^{2}, $0 \leqslant \theta \leqslant 1$; then θ^2 should also have a uniform distribution, which is impossible. For this simple problem, there have been many "solutions." Haldane (1932) proposes

$$\xi(\theta) \propto \theta^{-1}(1-\theta)^{-1}, \quad 0 < \theta < 1,$$

and others, including Jeffreys, use

$$\xi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}, \quad 0 < \theta < 1,$$

to express prior ignorance for θ. Thus we see, from its inception, Bayesian inference is a controversial subject, but not much was heard about it until the twentieth century, where during the last thirty years many advances have been made.

The current activity can be divided into Bayesian decision theory and Bayesian inference, where the former uses three sources of information (in making statistical decisions), namely the sample data s, the prior density $\xi(\theta)$ of the parameters, and the loss function which measures the consequences of making an incorrect decision. On the other hand, Bayesian inference only uses the prior density of the parameters and the sample data because with inference one is not interested in making decisions, only investigating the values of the parameters of the model or making predictions about future observations.

As mentioned earlier, the books by Lindley (1965), Zellner (1971), Box and Tiao (1973), Winkler (1972), and Press (1982) give introductions to Bayesian inference while DeGroot (1970), Ferguson (1967), Raiffa and Schlaifer (1961), Savage (1954) and Berger (1980) deal with decision making.

The early work in Bayesian statistics of this century was done by deFinetti (1930), Jeffreys (1939) and Ramsey (1931/1964). deFinetti and Ramsey develop a formal theory of subjective probability and utility which was continued by Savage (1954), Lindley (1971), and DeGroot, among others, while Jeffreys gave the foundation for Bayesian inference, which was continued by Lindley (1965), Box and Tiao (1973), Winkler (1972), Zellner (1971) and Press (1982). Most of these writers in the latter group employ improper type prior distributions for the parameters, but Raiffa and Schlaifer (1961) introduce conjugate prior distributions, which are proper probability distributions. In recent years, Bayesian statistics has been nurtured and developed by the Seminar on Bayesian Inference in Econometrics, which has been sponsored jointly by the National Bureau of Economic Research and the National Science Foundation. This small group under the leadership of Arnold Zellner has played a significant role in fostering new ideas in Bayesian theory and methodology. One activity they sponsor is the Savage dissertation award for the best doctoral dissertation in Bayesian inference in econometrics and another is the publication of several books devoted to new research. Among these are those edited by Aykac and Brumat (1977), Fienberg and Zellner (1975) and Zellner (1980).

We see Bayesian ideas have come a long way since Bayes, however with regard to statistical practice it will be some time before they are widely accepted for the routine processing of data. One reason Bayesian methodology will have to wait is because non-Bayesian ideas are firmly entrenched, but another is two controversial aspects of Bayes theorem. First is the subjective interpretation of probability, which is so often (but not always) used in Bayesian inference, and the other is the assessment of prior information.

Subjective Probability

The most widely accepted interpretation of probability is the limiting frequency of an event in infinite repetitions of an experiment. Most practicing statisticians use this interpretation and are skeptical of the other interpretations, such as the classical, logical, and subjective interpretations, which is one way to classify the various approaches to probability.

Bernoulli (1713) and DeMorgan (1847) advance the idea of subjective probability in an informal way. For instance DeMorgan says "By degree of probability we really mean, or ought to mean, degree of belief…. Probability then, refers to and implies belief, more or less, and belief is but another name for imperfect knowledge, or it may be, expresses the mind in a state of imperfect knowledge."

Formal theories of subjective probability were formulated by Ramsey (1926) and deFinetti (1937), but, according to Kyburg (1970), it was not until Savage (1954) that this interpretation began to have an impact on statisticians. A more recent formal presentation of subjective probability is given by DeGroot (1970) who axiomatically develops utility and probability.

Why is the Bayesian interested in the subjective interpretation of probability? We saw that in Bayes' original paper, the parameter θ was given a frequency interpretation, because he constructed it that way using a ball-tossing experiment, but in many applications it is difficult to not use a subjective interpretation (non-frequency) for the parameters θ of the probability model. For example, suppose θ is the variance component of a one-way random model, then treating θ as a random variable with values occurring as some experiment is repeated, is for some, to go too far. Barnett (1982), in his book on comparative inference, refers to this idea (giving θ a frequency interpretation) as a super experiment. Thus, in many applications, it appears unreasonable to give θ a frequency interpretation, and one must rely on his subjective interpretation that the probability distribution of θ expresses one's degree-of-belief about the values of the parameter.

If θ is the probability of a coin landing "heads" on one toss, and I believe every possible value of θ is equally likely, then I would use

$$\xi(\theta) = 1, \quad 0 \leqslant \theta \leqslant 1$$

as my density for θ, but if I believed the coin is "almost" unbiased, I would use

$$\xi(\theta) \propto \theta^{10}(1-\theta)^{10}, \quad 0 \leqslant \theta \leqslant 1$$

to express my belief. Other people would probably use other densities to express their beliefs about the values of θ, but the important thing to observe is that it is possible to express one's degree-of-belief about θ via probability density functions, and in doing so, one is employing a subjective notion of probability. Of course, it is one thing to express one's degree-of-belief about the parameter θ of a binomial experiment and quite another to do it for the variance covariance matrix of a multinormal population.

In the formal theory of subjective probability one must be coherent in the following sense: the individual must not exhibit irrational behavior when assigning his subjective probability to events.

He must follow certain rules so that it is not possible to contradict the usual rules of probability. The axioms of subjective probability begin with a

Printed by: dimitar.gueorguiev@nike.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

He must follow certain rules so that it is not possible to contradict the usual rules of probability. The axioms of subjective probability begin with a preference relation ("at least as likely as," see DeGroot, 1970) between the events and one must be consistent with his preferences. For example, if I prefer event A to event B, and prefer B to event C, then I must prefer A to C. If the axioms of subjective probability are adopted, one will assign probabilities to the events in such a way that the Kolmogorov axioms are satisfied.

Subjective probability is unavoidable, or so it seems, with the Bayesian approach to statistics, and in the remainder of this chapter, we will see how to deal with it when the ideas of prior probability are explained. For an interesting and highly critical account of the frequency interpretation of probability the reader should read chapter VII of Jeffreys (1961).

The Basic Components of Inference

Of the many definitions of statistics, I like Barnett's (1982), who says statistics is "the study of how information should be employed to reflect on, and give guidance for action in, a practical situation involving uncertainty." For this book, if one eliminates "and give guidance for action in" in Barnett's definition, the resulting statement is what this book is about, because this book considers inference (not decision making) and inference is an act of reflection about the parameters of the probability model, among other things.

When one does Bayesian statistical inference, one is using prior information and sample data in order to find the values of the parameters in the model which generated the data.

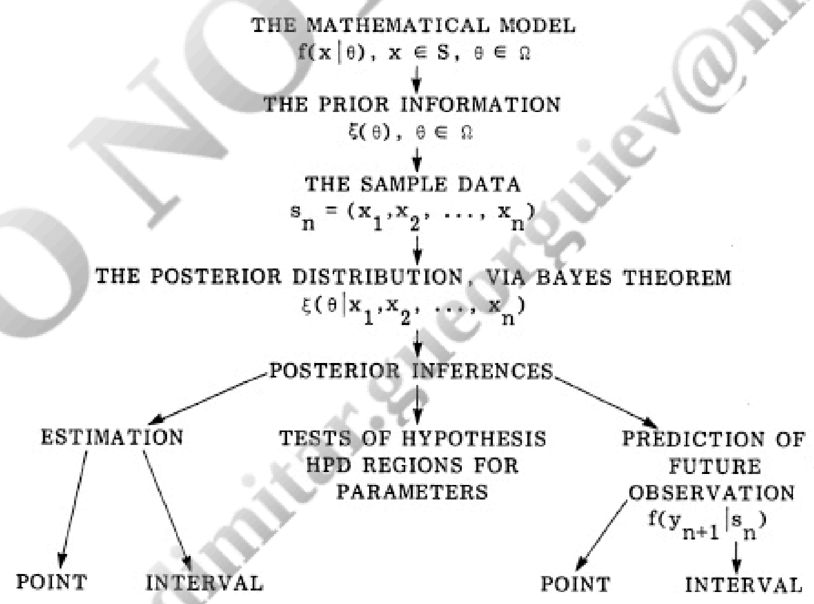
The components of statistical inference consist of the prior information, the sample data, calculation of the posterior density of the parameters and sometimes calculation of the predictive distribution of future observations. The prior information is expressed by a probability density $\xi(\theta)$, $\theta \in \Omega$ of the parameters θ of the probability model $f(x|e)$, $\theta \in \Omega$, $X \in S$, where f is a density of a random variable x and S is the support of f. The information in the data $x_1, x_2, \ldots, x_n$, where $x_1, x_2, \ldots, x_n$ is a random sample from a population with density f is contained in the likelihood function $\mathcal{L}(\theta | x_1, x_2, \ldots, x_n)$, $\theta \in \Omega$, which is the joint density of the sample data, then this is combined with the prior density of θ, by Bayes theorem, and gives the posterior density of θ. Of course, these ideas were shown in Chapter 1, thus one may describe an inference problem in terms of $(S, \Omega, \xi(\theta), f(x|\theta))$ and this problem is solved once the posterior density

$$\xi(\theta|x_1,x_2,\ldots,x_n) \propto \mathcal{L}(\theta|x_1,x_2,\ldots,x_n)\xi(\theta), \quad \theta \in \Omega$$

is calculated.

From the posterior density one may make inferences for θ by examining the posterior density. Some prefer to give estimates of θ, either point or interval estimates which are computed from the posterior distribution. If θ is one-dimensional, a plot of its posterior density tells one the story about θ, but if θ is multidimensional one must be able to isolate those components of θ one is interested in.

The following diagram illustrates the approach to statistical inference which will be adopted in this book:



Thus one begins the investigation with a probability model, which tells one how the observations s_n will be distributed (for each value of the parameter), and one's prior (before the data) information is expressed with the density $\xi(\theta)$, and after s_n is observed, the posterior density of θ can be calculated, via Bayes' theorem. The foundation of Bayesian inference is the posterior distribution of θ, because all other aspects of inference depend on this distribution. Perhaps one is more interested in estimation, in which case one must first decide what components of θ are of primary importance and the nuisance parameters eliminated. On the other hand if testing hypotheses is one's main objective, then one must decide how to conduct the test, i.e., to use an HPD region or a posterior odds ratio. Forecasting future observations will be done with the Bayesian predictive density $f(y_{n+1} | s_n)$, where y_{n+1}, is a future observation, $y_{n+1} \in S$, and one may provide either point or interval (region) predictions.

It is obvious this is an idealization of the actual way in which one would carry out a statistical analysis. For example, how does one adopt a probability model (say an AR model for a time series) for the observations, and how does one assess the prior information which is available for the parameters? These and other questions are to be examined in the following sections.

One difficulty with advocating such an ideal system of inference is that it ignores the way scientific learning progresses. Box (1980) describes, by means of a Bayesian model, the iterative process of criticism and estimation for building statistical models, and he uses the predictive distribution as a diagnostic check on the probability model. The diagram of inference, presented here, ignores that important aspect of model building, that is, it is diagnostic check on the probability model. The diagram of inference, presented here, ignores that important aspect of model building, that is, it is assumed the probability model is “correct” and is not to be changed during the analysis.

Box’s diagnosis should be carried out in practice but will not be adopted here, because it would carry us too far afield from our main goal, which is to demonstrate the technical mechanics of making posterior inferences. In any case, one will benefit from reading the Box paper.