Gottinger · Elements of Statistical Analysis

Hans W. Gottinger

Elements of Statistical Analysis

Walter de Gruyter Berlin · New York 1980

Library of Congress Cataloging in Publication Data

Gottinger, Hans-Werner.
Elements of statistical analysis.

Bibliography: p.
Includes index.
1. Mathematical statistics. 2. Probabilities.
I. Title.
QA276.G645 519.5 80-10104
ISBN 3-11-007169-X

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Gottinger, Hans-Werner:
Elements of statistical analysis / Hans W. Gottinger. – Berlin, New York : de Gruyter, 1980.
ISBN 3-11-007169-X

Satz: Tutte, Druckerei GmbH, Salzweg/Passau.
Druck: Karl Gerike, Berlin.
Bindearbeiten: Lüderitz & Bauer Buchgewerbe GmbH, Berlin.

Preface

This book has been designed as an introductory textbook on an elementary level with emphasis on application in the behavioral sciences, yet, it is of sufficiently methodological orientation to being used as a one-semester course for undergraduate studies that requires only a limited background in high school algebra, except for the more technical, starred chapters.
Furthermore, it can be used as a supplementary text in connection with broad coverage textbooks in statistical analysis, or even as a self-instruction manual, for beginning graduate students in comprehensive programs of economics, sociology, psychology, education, system and industrial engineering, or related fields. Equipped with this material the student should be able to work out simple problems for himself arising in his specific field of study. For this purpose a number of problem sets are given for self-instruction.
The present book emerged from a half-year lecture course, repeatedly taught at the University of Bielefeld, the University of California, Santa Barbara and at the Technical University Munich – in different departments, to students of different backgrounds.
Formally the book is organized in 10 chapters, for some chapters, e.g. 3*, 4* and 9*, technical supplements and survey-type material have been added to enhance better understanding of key concepts.
It attempts to be well balanced given its limited scope between methodology, statistical analysis and techniques. For the uninitiated among the readers it appeals more to intuition and common-sense reasoning, yet the material is presented in a reasonably rigorous fashion. Various options are left to the reader regarding further study on more advanced technical aspects, for this purpose a bibliography is added. Finally, this is to express my great appreciation to de Gruyter, Berlin for its cooperation in designing the book and for making possible last-minute changes in the text.

Bielefeld, März 1980 *Hans W. Gottinger*

Contents

Part I. **Foundations of Probability and Utility**

Part II. **Simple Random Processes and Probability Distributions**

Part III. **Topics in Statistical Inference, Information and Decision**

Part I.
Foundations of probability and utility

1. Methodology of Statistical Analysis

1.1 Conception of Statistics

All conceptions of statistics agree that statistics is concerned with the collection and interpretation of data. The data may be pertinent to an immediate practical action or decision, for instance, which best course of action – if available – to pursue under a situation of uncertainty or complete ignorance. But data may enhance knowledge without being immediately pertinent to a practical action, as in measurements of outcomes of natural laws or physical processes.

In application, the distinction between practical action and enhancement of knowledge is rarely clear-cut. Often, in real-life situations, it turns out that provision of knowledge is a *sine qua non* condition for choosing reasonable actions.

An essential ingredient of problems called „statistical" is that the relevant data are imperfect; we must always deal with uncertainty.

It is natural to ask why there should be a separate discipline of statistics. Certainly data are often collected and interpreted by people with little or no knowledge of statistics. The justification for statistics as a separate discipline lies in the hope that there will be general principles common to diverse applications. Some statistical principles consist of little more than common sense, and while statistics must ultimately be consistent with common sense, the implications of common sense in complicated problems are far from obvious. Much statistical reasoning can be conveyed by informal discussion of examples that illustrate simple uses and misuses of statistics, but for a deeper grasp of the subject, a more systematic and formal development is needed. Broadly speaking, the modern approach to statistics can be characterized by the words *inference* and *decision*, which refer to the processes of drawing conclusions and making decisions in the light of available data, or determining what additional data should be obtained before concluding and deciding. 'Conclusion' is used here in a technical sense, defined by J. W. Tukey (1960), as 'a statement which is to be accepted as applicable to the conditions of an experiment or observation unless and until unusually strong evidence to the contrary arises'.

1.2 History of Bayesian Statistics

A systematic treatment of utility and subjective probability according to Ramsey, de Finetti and Savage has stimulated the discussion on the foundations of mathematical statistics and its relationship to statistical inference. The core of this discussion is based on the 'behavioristic' interpretation of a statement (rather than of a 'theorem') implicitly derived by the English Clergyman Thomas Bayes (1763)*.

The proponents of this behavioristic interpretation of Bayes' statement are called Bayesians and their arguments have led to the Bayesian analysis in statistics. There are different kinds of Bayesians, but they all agree at least on the following point: It is possible to draw statistical conclusions from the conditional probability $P(H|E)$, that is the probability of a hypothesis H (to be true) given that the event E has been observed (to be true). The so-called 'Bayes Theorem' then is a trivial consequence of the product axiom of probability theory. However, it is more than a belief in this 'theorem' that distinguishes someone to be a Bayesian, it is the general acceptance of the idea to use a concept of intuitive probability in statistical theory and practice, to motivate this concept on a decision-theoretic basis and beyond that to find many applications in the experimental sciences.

There are two essential characteristics of Bayesian statistics and they can be listed in a simplified manner as follows.

(1) Probability evaluation is based on experience, intuition, and personal assessment combined with a number of consistency criteria relating to a rational person.
(2) Treatment of statistical data is continuously revised on the basis of new information, or evidence that is available to the decision maker (statistician).

Modern Bayesian statistics, in general, rests on three main construction blocks, consisting of:

(1) the game-theoretic studies of von Neumann and Morgenstern.
(2) the statistical works of J. Neyman and A. Wald.
(3) the subjective probability interpretations of F. P. Ramsey, B. de Finetti, B. O. Koopman, L. J. Savage et al.

Some of these elements have been developed within the classical statistical theory. A combination of all of these elements, however, forms the foundations of Bayesian analysis. Furthermore, there are two external factors that support this view, e.g. the philosophical attitude that most or all

* Th. Bayes, 'An Essay towards solving a problem in the Doctrine of Chances,' Philosophical Transactions of the Royal Society 53, 370–418 (Reprinted in: Biometrika 45, 1958, 293–315).
In crude form Bayes derived the statement that the probability of a 'certain cause' will be subject to change given that certain events will occur. In this statement the probability concept is used inductively for the first time by inferring from a small sample to the whole population.

scientific inferences result from 'inductive' rather than 'deductive' reasoning and the psychological viewpoint that 'statistics is a theory of the uncertain environment in which man must make inferences'. (Petersen and Beach, 1967), e.g. that human information processing is just an 'inconsistent' case of optimal information processing of data as required by statistical inference. Much of what Bayesian statistics has received as inputs in terms of new ideas is based on results obtained in experimental areas (such as psychology); to some extent, therefore, one could speak of a behavioral approach to statistical methodology. Although this process started from Bayes' fundamental work and reached the time of de Laplace, it has been cut off after de Laplace and only quite recently has been rediscovered by the 'Bayesians'. It is therefore illuminating to give a brief account of the main events in the development of statistical methodology. In the nineteenth century there was an increasing awareness among statisticians that a connection between probability theory and various methods of using data in a consistent fashion should result in a construction of a theory of statistical inference. Such a theory would permit predictions on the basis of a wise use of data and with tools provided by probability theory.

There were studies in this direction by Quetelet, W. Lexis, F.Y. Edgeworth, K. Pearson, culminating in the work of R.A. Fisher's *Statistical Methods for Research Workers* (1925). J. Neyman developed Fisher's ideas further and around 1940, say, statistical theory was firmly based on this standpoint – 'Fisher as seen through Neyman's eyes' (F.J. Anscombe, 1961) – which still prevails among eminent contemporary statisticians, although their number is slightly decreasing over the past decade. Since 1940 the theory of statistical decisions emerged paralleling the more orthodox theory of the Neyman-Pearson school.

In the more recent theory you structure any statistical problem as a decision problem where the statistician is engaged in a game against nature, and the only way of gaining information is by doing experiments. Again here two phases can be distinguished. The first phase was introduced by A. Wald's *Theory of Statistical Decision Functions* (1950). This theory still adopts a frequentistic interpretation of probability. The second phase is truly 'Bayesian', it emphasizes the point that the structure of a decision problem consequently requires a behavioristic interpretation of probability, that is a non-frequentistic concept of personal probability.

Now, what is it that makes the Bayesian method so attractive for many experimental situations that are not restricted to social or behavioral sciences but also extend to certain problems in the natural sciences (see I.J. Good, (1969)).

What are the specific prerequisites for the application of the Bayesian method, what specific kinds of information do we need?

- First of all, we need a statistical specification in an observational model in which observations are assumed to be realizations of random variables represented by a set of conditional probability distributions, conditional by a set of parameter values (states of nature). Let us assume, we have a finite number of states of nature, say m, and denote them by $\theta_1, \dots \theta_m$, and furthermore let us have a finite number of outcomes, denoted by $t_1, t_2, \dots t_n$. Then we can calculate for all m · n combinations of states and outcomes the direct probability of an outcome, given a state, denoted by $P(t|\theta)$.
- Second one needs a utility or loss function indicating the relative desirability of available decision acts for a given set of parameter values.
- Third, one needs a marginal probability distribution over the parameter space, i.e. an a priori (subjective) probability distribution.

Now the first condition has been universally accepted by all relevant schools of statistics, e.g. by the Bayesians as well as by the classical school – there is no disagreement about that. The second condition has been introduced by A. Wald, but only the third requirement is typical for the Bayesian method. It is this requirement which is most controversial for classical statisticians, and, as we shall see, centers around the validity and interpretation of Bayes' theorem.

Let us first deal with the third requirement (in natural conjunction with the first one), which, in general, leads to the inference rather than the decision problem, and then see how and why Bayes' theorem is essential for its formulation.

In the frequentistic interpretation of probability in terms of the limit of long-run frequencies it is impossible for the statistician to measure uncertain events (which are not repeatable) by probabilities.

In the frequentist's view an uncertain event on which no past history exists, is either considered to be not measurable by a probability or the probability of zero or one (and one does not know which) is assigned.

'If we ask the probability that the 479th digit in the decimal expansion of π is a 2 or a 3, most people would say 2/10, but the frequentist, if he answers the question at all, must say 0 or 1, but that he does not know which'. (J. Cornfield, 1967, p. 44). The frequency concept in connection with the construction of significance levels, errors of type I or II, confidence intervals etc. only answers the question as to how certain we feel on the basis of the given data (a posteriori distribution) but does not answer the question as to how certain we feel in advance when we still don't know the data (a priori distribution).

We could characterize an individual's state of uncertainty with respect to a proposition by the betting odds he would offer with respect to it. Consider an event (proposition) A and let P(A) be the probability of this event, for which you would receive \$1 in case the proposition A is true, otherwise you receive nothing. Let A be the proposition 'it will rain tomorrow' and somehow you arrive at a probability (estimate) of $P(A) = 1/3 = p$. Then, in other words, you would be willing to pay 33 cents in exchange for receiving \$1 provided the event A occurs or the proposition A is true. This is equivalent in ordinary language to saying that you bet on an amount \$1 at odds p to $1 - p$ on the occurrence of A. In some way p could be considered as your entrance fee to enter a betting contract. Of course, if p is greater than unity, you are certain to lose whether or not A is true. You wouldn't consider such a bet 'fair' and most likely refuse to accept such a bet. On the other hand, suppose you want to bet on 'rain' and pay $p = 20$ cents, and also bet on 'no rain' and pay $q = 30$ cents, in this case your betting partner wouldn't consider such a bet as fair, since whatever proposition turns out to be true he would have to pay \$1 and only receives 50 cents in return. Such a probability assignment could be termed 'incoherent'; to have a coherent assignment you have to select probabilities which sum up to one. Hence, fair betting implies coherent assignments of probabilities and this can be shown to satisfy Kolmogorov's finite additivity axioms of probability theory (see Chapter 2..1). [Also, the notion of conditional bets can be introduced by considering events which are not mutually exclusive, i.e. if an individual smokes he will develop lung cancer, if he doesn't then there is no bet.] The construction of personal probabilities via a betting contract is obviously related to the concept of conditional probability and this again will quite naturally lead to the formulation of 'Bayes' theorem'. This will be shown next. In the following exposition we draw heavily on Cornfield's review.

Let there be two classes of proposition, A and B, every class contains mutually exclusive propositions. For simplicity, let us first assume that class A contains two propositions: an individual has or has not developed lung cancer during some definite time interval. The class B may contain two propositions, either the individual is found to be a smoker or a non-smoker. Extension to more than two possibilities in each class presents no difficulty. Now the proposition space would consider of four points

$$A_1B_1,\ A_1B_2,\ A_2B_1,\ A_2B_2,$$

all A's and B's are not mutually exclusive. A_1B_1 would mean 'the individual has developed lung cancer and is a smoker'. $P(A_1)$, $P(B_1)$ would be the unconditional probabilities, but one could also define the conditional probability $P(A_1 | B_1)$, i.e. the probability of developing lung cancer given that one is a smoker. It could be defined by

$$(1) \qquad P(A_1 | B_1) = P(A_1B_1)/P(B_1).$$

Now on the basis of Kolmogorov's axioms of finite-additive probability together with (1) we could derive Bayes' theorem in a straight-forward fashion. By symmetry, and dropping the subscripts we have

(2) $$P(B|A) = P(BA)/P(A) = P(AB)/P(A),$$
$$\text{so that } P(AB) = P(B|A).\ P(A).$$

As can easily be seen, B is the union of the mutually exclusive events A_1B, $A_2B, \ldots$ so that $P(B) = P(A_1B \text{ or } A_2B \text{ or} \ldots)$, and by the additivity axiom

$$P(A_1B \text{ or } A_2B \text{ or} \ldots) = P(A_1B) + P(A_2B) + \ldots,$$

hence (3) $P(B) = \sum P(A_iB)$.
Now (2) can be changed to

(4) $$P(A_iB) = P(A_i)P(B|A_i).$$

If we start from (1), reformulate the numerator according to (2), and the denominator according to (3) and (4) we get

(5) $$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$$

This is 'Bayes' Theorem'.
Making a notational change by identifying the propositions $A_1, A_2, \ldots$ with states of nature $\theta_1, \theta_2, \ldots$, and the propositions $B_1, B_2, \ldots$ with the outcomes $t_1, t_2, \ldots$, we have 'Bayes' Theorem' in the form.

(6) $$P(\theta|t) = \frac{P(t|\theta)P(\theta)}{\sum P(t|\theta_i)P(\theta_i)}, \quad \text{where}$$

$P(\theta|t)$ is referred to as the a posteriori probability (posterior probability), $P(\theta)$ as the a priori probability (prior probability), and $P(t|\theta)$ as the likelihood.
(6) expresses the fundamental fact of 'learning by experience' in terms of the relation of prior and posterior probability.
It is the interpretation of this result which has triggered most of the controversy rather than the mathematical deduction, which is, without doubt, a correct one.
Let us first observe some properties of this relation.
(1) If $P(t|\theta) = 1$ and yet $\bar{t}$ (not t) has been observed (to be true), then $P(\theta|t)$, the posterior probability, is zero, i.e., an initially plausible hypothesis is rejected by the test.
(2) If $P(t|\theta_i)$ is the same for all i, then the posterior probability is equal to the prior probability, i.e., any additional information would not change the posterior probability.
(3) If $P(\theta) = 0$ then also $P(\theta|t) = 0$. If a proposition is initially false, then no information whatsoever will change the initial probability assessment.

I. In the philosophy of science an influential school of thought stressed the view that scientific conclusions based on past observations are not deductive. The theory of inductive inference originating with the work of David Hume in the last century has been elaborated and virtually extended to a philosophical school of thought by Rudolf Carnap.

Carnap's *LogicalFoundations of Probability* (1950) is just a straight-forward extension of his general principles of induction in scientific inference. Harold Jeffreys in his *Theory of Probability* (1961) devised five 'essential' rules of inductive inference under which 'Bayes' Theorem' could be subsumed as representing one important case of probabilistic inference.

II. Bayes' theorem has been accepted and used by Laplace, but some decades after Laplace the first critical voices have been heard. They centered around the construction of prior probabilities, in particular, how could one justify any assignment of probabilities to various states of nature. There were objections by Boole and Cournot, but much later – in the development of the theory of statistical inference under K. Pearson and R.A. Fisher –Bayes' theorem was not used at all and outrightly rejected by R.A. Fisher (1941). 'The theory of inverse probability [i.e. Bayes' theorem] is founded upon an error and must be wholly rejected. Inferences regarding populations from which known samples have been drawn, cannot by this method be expressed in terms of probability'. This questions the possibility of assigning prior probabilities to various states of nature. Frequentists, in particular, are troubled by the concept of prior probability. H. Cramér (1946) points out: '... the foremost weakness of this argument is that the prior frequency function $\pi(m)$ is in general completely unknown... Also irrespective of this, the argument suffers from the fundamental error that the true value of m is in most cases not the result of a random trial and may therefore never be regarded as a stochastic variable. Usually m is simply to be regarded as a fixed though unknown constant ... and on the whole under such circumstances no prior frequency function exists. Bayes' theorem is therefore practically useless for the theory of error and its use in this field should be replaced by the method of confidence limits.'

This and similar criticisms will probably in many cases be based'on special interpretations of the probability concept. These points of view imply, roughly, that one only accepts the probability of something if this something can be registered in experiments which can be repeated. The probability can then be approximated by the relative frequency of this something in a long series of trials. On the other hand, if the numerical value of a probability is interpreted as representing the degree of belief a prior probability statement (on some parameter or state of nature) should be fully legitimate.

Yet, it should be mentioned that while an advocate of subjective probability will find no 'ideological' barrier to apply Bayes' theorem in statistical inferences, the theorem can be used when making these inferences by a non-Bayesian. One cannot criticize Bayes' theorem on grounds that it is used by Bayesians, as come critics do, since provided non-Bayesians agree that statistical inferences should be based on an revision of data in the light of new information there is no effective alternative open to them other than Bayes' theorem. However, the trouble is, that they have to find a frequency interpretation for the prior probabilities, whereas the Bayesian is much more flexible in view of his probability concept. Let us see where the real source of difficulty is located.

If θ is a random variable with a well-defined frequency distribution, which is known then the frequency distribution is the prior probability function and there should be no controversial point between Bayesians and non-Bayesians in this case. Controversies will arise if θ is an unknown constant (not a random variable) and has no past history. Then according to the frequentist the probability is not defined, the Bayesian, however, would apply his subjective probability concept. It is obvious that the effect of the prior probability on the posterior probability will be diminishing to the extent that more and more information will become available through the likelihood which would modify the initial prior probability. As a consequence of this, two scientists (or statisticians) having initially quite different priors will eventually arrive at the same posterior probabilities when faced with a sufficiently large body of data – provided the priors are all non-zero. This fact has been rigorously proved in a paper by D. Blackwell and L. E. Dubins (1962).

As regards the nature of prior probability assignments the Bayesian would certainly utilize any information contained in samples of past data to construct his prior, in this case it is said that his prior is 'data-based' (A. Zellner, 1971, 2.3). This does not necessarily mean that all conditions will be satisfied that permit the specifications of a frequency distribution, e.g., if we deal with small samples of data, for instance. In other cases prior information may be obtained on the basis of introspection, casual observation or even from plausibility arguments, this could be referred to as a 'non-data-based' prior. It is clear that differences of opinion between statisticians are most likely to occur by the use of non-data based priors. To arrive at a prior probability judgment it is often convenient, and in the spirit of the Bayesian approach, to separate information (as represented by data or other sources) from probability analytically and to consider the process under which different degrees of subjective information will induce corresponding probability evaluations. One can then argue that the probability assessments represented by the prior have a sound information-theoretic basis. Work in this direction has been done by Gottinger (1973, 1974).

A particular problem which could arise is the case of complete ignorance or 'knowing a little'. In this case it has been suggested by H. Jeffreys (1961) that if the unknown parameter θ lies in some finite range its probability distribution should be taken as uniformly distributed. This proposal corresponds to the Laplacean principle of 'insufficient reason' where equal probabilities are assigned to completely unknown states.

As R.L. Plackett (1966) observed, when the number of observations is sufficiently large the likelihood will have a sharp peak at the maximum likelihood estimate of θ, so in forming the posterior distribution only a small interval of the prior distribution is relevant. Therefore it is sufficient to introduce a 'locally uniform' or 'gentle' prior distribution for an unknown parameter centering around the maximum likelihood estimate, but taking any form outside the range since these values get multiplied with only negligibly small likelihoods so that the posterior distribution is barely affected by this.

III. The concept of a loss function – as introduced by A. Wald – is essentially a counterpart of von Neumann-Morgenstern's utility function (1947) which came up around the same time. It is another basic element of Bayesian analysis, however, here the emphasis lies on 'decision' rather than 'inference'. A. Wald and some of his followers were inclined to think of any statistical inference problem in terms of a statistical decision problem. This view seems plausible for certain activities of the statistician (such as hypothesis testing which could be looked upon as preferring certain decision rules over others), but would not pertain to problems where the statistician only wants to observe and then draw conclusions on the basis of observations – such as choosing between rival cosmological theories, or in a medical diagnosis problem where conclusions may result in decisions (regarding the medical treatment of the person concerned) but they may also be valuable for themselves. In a statistical decision problem we consider m possible states of nature $\theta_1, \theta_2, \ldots, \theta_m$, and n possible outcomes $t_1, t_2, \ldots, t_n$. We assume that the statistician (decision-maker) can choose among a set of possible decision acts, denoted by $a_1, a_2, \ldots, a_r$. Now define a decision function as a real-valued function with the characteristic property

$$d(a_k, t_i) = \begin{cases} = 1, & \text{if } t_i \text{ results in } a_k \\ = 0, & \text{otherwise.} \end{cases}$$

Hence a decision function is a rule which assigns for a given state of nature an act a to the outcome t.

Consider a finite number, say p, of possible decision functions $d_1(a_k, t_i)$, $d_2(a_k, t_i), \ldots, d_p(a_k, t_i)$, among which you have to find the 'best' decision function. In order to establish a selection criterion A. Wald introduced the concept of a loss function $l(a, \theta)$ that could have a positive, negative or zero value. By choosing among decision rules you would prefer rules which result in the smallest losses. By definition, $l(a, \theta)$ is a random variable. The loss function, as a real-valued function, is in fact the negative counterpart of von Neumann-Morgenstern's utility function. Therefore, all their utility axioms that are required to prove the existence of such a function apply equally to Wald's loss function. One particular result of von Neumann-Morgenstern that matters here is the continuity property for gambles. This property is a direct consequence of the archimedean type axiom of utility theory. Roughly, the property implies that as long as there do not exist infinitely large positive or negative utilities (losses) it only matters that utilities (losses) are ordered according to their expected values.

Suppose you have three losses l_1^*, l_2^*, l_3^* with $l_1^* < l_2^* < l_3^*$. According to the continuity property we claim that $l_2^* = (p^* l_1^*, \bar{p}^* l_3^*)$ where $(\ldots, \ldots)$ is a gamble with $p^* \in (0, 1)$, $\bar{p}^* = 1 - p^*$.

Furthermore we would have

$$\begin{aligned} l_2^* &< (pl_1^*, \bar{p}l_3^*) \qquad \text{for } p > p^*, \\ l_2^* &> (pl_1^*, \bar{p}l_3^*) \qquad \text{for } p^* > p. \end{aligned}$$

On the right-hand side we observe the expected loss (gamble), the problem then is to minimize the expected loss:

$$E[l(d, \theta_j)] = \sum_i \left[\sum_k l(a_k, \theta_j) d(a_k, t_i)\right] p(t_i | \theta_i) = R(d, \theta_j),$$

the risk function.

Having defined the expected loss we would like to choose those decision rules that improve the risk function stepwise. We say a decision rule d_1 is at least as good as d_2 if and only if $E[l(d_1, \theta_j)] \leqq E[l(d_2, \theta_j)]$ for $j = 1, 2, \ldots, m$. If the strict inequality holds for at least one j then we say d_1 *dominates* d_2. A decision function is *admissible* if it is not dominated by any other decision function.

To compute the expected value of $l(a, \theta)$ over all possible states of nature, we set $E(l(a, \theta)] = \sum_j l(a, \theta_j) P(\theta_j | t)$ for every possible a. To choose an a which would minimize $E[l(a, \theta)]$ is called Bayes' decision rule, given the chosen prior probabilities. Thus there is a family of Bayes decision rules, corresponding to the possible prior distributions for the states of nature.

It can be proved that Bayes' decision rule is both necessary and sufficient for admissibility. This would put everyone – arguing for admissibility but denying the existence of prior probabilities – in an almost untenable position since every decision rule which is admissible is also a Bayes' decision rule relative to a particular set of prior probabilities. Finally, a few words should be said about the relationship of the decision and inference in Bayesian statistics.

In business and industry, actions must be and are taken in the face of uncertainty about nature. In such a framework the application of decision theory seems most natural and profitable, although technical problems such as assigning prior probabilities and specifying utilities will arise. The scientist, on the other hand, seeks to broaden and deepen his understanding of the laws governing his science. He performs experiment after experiment, and essentially revises his findings in the light of new information provided by additional experiments. Decisions are involved but often they are coupled with problems of inferences.

When the aim of a statistical analysis is that of describing or making inferences about nature, whether this will be the population of a country or the laws governing a physical or social process, in these cases the highly structured theory of decision making may not be so appropriate. One view of the Bayesian approach to statistics, motivated by such considerations as these, is that it provides a vehicle for the reduction of data, transforming problems into 'no-data' problems, the data being used to generate a posterior distribution. Summarizing, the Bayesian approach to statistical inference is based on an argument of the following form:

(i) It is the business of the scientific experimenter to revise his opinions in an orderly way with due regard to internal consistency and the data, and so

(ii) one has to develop techniques for the orderly expression of opinion with due regard to internal consistency and the data, but

(iii) the only orderly expression of opinion with due regard to internal consistency is the Bayesian one, and

(iv) the only orderly revision of those opinions with due regard for the data is through Bayes' theorem, therefore

(v) the Bayesian approach to statistical inference is the most 'natural' one.

1.3 A Quality Control Example

More than by any philosophical discourse on Bayesian analysis it is illuminating to introduce the key Bayesian ideas by a simple example, modified from one originally given by Schlaifer (1969). The purpose is only to give you some concrete feeling for the important ideas. Technical terms will be defined mostly in the context, and details will be developed later.

An automatic machine has just been adjusted by an operator, and we are uncertain as to how good an adjustment has been made. In principle it is possible to make an exhaustive and mutually exclusive list of *events* or *states of the world* that are relevant to the problem: one of these events surely obtains but we are uncertain as to *which one.* The word "event" can be interpreted informally as "something that might happen" or "something that might be true." But an event may refer to something that has occurred already but is unknown to us, and this is the situation in the example we are now presenting. The events of the example can be described by the probability p that the machine will turn out a defective part. For simplicity it is assumed that there are only four events (representing adjustments of the machine) and they can be described by values of p: p = .01, .05, .15, .25.

You can think of p in terms of betting odds (see Sec. 2.3). Whichever p is of the four possibilities – .01, .05, .15, .25 – we assume that it will remain constant during the production run now being contemplated, which consists of 500 parts.

If we knew that p = .01, which represents the best possible adjustment, we would be satisfied with the operator's adjustment. If, on the other hand, we knew that p = .25, we might be tempted to change the adjustment in the hopes of improvement. Suppose that there is a mechanic who can, without fail, put the machine in the best possible adjustment. We are told that the time needed by the mechanic to make the necessary adjustment should be valued at $ 10. The problem is to decide whether or not to incur this $ 10 cost.

We shall for the moment assume that just two *acts* or *decisions* might be taken: (1) *acceptance* of the adjustment, that is, do not check it; (2) *rejection* of the adjustment, that is, have it checked by the master mechanic.

For each possible combination of event and act, we assess the *expected net cash flow* that will ensue if that act is taken and that event obtains. To explain this, we shall assume first that $.40 is the *incremental* cost needed to rework a defective part, regardless of how many defective parts are produced. The incremental cost of a non-defective part is, of course, $0. Now if the probability p of a defective part is .01, the probability of a non-defective part is 1 − p = .99. To calculate the expected cost of a defective part for a production run of one, given the best adjustment, we *weight* $.40 and 0 by the respective probabilities, .01 and .99, as follows:

$$(.01)\,\$.40 + (.99)\,\$0 = \$.004.$$

If we equate the probability with long-run relative frequency, this equation can be interpreted as follows. In the long-run, .01 of the parts are defective and an incremental cost of $.40 is incurred, .99 are non-defective and the incremental cost is 0. *On the average in the long run*, then, the cost of defectives per part produced, the expected cost per part, is $.004. Since a decision is to be made about a production run of 500 parts, we then multiply $.004 by 500 to get $2.00, the expected cost of defectives per 500 parts produced, that is, per production run. The use of long-run frequencies is introduced to help to visualize the concept of expectation. Expectation and probability also have meaning even if there is a single unique choice, never to be repeated; then the idea is that of betting odds, not long-run relative frequencies.

Similar calculations for p = .05, p = .15, and p = .25 give expected costs for the act of acceptance as $10, $30, and $50.

For the act of rejection, the computation is even simpler. Regardless of the event that obtains, the mechanic achieves the best adjustment p = .01,, so that the expected cost of defective product is

$$500[(.01)\$.40 + (.99)\$0] = \$2.00.$$

In addition to this we must count the $10 for the mechanic's time, which also is the same regardless of the machine's actual adjustment. Hence the expected incremental cost for rejection is $2.00 + $10 = $12.00

All this information can be summarized in a *payoff table*. This is a two-way table in which the row headings are possible events, column headings are possible acts, and the entries are expected incremental profits or costs, as the case may be, for each event-act combination. Table 1–1 is the payoff table for the present problem.

Table 1–1: Payoff Table

Event	*Act*	
p	Acceptance	Rejection
.01	$ 2*	$ 12
.05	10*	12
.15	30	12*
.25	50	12*

With this information alone it is frustrating to decide whether to accept or reject. Acceptance is clearly the better act if p = .01 or .05, but rejection is better otherwise, as is indicated by the asterisks in Table 1. If the event is known, the best decision is obvious, but the problem is a problem of uncertainty as to which event obtains. Your decision depends on your assessment of the probabilities to be attached to the four possible events. How do you arrive at the needed probabilities? Suppose that there is extensive evidence on the history of the fraction of defective parts in 1000 previous long production runs under similar conditions in the past, and that this history is summarized in Table 1–2. The needed probabilities are assessed by the relative frequencies.

The basic criterion for decision can now be applied: choose that act for which expected cost is lowest (or, for which expected net revenue is highest).