of bridge can test samples of this component and obtain data to aid his assessment of its suitability. But he may also have information on the reliability of the supplier in terms of a *past record of that supplier's ability to provide reliable components* in similar circumstances. He may also be able to work out the possible *costs and consequences* of using a component that subsequently proves not to meet the required specifications. Both the *earlier experience* and the *potential consequences* are *relevant* to the decision on whether or not to use the component offered.

Other approaches to statistics are designed to incorporate such alternative types of information, and any attempt to survey the various approaches to the subject must therefore rest on a *wider concept of information than that of just sample data.* The broad subdivision of information into *three* categories (earlier experience, sample data, potential consequences), suggested by this example, is a useful one to pursue at this stage. Suppose we consider examples (i) and (ii) of the previous section in more detail.

The agricultural experimenter in example (ii) is unlikely to regard his problem as a mathematical one concerning the means of a set of normal distributions. He is presumably interested in using any knowledge gained about the relative merits of the different wheat varieties as a guide to future action. It may be that he hopes to be able to recommend one of the varieties for commercial production. The mere superiority of one variety in terms of its yield is unlikely to be sufficient justification for its adoption. To increase wheat production is an advantage, but it may be more expensive to grow the highest yielding variety and hence needs to be priced more highly, or it may involve much greater care and attention (in soil treatments and husbandry). Higher prices and greater effort can be disadvantageous in terms of profitability. The decision as to what variety to use is thus far more complex than merely a choice of that variety which produces the highest yield.

The application of statistical methods in this problem illustrates how the object of the study may need to be extended beyond describing (through an appropriate probability model) how different outcomes arise, to the wider aim of constructing a policy for action. Here we have an example of a situation where an assessment of the **consequences** of alternative actions is vital. These consequences must be quantified in some appropriate manner and the information they then provide may be critical to the choice of action. This *second type of information* is thus highly relevant, augments the information provided by the sample data, and demands the construction of statistical techniques through which it may be 'employed'.

Assessment of consequences, and their formal quantification through what is known as the concept of **utility**, must therefore form a part of any comparative study of different approaches to statistics. It is central to a particular approach known as **decision theory**. It is important, however, to recognise that information on consequences may be different in kind to that provided by sample data. Sample data are well defined, objective. Consequences may also lead to objective quantification (the costs of soil treatment if a particular variety of wheat is grown) but they need not do so. Farmers may react against the extra labour needed to care for a particular variety, over and above the actual cost of such labour. In the same way as situations may demand a subjective interpretation of probability, so an assessment of consequences may involve *subjective* (personal) value judgements.

In the sphere of human activity in particular it is often difficult to be objective. Bill cannot decide whether or not to accept an offer of a new job. It is clear that many items of information will be relevant to reaching a decision and they will involve imponderables and uncertainties. A change of salary may be obvious, but what about relative promotion prospects, the reactions to a change of work and home location, sale and purchase of a house, etc. Obviously, formal decision theory would not really be appropriate here—but the example is not so exaggerated! It is intriguing to ponder what information he might seek to assist in this decision; also it is obvious that his assessment of the consequences will be very personal and difficult to quantify. Much of the emphasis in utility theory and decision theory is on the construction of a rational model for human behaviour, in the sense of representing how individuals make (or should make) a choice from alternative possible actions in the face of uncertainty.

Even in the study of apparently objective situations we cannot escape the personal element. We saw this in the example of the seedsman. Often, a subjective assessment of consequences is forced on us by the sparseness of objective information. The seedsman knows that to market a higher yielding, but more expensive, wheat (in terms of seed cost and management) will exclude a proportion of the market (it may even be known that this currently accounts for 'about 20 per cent' of present sales). But perhaps it is also likely that certain growers who need to maximise their yields on limited land areas may in future be attracted to a newly introduced high-yielding variety (they might assess this as a 50 per cent chance of 40 per cent extra sales). The seedsman possesses, then, some measure of objective information, but to incorporate this factor in a statistical analysis it is necessary to *fully* specify the information in numerical terms and this can only be done by incorporating subjective or arbitrary values where knowledge is incomplete. Alternatively, subjective and personal elements *must* arise when the context of the problem is 'personal' rather than 'global'—(Bill and his new job, for example). Some writers would argue that this is right and proper: that the individual is making entirely personal decisions, and that it is therefore the personal (subjective) assessment of consequences that constitutes relevant information. The difficulty of quantification still exists, however!

We find a third type of information potentially arising in the earlier example (i) concerning radioactive decomposition of a substance. The aim here may be merely to characterise the substance in terms of its rate of decay, λ, which is the mean number of α-particles, per second, recorded by the Geiger counter. This can be achieved by representing the situation in terms of the proposed Poisson model and by using the sample data alone to yield an estimate of λ as the only unknown parameter. The reliability or accuracy of the estimate will depend on the extent of the data and the method used to process the data. We can define ways of measuring this accuracy.

Suppose, however, that on chemical or physical grounds we know that the substance *has affinities with other substances with known rates of decay*. Its own rate of decay should not be too dissimilar to these others!. As a result, we may 'feel quite confident' that λ is in the range $0.45 < \lambda < 0.55$, say. This knowledge is *a further form of relevant information* that we would hope to combine with the sample data to conduct a more refined estimation of λ. Such information is derived from *outside* the current situation. It may arise, as in this example, from the general accumulated knowledge of the investigator from other areas of experience: quite often from previous observation of similar situations. Information of this type is termed information a priori, or **prior information**. The particular branch of statistics designed to combine *prior information* with *sample data* is known as **Bayesian statistical inference**, but we shall see that prior information can also play an important role in *decision theory* and has implications in the *classical* approach to statistics.

In trying to incorporate prior information in a statistical analysis we again encounter the difficulty of quantifying it. It needs to be expressed in terms of *prior probability distributions*, e.g. in the radioactivity problem we might reinterpret the parameter λ as *an observation from some distribution of a random variable* Λ (relating perhaps to a 'super-situation' containing the present one). Prior information is now expressed through the distribution of Λ. But we need to be specific about this distribution—and it is hardly specified by feeling 'confident that $0.45 < \lambda < 0.55$'. Again subjective elements arise in expanding our limited knowledge to a complete specification; arbitrary and personal assessments might have to be introduced. For this reason, some statisticians have in the past claimed that prior information (and an assessment of consequences) has no place in a formal theory of statistics—see Pearson's remarks quoted in the next section—but the present day view is more eclectic.

In contrast to Pearson, an amusing and compelling plea for the use of *subjective* prior information is given in an example by Savage (1961c).

After reminding us that subjective opinions often rightly influence the way in which a statistical investigation is *approached* (in terms of its design), he claims that subjective principles should also be allowed to influence the *analysis* of an experimental situation. To illustrate this in relation to the use of prior information, Savage presents three different experiments that he says have the 'same formal structure', where the conclusions based on traditional (*classical*) statistical method would be identical, but where the unemployed subjective prior information compels him (and, he is 'morally certain', anyone) to react quite differently in the three situations. The three experiments are as follows.

(i) The famous tea-drinking lady described by R. A. Fisher (1966, pp. 11–25) claims to know if the milk, or tea, was poured first into the cup. For ten pairs of cups of tea, each pair having one of each type, she makes the correct diagnosis.

(ii) A music expert says he can distinguish a page of Haydn score from one of Mozart. He makes a correct assignation for ten pairs of pages.

(iii) A somewhat inebriated friend, at a party, says that he can predict the outcome when a coin is spun. It is spun ten times, and he is correct each time.

In each case, the significance level is the same, 2^{-10}, for a one-tail test of significance. But Savage reacts quite differently in each case. Some 'old wives tales' have some substance; it seems that this *may* be so in the tea-making situation. An expert *should* be able to tell a page of Haydn from a page of Mozart; it is not surprising that he does so, particularly when he is so confident in his ability. With regard to the coin-spinning, Savage says that he does not believe in extra-sensory perception, and is unimpressed 'by this drunk's run of luck'.

The use of prior information in the Bayesian approach also has implications for the interpretation of the probability concept. Either a *frequency-based*, or a *subjective*, concept may be appropriate depending on circumstances. It is hard to see how anything but a degree-of-belief attitude can be applied to the 'confidence that $0.45 < \lambda < 0.55$' in the present example; although similar statements in different contexts may be legitimately interpreted on a frequency basis (see Chapter 2). Writers differ on the centrality of a subjective view of probability to the structure of Bayesian inference—many would claim it is the only appropriate view; at the other extreme a few would restrict the use of Bayesian methods to situations where prior distributions have a natural frequency interpretation. We shall need to return to this point in more detail, later in this chapter and in subsequent chapters.

We have now distinguished three general forms of information that may, depending on circumstances, be *relevant* to a statistical study. These can be viewed to some extent on a temporal basis—the prior information accumulated from *past* (or external) experience, the sample data arising from the *current* situation, and assessments of consequences referring to (potential) *future* action. The circumstances in which these different forms of information are relevant depend on a variety of factors, as we have seen in the examples above. The ways in which the information is quantified and utilised depend also on many circumstantial factors—in many cases involving subjective evaluation. But even if we judge particular forms of information to be *relevant* to our practical interest (and ignore for the moment the problems of quantification), utilisation of the information depends on having available statistical procedures *specifically designed to incorporate the particular forms of information.* This is the 'down-to-earth' level at which we must also contrast and compare the different approaches to statistics.

We must examine what different tools and concepts have been developed within the **classical, decision-theoretic and Bayesian approaches** for the processing of information? The *relevance* of information depends, then, not only on the practical situation but also on whether procedures exist to process it. In decision theory, for example, we need to be able to process prior information and consequences, both of which are *relevant* to the decision theory approach.

1.4 STATISTICAL INFERENCE AND DECISION-MAKING

The definition of the object of statistics, given in Section 1.1, implies a distinction between a need to *describe* the practical situation, and a need to construct *rules for action* in the context of that situation. Is this a valid, and useful, distinction? The examples of the agricultural experiments and of the α-particles suggest that it might be. Let us now look at this question of *function* in more detail.

Consider the problem of weather forecasting. The interests of the meteorologist and of the man-in-the-street, in what today's weather will be, are quite different. The meteorologist's interest is scientific: he is concerned with providing an informed *description* of the likely situation. The man-in-the-street is involved with using this description to aid him in his *actions*: to decide whether to take an umbrella to work or whether to go trout fishing, for which early morning rain is an advantage, etc.

The distinction between these two modes of interest in a statistical study, the *descriptive* and the *action guidance* functions, arises again and again. We have seen it in the context of examples (i) and (ii) in Section 1.2. It is a useful distinction to draw for the present purpose of contrasting different approaches to statistics. Any statistical procedure that utilises information to obtain a *description* of the practical situation (through a probability model) is an *inferential* procedure—the study of such procedures will be termed **statistical inference**. A procedure with the wider aim of suggesting *action* to be taken in the practical situation, by processing information relevant to that situation, is a *decision-making* procedure—the study of such procedures, **statistical decision-making.**

This distinction has often been made in the literature. For instance, Smith (1965) says:

> Statisticians often play down something which is obviously true, when it does not quite accord with their line of thought. An example is the statement that there is no difference between inference and decision problems.
>
> A decision problem means the choice between several possible courses of action: this will have observable consequences, which may be used to test its rightness. An inference concerns the degree of belief, which need not have any consequences, though it may. This makes it more difficult to come to agreement on questions of inference than on decisions. For example, the question 'Shall I eat this apple?' is a matter of decision, with possible highly satisfactory or uncomfortable outcomes. 'Is this apple green?' is a question of belief. Of course, the two problems must be closely related, even though they are distinct.

Cox (1958) discusses this in greater detail.

> A statistical inference carries us from observations to conclusions about the populations sampled.... Statistical inferences involve the data, a specification of the set of possible populations sampled and a question concerning the true populations. No consideration of losses [consequences] is usually involved directly [but]... may affect the question asked.... The theory of statistical decision deals with the action to take on the basis of statistical information. Decisions are based on not only the considerations listed for inferences, but also on an assessment of the losses resulting from wrong decisions, and on prior information, as well as, of course, on a specification of the set of possible decisions.

Here, we see the difference of *function* affecting the forms of information regarded as relevant. Prior information will be seen to be relevant to both *inference* and *decision-making*.

Lindley (1965b) remarks that the decision-making problem is an extension of the inference problem; but that inference is fundamental to decision-making.

> ... the inference problem is basic to any decision problem, because the latter can only be solved when the knowledge of [the probability model]... is completely specified.... The person making the inference need not have any particular decision problem in mind. The scientist in his laboratory does not consider the decision that may subsequently have to be made concerning his discoveries. His task is to describe accurately what is known about the parameters in question. (pp. 66–67).

At a later stage, the more detailed discussions of this issue, such as that by Lindley (1977), will be relevant.

So we see decision-making extending the aims of inference: the descriptive function of inference being expanded to suggest rules for behaviour. As a result, we may expect decision-making procedures to range more widely than inferential procedures over the types of information they are designed to incorporate and process. Broadly speaking, inference utilises sample data and, perhaps, other information with a bearing on the description of the practical situation, e.g. prior information. Decision-making augments the inferential knowledge of the situation by also incorporating assessments of consequences.

Consider the case of a car manufacturer who possesses a variety of information on components of his vehicles. He may wish to use this information in various ways. Sample data and prior information may be used to describe the dimensions of a particular component for quoting in a 'reference manual'. But this description needs to be augmented by a knowledge of the effects and implications of using the component if, say, there is some legal obligation to meet prescribed standards or if the component is to complement the overall structure of the car. Formal decision-making procedures may need to be employed to process a quantified expression of these effects and implications. Such expression may be critical to the appropriate action to take. A machine that produces 'oversize' pistons renders the whole assembly inoperable. The effects are obvious and extreme; the machine *must* be modified. A door handle that is awkward to operate may be detrimental to the sales image of the car. The consequences of retaining or replacing this type of handle are by no means as critical, nor easily quantified. (Advertising stratagems may be all that are required: 'What ingenious door handles! The children will never open them!')

This distinction between the reflection function and the action guidance function must show itself also in the statistical procedures used in the one context or the other. In particular, a decision-making procedure needs formal tools or methods to handle potential consequences, expressed as *utilities* or *losses*.

Most statisticians accept the dual nature of statistics implied by these two functions. There can, on occasions, be an advantage in separating out the message of sample data alone from the modifications that one must make in the light of additional prior, or consequential, information. This is important, for example, in appreciating the role of *classical statistics* in relation to some other approaches. If prior information exists however, it must surely be important to seek to incorporate it through the methods of *Bayesian inference*. We need now to examine such distinctions in more detail.

1.5 DIFFERENT APPROACHES

The object of this book is to present and contrast the different approaches to statistics. In doing so, it is essential to retain the distinction between inference and decision-making. It is apparent that most statisticians regard this distinction as fundamental (Smith, 1965), also that a large number are committed to a view of statistics that embraces some decision-making aspect. The vast literature on *decision theory*, at all levels of mathematical sophistication, applicability or philosophical comment, bears witness to this.

The distinction between inference and decision-making, coupled with the earlier subdivision of different types of relevant information, puts us in a position now to make a preliminary classification of the different approaches to statistics. For the moment we distinguish *three* main approaches (summarising their distinguishing features in anticipation of fuller discussion later).

(i) **Classical Statistics**, originates in the work of R. A. Fisher, J. Neyman, E. S. Pearson, and others. This includes the techniques of *point* (and *interval*) *estimation, tests of significance* and *hypothesis testing*. At first sight it might be judged to be an inferential approach, utilising sample data as its only source of relevant information—although we will find it necessary to qualify this assessment in certain respects later. The distinction between *significance testing* and *hypothesis testing* is crucial to the issue of the relative inferential/decision-making role of classical statistics. Any particular approach needs to incorporate concepts, tools and interpretations for its 'internal management'. In these respects classical statistics leans on a *frequency concept* of probability. It represents the sample data through what is termed their *likelihood*, and sets up certain criteria based on *sampling distributions* to assess the performance of its techniques. For instance, point estimators may be required to be *unbiased* or *consistent*, hypothesis tests are based on 'tail-area probabilities' and at a more general level data are shown to be best handled in predigested form as *sufficient statistics* where possible. The methods embody *aggregate* measures of their own 'reliability' or 'accuracy' such as *standard errors* or *efficiency* determinations. (See Chapter 5.)

Some comment is needed on the use of the term 'classical' to describe this approach. It is commonly so described, in recognition of its traditional role as a formal principle of statistical method more widely applied and of somewhat earlier origin (in terms of detailed formal development) than either *Bayesian inference* or *decision theory*. But the approach is also labelled in many other ways: *sampling-theory, frequentist, standard, orthodox*, and so on. We shall retain the term '*classical*' throughout the book, but this is not to be confused with the similar label attached to a particular view of the probability concept discussed in Chapter 3. It is interesting to note that Buehler (1959) calls the *Bayesian* approach 'classical'; the classical approach 'modern'—an extreme example of the non-standardisation of terminology in the area of comparative statistics.

(ii) **Bayesian Inference** is again essentially an inferential procedure, but admitting (indeed demanding) the processing of prior information as well as sample data. The prior information is modified by the sample data through the 'repeated use of Bayes' theorem' (Lindley, 1965b, p. xi) to yield a combined assessment of the state of knowledge of the practical situation. The *likelihood* again plays a crucial role. Inferential statements are expressed through *posterior probability distributions*, and hence embody their own measure of accuracy. The idea of *sufficiency* is again relevant, but *not so* the method of sampling used to obtain the sample data.

The expression of prior information through *conjugate* families of prior distributions, when appropriate to the practical problem, is of mathematical and interpretative convenience. This approach cannot rest on a frequency interpretation of probability *alone*; a *subjective* interpretation is almost inevitable, and probabilities tend to be regarded as conditional. Central to the basic development of Bayesian methods is the notion of *coherence* (an idea that goes back to Ramsey (1931/1964): see, for example, Lindley, 1971c, pp. 3–6). This is a concept, that seeks to express in precise terms what is required of individuals if they are to react 'rationally' or 'consistently' to different prospects in situations of uncertainty. The use of Bayesian methods is not restricted to situations where tangible prior information exists; the formal expression of *prior ignorance* is important and arouses great interest and some controversy. An assumption of *exchangeability* (an operational symmetry in the parameterisation of the model) can facilitate the solving of multi-parameter problems. (See Chapter 6.)

(iii) **Decision Theory**, stemming from the work of Wald was first presented in his detailed treatment (1950). As the name suggests, this approach is designed specifically to provide rules for action in situations of uncertainty, i.e. *decision rules*. It inevitably incorporates assessments of the *consequences* of alternative *actions*, expressed through the mathematical theory of *utility* in the form of *losses* or *loss functions*. The value of any decision rule for prescribing action on the basis of sample data (and any prior information) is measured by its *expected loss*, or *risk*. The aim is to choose a decision rule with 'minimum risk'; and the concepts of the *admissibility* of a decision rule and of *complete classes* of decision rules are central to the study of optimality.

There is no derived probabilistic assessment of the accuracy of decision rules; their relative merits are measured in aggregate terms by their associated risks. This approach may be regarded as the stochastic extension of the deterministic *games theory*, due to von Neumann and Morgenstern (1953; first published 1944): it is concerned with 'games against nature', and a *minimax* principle is one basis for the choice of an 'optimum' decision rule.

In as far as prior information is incorporated, the methods used are essentially those of Bayesian inference. No particular philosophical view of probability is inevitably implied in decision theory; usually probability statements are frequency-based, although when prior information is processed a subjective attitude is often adopted. Whilst decision theory does not depend on, or demand, the use of Bayesian methods or a subjective interpretation of probability, the study of optimum decision rules is simplified on the Bayesian approach. At the formal level, adoption of prior distributions allows us to delimit (in many situations) the range of admissible decision rules, irrespective of whether or not any particular prior distribution is appropriate to the real problem in hand. At a fundamental level, Ramsey (1931/1964) shows that Bayesian decision theory is the *inevitably correct* way in which decisions *should* be made if we were entirely rational: in that *coherence* implies the existence of prior probabilities, and of utilities, and the need to maximise expected utility. See Lindley (1971b) for a down-to-earth explanation of this *normative* role of Bayesian decision theory. But not everyone is entirely 'rational' in their actions and reactions.

Attitudes to decision theory vary from one statistician to another. Some see it as an objective procedure for prescribing action in real and important situations where it is possible to quantify consequences, others as a tentative formal model of the way people behave (or ought to behave) in the day-to-day running of their lives. (See Chapter 7.)

On this simple classification the *main characteristics of the three approaches* may be summarised in the manner indicated in Table 1.5.1:

This broad classification of approaches to statistical inference and decision-making oversimplifies the true structure. At this stage the descriptions under (i), (ii) and (iii) are intended to provide merely an intuitive feeling for distinctions between the approaches, and a first contact with some of the vocabulary of inference and decision-making.

A matter that is of importance in whatever approach is used for inference is the question of whether or not *statistical* (inferential) *association* or relationship has a *causal* origin. Thus, road accident figures are seen to increase over the years (until recent times) with an increase in the numbers of lorries and trucks on the road. This is a statistical relationship. But can we go further and infer that the increase in commercial traffic *causes* the increase in the number of accidents? Consider another example.

In the late 1940s, the Medical Research Council in England expressed concern about the apparent large increases in cases of lung cancer. Doctors took note of this concern and data were collected on admissions to hospitals in the London area of patients suffering from respiratory problems. Their subsequent diagnoses were examined and those with lung cancer were compared with the others to see whether there was any evidence of factors distinguishing the two groups. It came as a surprise to find a much higher proportion of cigarette smokers in the

Table 1.5.1

Approach	Function	Probability concept	Relevant information
Classical	Inferential (predominantly)	Frequency-based	Sample data
Bayesian	Inferential	'Degree-of-belief'; subjective. Possible frequency–interpretable components	Sample data Prior information
Decision theory	Decision-making	Frequency-based	Sample data Consequential losses or utilities
		('Degree-of-belief': subjective, when prior information is incorporated)	(Prior information)