

UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

A BAYESIAN INDIFFERENCE PROCEDURE

by

Melvin R. Novick

and

W. J. Hall

January 1963

This research was supported primarily by the Office of Naval Research under contract No. Nonr-855(09) for research in probability and statistics at the University of North Carolina, Chapel Hill, N. C. Reproduction in whole or in part is permitted for any purpose of the United States Government. Supplemental support was received from the National Science Foundation, under Grant G-5824 and from the National Institutes of Health under Grant GM-10397-01.

Institute of Statistics
Mimeo Series No. 345

TABLE OF CONTENTS

		Page
CHAPTER I	INTRODUCTION	1
CHAPTER II	THE PARAMETRIC FORMULATION	4
CHAPTER III	DEFINITIONS OF PROBABILITY	14
	3.1 Semantics and Syntactics	14
	3.2 Frequency Probability	15
	3.3 Logical Probability	17
	3.4 Personal Probability	18
	3.5 Fiducial Probability	20
CHAPTER IV	AN INDIFFERENCE PROCEDURE	24
	4.1 Indifference, Invariance and Improper Densities	24
	4.2 Natural Conjugate Bayes Densities	27
	4.3 Minimum Prior Data	31
	4.4 Minimum Bayes Information	37
	4.5 Minimum Necessary Sample	42
	4.6 Minimum Fiducial Information	45
	4.7 The Natural Bayesian Structuring	49
CHAPTER V	NORMAL MODELS	53
	5.1 Univariate Normal Model - Location Parameter Unknown	53
	5.2 Univariate Normal - Scale Parameter Unknown	55

	Page
5.3 Univariate Normal - Location and Scale Parameters Unknown	58
5.4 Bivariate Normal Model - Regression Parameter Unknown	60
5.5 Bivariate Normal Model - Three Parameters Unknown	61
5.6 Bivariate Normal Model - Five Parameters Unknown	63
5.7 A Log - Normal Model	65
CHAPTER VI POISSON MODELS	67
6.1 Random Poisson Sampling	67
6.2 Poisson Process Model	68
6.3 Discussion of the Indifference Rule for the Poisson Model.	69
CHAPTER VII MULTINOMIAL MODELS	71
7.1 Binomial Models	71
7.2 Trinomial Models	75
7.3 The General Multinomial Model	77
7.4 Validation and Semantic Interpretation of the Multinomial Models	79
CHAPTER VIII UNIFORM AND RELATED MODELS	82
8.1 Uniform Model - Scale Parameter Unknown	82
8.2 Uniform Model - Location Parameter Unknown	84

	Page	
8.3	Uniform Model - Location and Scale	
	Parameter Unknown	85
8.4	A Positive Exponential Model	87
CHAPTER IX	NEGATIVE EXPONENTIAL AND RELATED MODELS	89
9.1	Negative Exponential Model - Scale	
	Parameter Unknown	89
9.2	Negative Exponential Model - Location	
	Parameter Unknown	90
9.3	Negative Exponential Model - Location and	
	Scale Parameters Unknown	90
9.4	A Pareto Model	92
9.5	A Weibull Model	93
CHAPTER X	LIMITATIONS, EXTENSIONS AND REFINEMENTS	95
10.1	Limitations of the Procedure.	95
10.2	Evaluation of the Four Principles	97
10.3	Transformations of the Random Variables	97
10.4	Fiducial Inference	99
	REFERENCES	101

ACKNOWLEDGEMENTS

The authors wish to acknowledge the assistance of Professors W. L. Smith, W. Hoeffding, R. D. Bock, and L. V. Jones in offering suggestions which contributed to the work in this paper. Critical comments by Professors L. J. Savage and D. V. Lindley on the preliminary report of this research were very helpful.

Mrs. Doris Gardner handled the typing assignment in a most expeditious and efficient manner. Miss Martha Jordan assisted in the administrative details associated with the project.

CHAPTER I

INTRODUCTION

In 1763 there appeared in the Philosophical Transactions of the Royal Society a paper entitled, "An essay towards solving a problem in the doctrine of chances" (Bayes 1763). This paper, the work of the Reverend Thomas Bayes, had been transmitted to the Royal Society, following the author's death, by the author's friend, Richard Price. It has been suggested that Bayes failed to publish this paper because of his fear that the postulate which was the basis of his work "would be thought disputable by a critical reader" (Fisher 1956). This postulate was however later accepted, without question, by Laplace and occupied a central position in his Théorie Analytique des Probabilités (Laplace 1820). Later generations have found the axiomatic nature of Bayes postulate to be, at best, questionable. Beginning, perhaps, with Boole (1854) there appeared numerous criticisms of the Bayes postulate and later numerous attempts to circumvent or modify it. Few problems have attracted the efforts of a more distinguished roster of statisticians, and certainly none has given rise to more acrimonious debate.

The Bayes postulate was stated for use in what has come to be termed the theory of inverse probability, i.e. the theory which deals

with the problem of assigning probabilities to states of nature or causes after having observed experimental results or effects (Fisher, 1930). This theory makes use of Bayes theorem which, in the form given by Jeffreys (1961), is

$$\text{Posterior probability} \propto \text{Prior probability} \times \text{Likelihood} \quad (1.1)$$

Bayes theorem is an immediate consequence of the definition of conditional probability and hence is not subject to question. But a posterior probability can be obtained from (1.1) only if both the likelihood and the prior probability are assumed known. The Bayes postulate (or Bayes-Laplace indifference rule) specifies that when it is assumed that before the observational data are available, nothing is known about the true state of nature, a uniform prior distribution over the possible states of nature should be assumed. This specification yields an apparent symmetry in that it would seem that all possible states of nature are being considered as "equally likely" a priori. The major difficulty arising from the Bayes postulate is that when the states of nature are indexed by a continuously variable parameter, it is not invariant under reversible transformation of the parameter space.

This paper presents an approach to the problem of specifying prior distributions within a parametric structuring and under indifference conditions when the formulation is designed to serve as a model for scientific inference. Heavy reliance is placed upon the works of D. V. Lindley, of H. Jeffreys, of H. Raiffa and R. Schlaifer,

and particularly of R. A. Fisher. An attempt is made to modify and expand principles and proposals taken from these works and to integrate them with other principles to furnish an acceptable procedure for specifying prior distributions under indifference for a wide range of statistical models. The procedure suggested in this paper appears to overcome the lack of invariance exhibited by Bayes postulate and to satisfy some heuristic criteria which may be reasonably demanded of an indifference rule. The question of the interpretation of mathematical statistical statements in applications is considered, and a dual mode of interpretation of the Bayesian model is proposed.

The indifference procedure is applied to one and two parameter normal models, uniform and exponential models, Poisson and multinomial models, normal regression models, and several other models. It is not proposed or supposed that this paper contains any final statement on the general problem of indifference specifications. It is suggested, however, that the proposed procedure yields reasonable specifications in the relatively wide range of cases considered and may be capable of further extension and refinement.

CHAPTER II

THE PARAMETRIC FORMULATION

In this chapter we introduce and discuss the parametric formulation of the statistical inference model and contrast the classical and Bayesian approaches to statistical inference. We associate the real- or vector-valued random variable X , taking values x in the Euclidean space \mathcal{X} , with the possible outcomes of an experiment performed on experimental units drawn from a defined population. We assume that a distribution function $M(x)$ and a corresponding density function $m(x)$ (with respect to a fixed measure defined on the Borel sets of \mathcal{X}) may serve as an adequate probabilistic representation of the relative frequencies of the possible outcomes of the experiment. When M is known, no statistical problem exists, as the probability of any collection of outcomes may be calculated. When M is unknown, a statistical problem exists and may be approached in various ways.

An important consideration is the character of the assumptions which the statistician is willing to make about M . A common assumption is that M belongs to a class \mathcal{M} of distribution functions and that these distributions may be indexed by a real- or vector-valued parameter $\theta \in \Theta$. By this is meant only that there exists a one-to-one correspondence between the points of the parameter space and the distributions in \mathcal{M} . The parametrization of a class of distribu-

tions, however, is not unique. Any reversible transformation on \mathcal{H} may be taken as an indexing parameter for \mathcal{M} . The choice of a particular parametrization has been subject to convention and convenience but not to logical specification.

For example, let \mathcal{M} be the class of normal distributions with density functions indexed by the parameter σ , and written as

$$m_{\sigma}(x) = \sigma^{-1}(2\pi)^{-1/2} \exp(-x^2/2\sigma^2), \quad -\infty < x < \infty, \quad 0 < \sigma < \infty, \quad (2.1)$$

In this particular formulation the chosen parameter σ corresponds to the standard deviation of the distribution. \mathcal{M} could, however, have been indexed by $\alpha = \sigma^2$, the parameter corresponding to the variance, by $\beta = \sigma^{-2}$, the parameter corresponding to the precision, or by $\lambda = -\log \sigma/\sqrt{2\pi e}$, the parameter corresponding to the (Shannon) information of the distribution. Often a parameter is chosen to correspond with a moment or other important characteristic of the distribution, or otherwise simply for convenience. A parameter usually does indicate some kind of ordering of the densities, however, and therefore it is reasonable to consider only a somewhat restricted class of possible transformations. We shall assume that there is some natural parameter $\theta = (\theta_1, \theta_2, \dots, \theta_r)$ and shall restrict ourselves to monotone and differentiable transformations $\lambda_i = \varphi(\theta_i)$ such that $\varphi'(\theta_i)$ may be expressed as the finite sum $\sum a_j \theta_i^{b_j}$ where a_j and b_j are real numbers. This last requirement will not be necessary for some examples. For simplicity of development we shall further assume that \mathcal{H} is an n -dimensional interval. Finally, we assume that the scale of measurement of the random variable is fixed up to a linear transformation.

The specification of the class \mathcal{M} involves an assumption about M which presumably is made on the basis of prior knowledge. This knowledge may have been gained in a variety of ways. We may have sufficient knowledge concerning the physical process generating the random variable to completely specify \mathcal{M} . For example, the assumptions of the Poisson process may be valid and hence \mathcal{M} may be the class of Poisson or negative exponential distributions, depending on the nature of X . We may have considerable experience with similar experiments or processes and hence may find it useful to use this information as a basis for specifying \mathcal{M} by analogy. Finally \mathcal{M} may be obtained by empirical curve-fitting based on previous observations of X .

The specification of the population, the experiment, the sampling rule, the random variable, the restriction of M to a particular class \mathcal{M} , the specification of the dominating measure and associated metric, and the specification of the parameter element θ will be referred to as the parametric structuring of the statistical inference problem. For purposes of this paper sampling will be restricted to simple random sampling, and the dominating measure will always be either Lebesgue or counting measure. The distribution function M_θ will be called the model distribution, and the corresponding density m_θ will be called the model density. Thus, the model density will either be the usual probability density function of a continuous random variable or the usual frequency function of a discrete random variable.

The assumptions used to specify the model are not, within the inference procedures to be discussed, subject to modification on the basis of the data. Ideally these assumptions would represent only certain and exact knowledge. The statistician, in employing a model, is using a necessary, convenient and useful fiction. He must compromise between accuracy and convenience in choosing a model and then subject it to continuing re-evaluation based upon inference drawn from outside the model. Thus all inferences made using these procedures are conditional inferences, conditional upon the accuracy of the model.

On the basis of the information available to him, the experimenter will design an experiment and specify measurements to define a random variable. At times either the experiment chosen or the measurements specified will not be such as to utilize all of the prior certain knowledge. For example, the experimenter may know that the process under study could yield measurements which would be very nearly normally distributed. However, for convenience or by necessity, he may choose to limit himself to measuring whether a resultant value is greater than or less than some fixed value, that is, he specifies a binomial model instead of a normal model.

Alternatively, a normal model may be appropriate, but all observations outside certain limits are truncated (or censored) so that a truncated normal model is in fact appropriate for the experiment as actually performed. A third example is provided by a multinomial experiment in which draws are made from an urn containing balls of

several colors, but a record is made only as to whether or not they are black; or sampling from the urn may be inverse rather than direct, inducing a negative binomial rather than a binomial model.

The term experimental model may be used to refer to the model appropriate to the experiment as performed and measurements as taken. The term informational model may be used to refer to the model appropriate to an experiment and resultant measurements which are such as to incorporate all prior certain knowledge. Fraser (1962) has also expressed this kind of idea in a discussion of a paper by A. Birnbaum. For the purposes of this paper it is assumed that the experimental model is identical with the informational model. More generally it would seem necessary that the Bayesian parametric structuring be determined with reference to the informational model when the experimental and informational models differ. The technical problems involved in this case will not be studied here.

The classical approach to the statistical inference problem in its parametric formulation has been to consider the primary statistical problem to be that of drawing inferences about θ , without recourse to prior distributions or considerations of actions or attendant losses. The inferences are usually expressed in the form of point estimates, interval estimates, or tests of hypotheses. For convenience, the term classical approach will be used to denote the theory employing just these and related techniques. It is thus distinguished from both the decision-theoretic approach and the Bayesian approaches.

From a purely mathematical point of view, the theory underlying the classical approach seems unimpeachable. Yet there has been a continuing and particularly vocal criticism of these techniques by those who find a discrepancy between the valid interpretations of classical theory and the requirements of scientific reporting or individual decision making. A general agreement with a number of these criticisms underlies a motivation for the research reported here; however, neither an appraisal nor a review of all of these criticisms is attempted. The major advantages and disadvantages of the Bayesian approach as compared to the classical approach will be sketched, though not in detail. No further reference will be made to the decision-theory approach as it is not primarily designed for purposes of scientific inference (Lindley, 1950).

In the Bayesian approach the parametric formulation is as in the classical approach. In addition, it is assumed that there is specified, a priori, a probability measure on the Borel sets of (\mathcal{H}) , and that the model density $m_{\theta}(x)$ is regarded as the conditional density of X given θ . The dominating measure is usually, and here, assumed to be Lebesgue measure and the corresponding density function will be denoted by $b(\theta)$. The joint density of (X, θ) , of continuous or mixed type, is thus given by $m_{\theta}(x) b(\theta)$. (We do not distinguish the random variable θ and its values.)

The joint density function of n independent observations, as a function of the vector of observations for fixed θ , is denoted by

$$m_{\theta,n}(\underline{x}) = \prod_{i=1}^n m_{\theta}(x_i), \quad n = 1, 2, 3, \dots \quad (2.2)$$

This same expression, as a function of θ for fixed \underline{x} , is called the likelihood function. The prior unconditional density of a single observation, to be called the (prior) fiducial density of X , is defined by

$$f(x) = \int m_{\theta}(x) b(\theta) d\theta \quad (2.3)$$

Given a vector of n observations \underline{x} , the posterior density of θ given \underline{x} , is given by

$$b_n(\theta) = \frac{m_{\theta,n}(\underline{x}) b(\theta)}{\int m_{\theta,n}(\underline{x}) b(\theta) d\theta}, \quad n = 1, 2, 3, \dots, \quad (2.4)$$

(where \underline{x} has been suppressed in the notation $b_n(\theta)$). This formula is known as Bayes theorem. The posterior fiducial (unconditional) density of X is given by

$$f_n(x) = \int m_{\theta}(x) b_n(\theta) d\theta, \quad n = 1, 2, 3, \dots \quad (2.5)$$

The definition of the random variable X and the selection of the indexing element θ are often such as to restrict the domains of X and θ to some subset of their respective spaces. The region \mathcal{X} (or \mathcal{H}) of the definition of X (or θ) will be referred to as the spectrum of X (or θ). The spectrum is thus the region of (possibly) positive density. The sequence $\{b_n(\theta)\}$, $n = 0, 1, 2, \dots$, $b_0(\theta) = b(\theta)$, will be called the sequence of Bayes densities. The sequence $\{f_n(x)\}$, $n = 0, 1, 2, \dots$, $f_0(x) = f(x)$, will be called the sequence of fiducial densities. The letter b will be used to denote

any Bayes density, m to denote any model density, and f to denote any fiducial density without implying the identity of densities thus denoted.

Given a prior Bayes density $b(\theta)$, it is possible to combine the implied evaluation of prior information with the obtained experimental results to obtain posterior evaluations of θ (or X) through the posterior Bayes density (or fiducial density). Point or interval estimates of any parameters or tests of simple or composite hypotheses about any parameters may thus be obtained by an appropriate integration. Such posterior evaluation is, under very general conditions, consistent, in that the posterior Bayes distribution approaches, with probability one (M), a point distribution on the true parameter value and in that $f_n(x)$ approaches, with probability one (M), the true density $m(x)$. An important necessary condition is that $b(\theta)$ be positive, almost everywhere, throughout (H) (that is, that Lebesgue measure be dominated by the prior Bayes measure). Prior Bayes densities having this property will be called adaptive. This requirement corresponds to Jeffreys' convention that no possibility be excluded a priori. A general discussion of the question of consistency may be found in Le Cam (1958); however, direct verification of such consistency is not difficult for any of the examples treated here.

A second desirable property of Bayes procedures is that they depend on the sample only through a sufficient statistic. By the Neyman factorization theorem we have

$$m_{\theta,n}(\underline{x}) = h_n(\underline{x}) k_{\theta,n}(s)$$

where $s = s(x) \in S_n$ (the spectrum of the sufficient statistic) is sufficient for θ and $h_n(\underline{x})$ does not depend on θ . We assume the existence of a sufficient statistic of fixed dimensionability for n sufficiently large. Then the factor $h_n(\underline{x})$ cancels out in (2.4) to yield

$$b_n(\theta) = \frac{k_{\theta,n}(s) b(\theta)}{\int k_{\theta,n}(s) b(\theta) d\theta} \quad (2.7)$$

The function $k_{\theta,n}(s)$ is called a kernel of the likelihood (Raiffa and Schlaifer (1961)).

A third desirable property of Bayes procedures is that they depend only on the joint likelihood. Thus, given a prior Bayes density $b(\theta)$, and a likelihood function $m_{\theta,n_1}(\underline{x})$, the resultant posterior Bayes density $b_{n_1}(\theta)$, if used as a prior Bayes density with a new likelihood function $m_{\theta,n_2}(\underline{x})$, will yield the same final posterior Bayes density as if the data yielding the two separate likelihood functions were pooled to yield a single likelihood and combined with $b(\theta)$ to yield a posterior Bayes density. More generally the posterior Bayes density does not depend upon the order in which the data were received, the sampling or sequential stopping rule employed, or considerations of what results might have been obtained from the experiment, but only upon the experimental model and the actual results obtained, as summarized in the likelihood kernel. (This is in keeping with the likelihood principle, e.g., Birnbaum, (1962).)

The problem of multiple comparisons does not arise in a Bayesian analysis. Since the joint posterior Bayes density of all parameters is available, the statistician is free to transform this joint density to one on any other set of parameters more appropriate to the statements he may wish to make. A suitable region in this new space may then be defined so that the integral of $b_n(\theta)$ over this region is not less than some fixed level. Individual statements can then be made marginally about each of the parameters involved with total probability as given by the joint statement.

Finally, and most importantly, the Bayesian approach permits direct probability statements about any parameter of interest. The primary criticism of classical theory is that it does not permit direct probability statements about θ but rather requires inferences about θ based to a large extent upon a logic which Fisher (1956) points out to be a simple disjunction. For example in classical hypothesis testing if the data fall in the rejection region, having small probability under the null hypothesis, either a rare event has occurred or the null hypothesis is not true. This constitutes a disjunction, and hence we infer that the null hypothesis is not true. However, no direct statement about θ is made nor is one possible, and even after the conclusion of the disjunction is asserted, nothing is said about the relative credibility of various subsets of (H) . In the Bayesian approach the credibility of any hypothesis set can be determined. On the other hand the advantages of the Bayesian approach are all dependent upon the availability of a prior Bayes density.

CHAPTER III

DEFINITIONS OF PROBABILITY

3.1 Semantics and Syntactics

In this chapter we discuss the meaning of probability and introduce a dual mode of interpreting the Bayesian model. The Bayesian approach to the statistical inference problem has been subject to criticisms other than those involving the difficulties in specifying the a priori probabilities. These criticisms have not dealt with the mathematical validity of the theory, which seems as unimpeachable as the mathematical validity of the classical theory, but rather with its applicability. Thus, given the availability of a prior Bayes density, the question of relative applicability is essentially the sole basis for comparing the relative merits of the two theories. To clarify this point it is convenient to review some fundamental principles of mathematical logic and to introduce certain definitions. Mathematics proceeds by means of the axiomatic method by which a purely abstract theory is developed deductively on the basis of the specification of certain elementary entities, the assumption of certain relations among these entities and the acceptance and use of a formalism of logical deduction. The sole necessary requirement of a valid mathematical theory is that it be consistent in that no logical contradictions are derivable in it. It is in this sense that it has been

said that the mathematical validity of both the classical and Bayesian theories seem unimpeachable.

The axiomatic development of probability theory requires a definition of probability within the mathematical system. This will be termed the syntactic definition of probability. A syntactic definition of probability based upon the theory of measure was first proposed by Kolmogorov (1933) and advocated by Cramer (1946), Doob (1941), and others, and is now the generally accepted axiomatic formulation. Within this theory, probability is simply a normed measure defined on a fixed σ -algebra of sets.

The use of any abstract logical system in the study of the physical world requires the establishment of a correspondence between the elementary units of the abstract theory and the physical elements under study. The axioms of the abstract theory must further be chosen so as to conform with known simple behavior of the physical elements. The establishment of this correspondence between axiomatic probability theory and the physical world involves the specification of what will here be termed the semantic definition of probability. Classical theory has been associated with what is known as the relative frequency (semantic) definition of probability. Bayesian theory has been associated with essentially two semantic definitions.

3.2 Frequency Probability

The frequentist view of the application of the probability calculus has been dominant throughout this century. Indeed the feeling had been so intense that this was the only legitimate application of probability theory that it is only within the last several years that any

widespread effort has been made to employ techniques applicable to any other definition. Two unfortunate results ensued from this preoccupation with frequency interpretation. The first was the confusion between syntactic and semantic probability which led to the attempt to give a syntactic definition of probability in terms of relative frequency. The second was that the number of new developments in the Bayesian syntax since the time of Laplace has been severely limited. Important contributions have been made, particularly by Jeffreys and more recently by a number of writers, but the work of a few men could not match that of an army of frequentists. In confronting an applied inference problem today the statistician may be forced to employ a classical procedure whether or not he considers it really appropriate because the corresponding Bayesian procedure has not been derived.

The frequentist theory requires that every probability be interpretable as the limit of a relative frequency. If it is said that the probability is p that a ball drawn from a hypothetical urn containing an infinite number of balls will be red, then this must mean only that the observed proportion of red balls from an arbitrarily large sample of balls must almost always be arbitrarily close to p . No other interpretation is permitted. For detailed exposition of this theory one may refer to Carnap (1945), Von Mises (1939), and Reichenbach (1935). For interesting evaluations of its limitations in statistical inference one may read Savage (1954), Raiffa and Schlaifer (1961), Birnbaum (1962), and the discussions by Good following the last paper.

3.3 Logical Probability

The first semantic definition applicable to Bayesian theory has been termed logical probability and may be associated with the works of Bayes (1763), Laplace (1820) and Jeffreys (1961). The second has been termed personal probability and may be associated with the works of Ramsey (1926), De Finetti (1937), Good (1950), Savage (1954) and Schlaifer (1959). Comprehensive explications of these semantic definitions may be found in the works of Jeffreys and Savage. A short but stimulating survey of semantic definitions, together with a vigorous discussion of the axiomatic method, may be found in Good (1950).

In simplest terms logical probability involves the making of numerical statements concerning the relative credibility of hypotheses (parameters), given all prior and experimental data. A satisfactory explication of logical probability is dependent upon the availability of a satisfactory Bayesian postulate. A proposed theory of logical probability is tested by verifying its internal consistency and by verifying that simple derivations from the Bayesian syntax lead to intuitively plausible interpretations. Having validated these simple derivations we would then be inclined to believe that more complex derivations would yield meaningful and valid results, though in these cases our intuitions may not be sufficiently schooled to judge. It is because of this limitation on the scope of our intuition that we require a theory. For this reason each mathematical result obtained requires semantic validation. A single important example which

grossly offended our intuition would be sufficient to question the generality of a theory of logical probability provided we were convinced that our intuitions could not be enlightened to eliminate the offence. Fortunately, the ease with which intuition can adapt to interpret and explain mathematical derivations is remarkable.

3.4 Personal Probability

Personal Probability has to do with the risk-taking behavior of an idealized rational person. If it is said that your personal probability that a ball drawn as in 3.2 will be red is p , then this means that $p/(1-p)$ is the odds that you would "barely be willing to offer" for a red ball against another ball (Savage, 1962).

A convenient way of classifying a Bayesian theory as pertaining to logical probability or personal probability may, following Savage, be based upon the theory's proposed method of specifying the prior Bayes density. If the specification is based upon the experimental or introspective evaluation of a person's a priori beliefs concerning the relative credibility of the available hypotheses, then the theory will be considered to be a theory of personal probability. If the specification is based upon consideration of symmetry or indifference in the spirit of Bayes postulate and modifications based upon prior data, then the theory will be considered to be a theory of logical probability.

The problem studied in this paper is that of specifying the prior Bayes density under indifference when it is supposed that the resultant statistical model is to be used for the evaluation of scientific data. It is important to note that we are dealing with a problem

in the area of statistical inference theory as opposed to statistical decision theory (Lindley, 1956). While it is possible that some ultimate action may be taken on the basis of the inferences drawn, it would be desirable that these inferences be made independently of any considerations of possible actions or resultant losses. The proposition (Savage, 1962) that a person faced with a choice of actions must rationally behave "as if" he had some prior distribution and some loss function seems reasonable, but it seems equally reasonable that, for the problem discussed in this paper, the prior distribution must be free of personal bias.

A second consideration precludes the adoption of the personal approach. Often it is desirable to assume no prior knowledge, not because there is none, but because there is too much. Several theories may have been proposed and several sets of contradictory data may be available, each supporting a different theory. In this context it may be convenient to disregard all prior data, in which case the indifference requirement is clearly one of symmetry with respect to the possible theories. This consideration really only underscores the idea that scientific inference involves group inference; hence prior Bayes distributions cannot properly be determined on the basis of personal bias, and thus, for the purposes of scientific inference the requirement is for a theory of logical probability with respect to statements about parameters.

We have purposely eschewed the use of the commonly used terms subjective and objective probability. Bayes methods, whether serving

as a model for logical probability or for personal probability, may be either subjective or objective depending upon how the prior Bayes density is determined. In the case of personal probability, if the prior Bayes density is determined by the introspective evaluation by the person concerned of his beliefs, then the method may rightly be called subjective. If however his beliefs are obtained experimentally through a measurement of his behavior, then the method may rightly be called objective. In the case of logical probability if specific prior data are combined with a satisfactory Bayesian postulate to produce a prior Bayes density we may term the method objective. If some "general consensus of opinion" concerning current available data about the parameter is used to modify a prior density obtained from a satisfactory Bayesian postulate we may term the method subjective. The term subjective probability is taken (by some writers) to be synonymous with the term personal probability

3.5 Fiducial Probability

Classical theory and current Bayesian theories deal primarily with techniques for making inferences about θ . But θ is, generally, nothing more than an arbitrary indexing element of the class \mathcal{M} , which may happen to correspond to an important characteristic of the distribution, and as indicated in example (2.1) the correct choice of parameter is often not clear. In light of the structuring defined above, it would seem that the more fundamental subject of interest is the random variable X or values thereof in future repetitions of the experiment. This concept was introduced by Novick (1962), and later

and independently a similar idea was suggested by Fraser (1962).

To understand nature is to predict successfully, and the proof of understanding is valid prediction. We therefore submit that a realistic prime object of inference is to provide a prediction model (which we shall term fiducial) to be substituted for the true but unknown model specified by M . In line with the classical approach we may describe this approach as one of estimating the true model density function $m(x)$ by the fiducial density function $f_n(x)$. The point is not whether it is possible, by use of classical or Bayesian techniques, to make statements about X , or to estimate $m(x)$, but rather whether the assertion of these statements is considered to be of first importance in formulating the inference procedure. In the Bayesian approach we propose that any indifference principle be consistent with this concept. The technique for utilizing this idea will be presented in Chapter IV. Of course, fiducial prediction does not preclude Bayesian logical probability inferences about parameters.

The use of the posterior marginal (fiducial) density of X first appeared as the law of succession, Laplace (1820), and has played an important part historically in the evaluation of proposed indifference rules. The idea of asserting statements about θ and statements about X in a completely symmetric manner was proposed by Fisher (1933) in his discussion of the fiducial distribution of the $(n+1)$ -st observation. Our choice of the term fiducial for this density is based upon our interpretation of this neglected aspect of Fisher's work.

Although fiducial statements may be considered as logical proba-

bility statements about the $(n+1)$ -st observation, they are capable of certain direct frequency interpretations. If after each observation from one fixed urn we use the fiducial density to predict the next observation, for example by predicting that the next observation will fall in the central $(1-\alpha)$ -interval of the fiducial density, then the continuation of this process will lead to a long run relative frequency equal to $(1-\alpha)$ of $(n+1)$ -st observations falling in the predicted $(1-\alpha)$ -interval, provided the conditions for consistency are satisfied.

More importantly, at least one example exists in which an exact (not long run) frequency interpretation of fiducial statements is possible. Suppose we had available a large number of urns from which we could draw objects having associated with them a random variable which was normally distributed with unit variance but with unknown means, which might be different for the different urns. We then draw n observations from an urn, compute a central fiducial $(1-\alpha)$ -interval and then determine whether or not the $(n+1)$ -st observation from that urn falls within that interval. If the prior Bayes density of the mean μ for this urn is uniform (and hence improper - see Chapter IV), then the posterior fiducial density of X_{n+1} is normal with mean \bar{X}_n (sample mean of X_1, \dots, X_n) and variance $(n+1)/n$. By direct elimination methods we would find that the true distribution of $\bar{X}_n - X_{n+1}$ is normal with mean zero and variance $(n+1)/n$, independently of μ . Hence it follows that the probability that the $(n+1)$ -st observation falls in the central $(1-\alpha)$ -interval of the fiducial density determined by the first n observations is exactly $1-\alpha$, and hence the relative frequency

of occurrences of this event over all urns is exactly $1-\alpha$. We shall find that the principles proposed in this paper lead to the above prior Bayes density for μ .

Fraser (1960) has used a similar derivation for the one-parameter normal case to demonstrate a quite different frequency interpretation of fiducial probability. The basic difference between fiducial probability as proposed above and as considered by Fisher and Fraser is that the system proposed herein is avowedly Bayesian, whereas Fisher has denied that his fiducial probability is related to the Bayesian approach except by coincidence, and Lindley (1958) has verified the extent of this coincidence.

CHAPTER IV

AN INDIFFERENCE PROCEDURE

4.1 Indifference Invariance and Improper Densities

In this chapter we review, re-interpret and extend a number of principles from the Bayesian literature and integrate them into a Bayesian indifference procedure.

The Bayes postulate was first stated for a problem which involved a class of densities most naturally indexed by a parameter whose spectrum was a bounded interval. For this case a proper uniform density could be defined on this space and taken as the prior Bayes density. Such a density expresses a kind of indifference. Other problems, however, involved densities more naturally indexed by parameters having spectra of infinite extent, a simple example being the normal model

$$m_{\mu}(x) = (2\pi)^{-1/2} \exp -\left[\frac{(x-\mu)^2}{2}\right], -\infty < x < \infty, -\infty < \mu < \infty, \quad (4.1)$$

in which the parameter μ takes values in an infinite interval. A second case is exemplified by (2.1) in which the parameter σ takes values in a semi-infinite interval. In these cases it is not possible to define a proper uniform density over the parameter space; however, an improper uniform density may be defined instead.

We define an improper or unnormed density as any non-negative (measurable) function g whose integral does not converge. A constant c is an improper uniform density on any interval of infinite

extent. Since the posterior Bayes density will be independent of any constant multiplicative change in the prior density it is unimportant what positive value c takes. Indeed we shall always use the constant c to indicate an improper uniform density even when considering more than one such density and even when the constants cannot be the same.

Improper densities have long been used in Bayesian theory. Jeffreys extended Bayes postulate to parameter spaces of infinite extent by postulating that a parameter on $(-\infty, \infty)$ should have an improper uniform prior density and that a parameter θ on $(0, +\infty)$ should have the improper density θ^{-1} , or equivalently that $\lambda = \log \theta$ should have a uniform density on $(-\infty, \infty)$. The use of improper densities has not seemed unreasonable, particularly in that, if properly chosen, the posterior densities are proper after an appropriate number of observations. It is, for example, easily verified that with the improper uniform prior density for μ in (4.1) the posterior density of μ , given a single observation x , is proper. Similarly in the two parameter normal case, when the parameters (μ, λ) corresponding to the mean and the logarithm of the standard deviation are chosen, we find that a joint uniform density on (μ, λ) will lead to an improper joint density after one observation, but a proper conditional density of μ after one observation, and a proper joint density of (μ, λ) after two observations. This second example seems consistent with the idea that one observation yields no information about the degree of spread of a normal distribution with unknown mean.

Another motivation for considering improper densities is the desire to obtain closure for certain parametric classes of densities. Thus if $g_\sigma(y)$ is the normal density with zero mean and standard deviation σ ($0 < \sigma < \infty$), it is useful to extend the definition of the density to permit a limiting density as σ tends to infinity. Since the relative density at any two points, $g_\sigma(y_1)/g_\sigma(y_2)$, tends to a finite limit, namely unity, we say that the limiting (improper) density is defined as the (improper) uniform density. In general, if g_θ , $\theta \in \mathbb{H}$ is a proper density except at a boundary point θ_0 , but $g_\theta(y_1)/g_\theta(y_2)$ tends to a finite limit $r(y_1, y_2)$ as θ tends to θ_0 , then any improper density g for which $g(y_1)/g(y_2) = r(y_1, y_2)$ is called a limiting improper density. Similarly, if for some other value of $\theta' \notin \mathbb{H}$, $g_\theta(y_1)/g_\theta(y_2)$ is a finite function $r'(y_1, y_2)$, then any improper density g' for which $g'(y_1)/g'(y_2) = r'(y_1, y_2)$ is called an improper density corresponding to the point θ' . In this manner we may supply natural closures and extensions to parametric classes of proper densities. We shall later show how these extensions are uniquely determined. Even if we permit improper densities, the problem of invariance of Bayes postulate is hardly less crippling. Jeffreys' proposal that a parameter θ on $(0, +\infty)$ have the prior Bayes density θ^{-1} does remain invariant under power transformations on θ since, with $\omega = \theta^x$, ω also has prior Bayes density ω^{-1} for ω in $(0, \infty)$, but this is the extent of the invariance obtainable. Jeffreys (1948) also proposed that the initial probability density be taken as proportional to the square root of the determinant of the Fisherian information matrix. This rule however leads to contradictory specifica-

tions in some cases and the arguments employed by Jeffreys to resolve individual cases must be considered ad hoc. There does not appear to be any general way of specifying prior densities which will be invariant under any but severely limited classes of transformations. Thus it might seem that the problem of lack of invariance is insolvable.

An alternative to finding a prior density exhibiting invariance under transformation would be to determine which parameter should be subject to Bayes' or some analogous postulate. If some rationale could be discovered which demanded that μ and not μ^3 , say, in (4.1) should be uniformly distributed then the problem would indeed have been solved. Now it might be tempting to say that a uniform prior density should be taken for whatever parameter seems of special interest to the statistician (Lindley 1957, Kerridge 1961). But clearly this is not an acceptable basis for a theory of logical probability. Again reference must be made to the fact that scientific inference is a problem of group inference and just as the determination of the prior distribution cannot be subject to personal bias, it cannot be subject to personal interest. Basically a parameter is merely an indexing element of a class of densities. Hurzurbuzar (Jeffreys, 1961) has proposed a particular choice of parameter to be subjected to an indifference postulate in exponential class models, but little rationale has been given for this choice.

4.2 Natural Conjugate Bayes Densities

Raiffa and Schlaifer (1961) have furnished a rich formalism for the Bayesian analysis. For the purposes of this paper the essential

feature to be emphasized is the theory of natural conjugate Bayes densities (NCBD). The function $k_{\theta,n}(s)$, defined in 2.6, has been termed a kernel of the likelihood. Suppose that $\int_{\mathcal{S}_m} k_{\theta,m}(s') d\theta$ exists (finite) for $m = m_0 + 1, m_0 + 2, \dots$ and for $s' \in \mathcal{S}_m$, where \mathcal{S}_m is the space of the sufficient statistic s when the sample size is m . Suppose also that the definition of $k_{\theta,m}(s')$ can be extended in some natural way for all $m > m_0$, $s'_m \in \mathcal{S}_m^*$ where \mathcal{S}_m^* is the smallest interval containing $\bigcup_{i=1}^m \mathcal{S}_i$. We shall then term $k_{\theta,m}(s')$ a kernel of the NCBD defined by

$$b_m(\theta) = \frac{k_{\theta,m}(s'_m)}{\int k_{\theta,m}(s'_m) d\theta} \quad (4.2)$$

for all real $m > m_0$ and $s'_m \in \mathcal{S}_m^*$, and $b_m(\theta)$ will be called a natural conjugate Bayes density for M_θ . If $k_{\theta,m}(s'_m)$ is not integrable we will still consider it to be a kernel of an improper density. The existence or non-existence of this integral for various values of m generally depends upon the choice of parameter.

For the models studied in this paper there exists a natural selection of the sufficient statistic which completes the definition of (4.2) for $m > m_0$. We shall use the symbol x in the kernel of the likelihood and the symbol z as the analogous quantity in the kernel of the NCBD, similarly m will replace n and $\bar{z} = \sum z_i/m$ will replace $\bar{x} = \sum x_i/n$. The analogs of the sufficient statistics are thus the parameters of the NCBD. For convenience we shall consider the sample size n to be a component of the sufficient statistic, and hence m will be considered

to be one of the parameters of the NCBD.

To exhibit the general properties of the NCBD it is convenient to consider the example of the Bernoulli model. We have, as the joint likelihood of n observations,

$$k_{p,n}(\underline{x}) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}.$$

Then $k_{p,n}(s) = p^s (1-p)^{n-s}$ where $s = \sum x_i = 0, 1, \dots, n$. We then define the NCBD as

$$b_m(p) = \frac{p^z (1-p)^{m-z}}{\beta(z+1, m-z+1)} \quad (4.3)$$

for $0 \leq z \leq m$, $0 < p < 1$, $m \geq 0$. The extension of the definition for non-integral m and z seems natural enough. The density (4.3) is a beta density of the first kind with parameters z and $m-z$. If (4.3) is taken as the prior Bayes density then the posterior density of p given a sample of size n and x successes is

$$b_{m+n}(p) = \frac{p^{x+z} (1-p)^{(m+n)-(x+z)}}{\beta(x+z+1, \int m+n - \int x+z + 1)},$$

which again is a beta density, now with parameters $(x+z, \int m+n - \int x+z)$. The advantages of the use of a NCBD as a prior Bayes density when a NCBD exists becomes apparent from this example.

1. They are analytically tractable in that we can obtain a closed form expression for $f(x)$ and $b_n(\theta)$.

2. The kernel of the prior density combines with the kernel of the likelihood in exactly the same way that two sample likelihood kernels combine.
3. The NCBD class is closed under random sampling; hence the entire class of prior and posterior densities may be denoted by (4.2) for $m > m_0$ and $s'_m \in S_m^*$. The class of fiducial densities is similarly determined.

A more formal and extended discussion of these points may be found in Raiffa and Schlaifer (1961).

Let m_0 (an integer) be the smallest value such that b_m is defined for $m > m_0$. To demonstrate the technique for extending the NCBD class for non-negative integral values of $m \leq m_0$ we now consider a second example. For the two parameter normal model with parameters μ , the mean, and λ , the logarithm of the standard deviation, we shall obtain

$$b_m(\mu, \lambda) = \left(\frac{2 \int \frac{\sum z_i^2 - m \bar{z}^2}{2} \int^{\frac{m-1}{2}} e^{-(m-1)\lambda} e^{-e^{-2\lambda} \int \frac{\sum z_i^2 - m \bar{z}^2}{2}}}{\Gamma\left(\frac{m-1}{2}\right)} \right) \times \left(\frac{\int \frac{m e^{-2\lambda}}{\sqrt{2\pi}} e^{-e^{-2\lambda} \left(\frac{m}{2}\right) \int \mu - \bar{z}} \int^2}{\sqrt{2\pi}} \right)$$

for $m > 1$, $-\infty < \bar{z} (= \frac{\sum z_i}{m}) < \infty$, $0 < v = \left(\frac{m \sum z_i^2 - (\sum z_i)^2}{m} \right) < \infty$.

For $m = 1$, consistent with definitions to be presented in 4.3 we shall

have $\sum z_i^2 - m \bar{z}^2 = 0$, i.e. $\sum z_i^2 = (\sum z_i)^2$. Then $b_1(\mu_1, \lambda_1) /$

$$b_1(\mu_2, \lambda_2) = \frac{\int e^{-\lambda_1} / e^{-\lambda_2} \int e^{-e^{-2\lambda_1}(\frac{1}{2})} \int \mu_1 - \frac{z}{\lambda_1} dz}{\int e^{-e^{-2\lambda_2}(\frac{1}{2})} \int \mu_1 - \frac{z}{\lambda_2} dz} . \text{ Hence define}$$

$$b_1(\mu, \lambda) = \frac{\int e^{-2\lambda} e^{-e^{-2\lambda}(\frac{1}{2})} \int \mu_1 - \frac{z}{\lambda} dz}{\int e^{-2\lambda} e^{-e^{-2\lambda}(\frac{1}{2})} \int \mu_1 - \frac{z}{\lambda} dz} . \text{ Then } b_1(\lambda) = 1 \text{ and}$$

$b_1(\mu|\lambda) = b_1(\mu, \lambda)$. For $m = 0$ we have $b_0(\mu_1, \lambda_1) / b_0(\mu_2, \lambda_2) = 1$, hence define $b_0(\mu, \lambda) = c$. The function $b_m(\mu, \lambda)$, $0 < m < 1$, will not be defined as such definition is not required for the theory to be presented in this paper. The second factor in the expression for $b_m(\mu, \lambda)$ is the conditional density of μ given λ , for $m > 1$. We may extend its definition in a natural way to be made explicit later to values $0 < m < 1$.

The class of fiducial densities is defined by

$$f_m(x) = \int m_\theta(x) b_m(\theta) d\theta \quad \text{for } m = 0, 1, \dots, m_0 \text{ and } m > m_0 .$$

Thus in extending the domain of the NCBD class we automatically extend the domain of the class of fiducial densities.

4.3 Minimum Prior Data

Raiffa and Schlaifer do not explicitly propose a Bayesian postulate. Indeed, they would seem to indicate that they believe that a consistent one is not possible (Raiffa and Schlaifer, 1961, p. 66, Schlaifer 1959, p. 445). Yet the formalism which they have developed leads to a princi-

ple, which though it is not itself sufficient, will prove useful in developing a consistent Bayesian procedure. This principle is at least suggested by the choice of certain limiting operations taken (e.g. Raiffa and Schlaifer, 1961, p. 278, p. 292, p. 302, etc.)

The analytic tractability of certain classes of Bayes densities for certain models has long been known and the restriction of prior densities to these classes has long been advocated on this basis alone. The formalism of Raiffa and Schlaifer, however, permits a most fruitful heuristic justification. If we require that all prior densities be determined by prior data we might, for example, consider the quantities m and z in (4.3) to refer to some prior sample of size m with z prior successes. It would be natural then to see what happened as $m \rightarrow 0$ and $z \rightarrow 0$, with the restriction $z \leq m$, i.e. to see what happened as the prior data went to naught. We obtain

$$\lim_{\substack{m \rightarrow 0 \\ z \rightarrow 0 \\ z \leq m}} \frac{p^z (1-p)^{m-z}}{\beta(z+1, m-z+1)} = 1, \quad (4.4)$$

i.e. the uniform density, which agrees with Bayes postulate.

This type of procedure is quite generally applicable to natural conjugate Bayes densities, though we shall in some cases consider different limiting values than those taken by Raiffa and Schlaifer. The choice of limits taken and the definition of the limiting values will always be such as to be consistent with the principle that the kernel of the Bayes density was determined by prior observations.

So that the sample size parameter may have fractional values and so that limits, as the sample size approaches zero, may be precisely defined, we shall introduce explicit definitions of sums, products, maxima and minima when the index has a non-integral upper limit; any other convenient definitions leading to the same limiting behavior might do as well.

Let m be positive and let m_* be the largest integer less than or equal to m . For any function $\omega(z)$, let $\omega_i = \omega(z_i)$, given z_1, z_2, \dots , and suppose the supremum to the range of ω is a , and the infimum is b . We then define:

$$(1) \sum_{i=1}^m \omega_i = \sum_{i=1}^{m_*} \omega_i + (m - m_*) \omega_{m_*+1} ;$$

$$(2) \prod_{i=1}^m \omega_i = e^{\sum_{i=1}^m \omega_i^f} = \left(\prod_{i=1}^{m_*} \omega_i \right) (\omega_{m_*+1})^{m-m_*}$$

where $\omega_i^f = \log \omega_i$ and $\omega_i > 0$;

$$(3) \max_{i \leq m} \omega_i = \max_{i \leq m_*} \omega_i + (m - m_*) \left(\max_{i \leq m_*+1} \omega_i - \max_{i \leq m_*} \omega_i \right)$$

where $\max_{i \leq 0} \omega_i = a$;

$$(4) \min_{i \leq m} \omega_i = - \max_{i \leq m} (-\omega_i)$$

$$= \min_{i \leq m_*} \omega_i - (m - m_*) (\min_{i \leq m_*+1} \omega_i - \min_{i \leq m_*} \omega_i)$$

where $\min_{i \leq 0} \omega_i = - \max_{i \leq 0} (-\omega_i) = b$.

Hence taking limits as m tends to zero, we have:

$$(1') \quad \lim_{m \rightarrow 0} \sum_{i=1}^m \omega_i = 0 ,$$

$$(2') \quad \lim_{m \rightarrow 0} \prod_{i=1}^m \omega_i = 1 ,$$

$$(3') \quad \lim_{m \rightarrow 0} \max_{i \leq m} \omega_i = a$$

$$(4') \quad \lim_{m \rightarrow 0} \min_{i \leq m} \omega_i = b$$

In particular, if the range of z is $(-\infty, +\infty)$, then

$$\lim_{m \rightarrow 0} \max_{i \leq m} z_i = -\infty , \quad \lim_{m \rightarrow 0} \min_{i \leq m} z_i = +\infty .$$

These limits and the value $m = 0$ will be referred to as the null values of the parameters of the NCBD. We shall refer to this procedure of obtaining a prior Bayes density by allowing the parameters of the Bayes density to approach null values (which are consistent with the interpretation of no prior data) as the principle of Minimum Prior Data (MPD). In typical cases the application of this principle leads to a prior Bayes density which may be included in the NCBD class by the extension technique described in 4.2. The choice of prior density obtained by this extension technique will always be taken so that the posterior density after n observations with sufficient statistic s will be identical to the posterior density obtained by employing the general form of the NCBD as the prior density to obtain a posterior density and then evaluating that posterior density for null values of the parameters of the prior NCBD. To illustrate, the NCBD for the location-parameter normal model is

$$b_m(\mu) = \frac{\sqrt{m}}{\sqrt{2\pi}} \exp(-m(\mu - \bar{z})^2/2) \quad \text{for } m > 0.$$

The extension technique yields $b_0(\mu) = C$. Now if $b_n(\mu)$ is taken as the prior Bayes density the posterior Bayes density is

$$b_{m+n}(\mu) = \frac{\sqrt{m+n}}{\sqrt{2\pi}} \exp \left[- (m+n) \left(\mu - \frac{\sum z_i + \sum x_i}{m+n} \right) / 2 \right]$$

which, upon evaluation for null values $\sum z_i = 0$, $m = 0$, is

$$b_n(\mu) = \frac{\sqrt{n}}{\sqrt{2\pi}} \exp - n \left(\mu - \frac{\sum x_i}{n} \right) / 2$$

which is identical to the posterior Bayes density obtained employing the prior Bayes density $b_0(\mu) = C$.

Unfortunately, this heuristic principle cannot stand on its own as an invariant Bayesian postulate. Suppose $\theta = \varphi(p)$ is an allowable transformation for the parameter of the Bernoulli model. Let $p = \pi(\theta)$ be the inverse transformation; then a more general form of the likelihood is

$$m_{\theta, n}(\underline{x}) = \left[\pi(\theta) \right]^x \left[1 - \pi(\theta) \right]^{n-x}$$

and the NCBD for θ is

$$b_m(\theta) = \frac{\int \pi(\theta)^z \int 1 - \pi(\theta) \int^{m-z}}{\int \int \pi(\theta)^z \int 1 - \pi(\theta) \int^{m-z} d\theta}, \quad \text{for } m \geq z > 0.$$

This implies

$$b_m(p) = \frac{p^z (1-p)^{m-z} \varphi'(p)}{\int_0^1 p^z (1-p)^{m-z} \varphi'(p) d p} \quad (4.5)$$

Then

$$\lim_{\substack{m \rightarrow \infty \\ z \rightarrow \infty \\ z \leq m}} b_m(p) = \frac{\varphi'(p)}{\int_0^1 \varphi'(p) d p} = \frac{\varphi'(p)}{\varphi(1) - \varphi(0)}$$

which will not be consistent with (4.4) unless θ is a linear function of p . The MPD principle is thus subject to exactly the same difficulty as the original Bayes postulate.

Raiffa and Schlaifer have suggested that when there exists some prior knowledge concerning the parameters that prior distributions might be chosen so as to conform with subjective betting odds that a person might place on various values. However, accepting the idea that the prior density is to be made up from knowledge gained from prior observations, it would seem more natural, in developing a model for scientific inference, to select a prior density which corresponds to an equivalence between our prior knowledge and the results of some prior hypothetical experiment. That is, we might say that our current knowledge concerning the parameter is approximately that which we would have if we observed an experiment of a certain size and obtained certain results, with no information prior to that. Since we have extended the definition of the class of natural conjugates to non-integral m , we permit a continuum of prior information so that this class may be sufficiently rich to adequately characterize prior data obtained from non-independent observations or non-equivalent experiments.

4.4 Minimum Bayes Information

In light of the parametric formulation adopted it is possible to consider the Bayes problem to be that of making the minimal possible assumption about the process under study subject to the assumptions implicit in the model. This may be thought of as choosing a prior Bayes density which is minimally informative. The amount of information in a density function may be associated with the lack of spread in that function. For example, a density which was almost entirely concentrated over a small interval would tell us rather precisely what the value of an observation from the associated population might be, whereas a density function which was relatively flat over the spectrum of X would not yield very precise predictions. In the first instance we would consider the density function to be very informative; in the second case we would consider the density function to be very uninformative.

One possible measure of the degree of spread or lack of information in a distribution is the variance of that distribution. It is not, however, difficult to construct rather concentrated densities which have infinite variance. A second possible measure of information is that due to Fisher who defines the information in a sample as the expected value of the square of the derivative of the logarithm of the likelihood function. The important use of this function is in maximum likelihood estimation and in the Cramer-Rao lower bound for the variance of an unbiased estimator. As a true measure of information it is hardly satisfactory since it defines the amount of information obtained as being directly proportional to the number of observations. This con-

tradicts the well-accepted principle of decreasing marginal utility of successive equidistributional observations, that is, the fact that the first n observations from a process yield more information about that process than the conditional increment in information afforded by the second n observations given the information from the first n observations. The variance is preferable in this respect; for example the variance of the mean is given by the variance of the random variable divided by n , and hence the marginal decrement in variance is a decreasing function of n , assuming the variance of X is finite.

Another measure of the amount of information in a distribution was taken from the work of Shannon (1949) and applied in a Bayesian context by Lindley (1956, 1957, 1961, 1962) and by Kerridge (1961). A related application by the physicist E. T. Jaynes (1957) seems to have gone unnoticed by statisticians. If X has density $f(x)$ (with respect to Lebesgue measure); then the information in f is defined as

$$I(f) = \int f(x) \log f(x) dx \quad (4.6)$$

or by the corresponding sum in the discrete case. The convention $f(x) \log f(x) = 0$ when $f(x) = 0$ is assumed. For a joint density $h(x, y)$ the information is defined by

$$I(h) = \iint_{\mathcal{X} \times \mathcal{Y}} h(x,y) \log h(x,y) dy dx \quad (4.7)$$

with the restrictions as above. The important defining relation of the information measure is

$$I(h) = I(g) + \mathcal{E} I(f|Y) \quad (4.8)$$

where $I(g)$ is the information in the density $g(y)$ of Y and $\mathcal{E} I(f|Y)$ is the expected value with respect to the density $g(y)$ of the information in the conditional density of X given Y . It has been shown by Shannon that, in the discrete case, $I(f)$ is the unique function, for fixed logarithm base, satisfying this condition and a mild continuity property. It has also been shown that

$$I(h) \geq I(f) + I(g) \quad (4.9)$$

with equality if and only if X and Y are independent.

If X has a continuous density with spectrum \mathcal{X} of finite Lebesgue measure d , then the density $f(x) = d^{-1}$, $x \in \mathcal{X}$, with information $I(f) = -\log d$ is the (proper) density over \mathcal{X} having least information. If X has a discrete density defined over a finite set of N points then the density $f(x) = N^{-1}$ with information $I(f) = -\log N$ is minimally informative over this spectrum. For continuous X , $-\infty < X < \infty$, $\text{Var}(X)$ fixed and $\mathcal{E}(X)$ arbitrary, the normal density is minimally informative. For continuous X , $0 < X < \infty$ and $\mathcal{E}(X)$ fixed the exponential density is minimally informative. For discrete X , $x = 0, 1, 2, \dots$ and $\mathcal{E}(X)$ fixed the geometric density is minimally informative. Finally, for discrete X , $x = 0, \pm 1, \pm 2, \dots$, $\text{Var}(X)$ fixed, $\mathcal{E}(X)$ arbitrary, the density proportional to $e^{-a(x-m)^2}$, $a > 0$, is minimally informative. The last density might be called the discrete normal density. We note that many of the Bayes densities encountered in this paper are minimally informative densities.

The infimum of the information function on an infinite set is $-\infty$ and may be approached, for example, by taking a normal distribution and letting its variance go to infinity. In the continuous case the supremum of the information function is infinity and may be approached by letting the above variance tend to zero. In the discrete case the maximum information is zero, which is attained by a density assigning probability one to a single point. The fact that the limiting information as a density approaches a point density in the continuous case is not the same as the information of a point density in the discrete case is cause for contemplation. Lindley (1956) has shown that $E_x I(b_n) \geq I(b)$ where expectation over \mathcal{X} is taken with respect to $f(x)$. Finally, Rajski (1960) has shown that information exists (finite) under a mild regularity condition provided there exists an $\epsilon > 0$ such that

$$\int_{-\infty}^{\infty} |x|^\epsilon m(x) dx < \infty$$

For improper densities or densities not having moments of order ϵ the Shannon information measure is not an adequate measure of spread as the Shannon information need not be finite. It is apparent that some improper densities are more spread out than others, and perhaps the one with the most spread is the improper uniform density. The inadequacy of the Shannon information measure is nicely illustrated by the following example: Consider the density

$$m_\theta(x) = \frac{\log \theta}{x (\log x)^2}, \quad 1 < \theta < x < \infty \quad (4.10)$$

It may be verified that this is a proper density, having no moments, and that its information is minus infinity for all θ , $1 < \theta < \infty$. By taking θ very close to one we find that almost all of the probability is concentrated near one. One would not happily say that this density was non-informative. While such densities and improper densities may be uninformative in the Shannon sense, it is clear that in a wider sense some are more uninformative than others. In particular, we accept the proposition that the improper uniform density is most uninformative. In doing this we are in effect assuming two levels of measurement of information. When Shannon information is sufficient for our purposes, we shall employ it; when it is not, i.e. when dealing with improper densities, we shall introduce other ideas to discriminate among densities.

A natural application of this theory to the Bayes problem would be to require that a minimally informative prior density be chosen for the parameter. We shall refer to this as the principle of Minimum Bayes Information (MBI). If the spectrum of θ is finite, this reduces to the Bayes postulate. Lindley and Kerridge have specified that in all cases the uniform density be taken, and Kerridge has displayed an interesting property of this specification. Since we have referred to the uniform (proper or improper) density as most uninformative, we shall require, as our MBI principle, that a uniform density be taken a priori. More specifically we shall require a uniform density on the joint space of all unknown parameters of the model. The invariance problem remains, however, as we have not specified what parameters should have a uniform density. Finally we note that the MBI principle is consistent with the

restriction to NCBD's since a uniform prior is always a limiting form of a NCBD if a NCBD exists. Moreover an MBI density is always adaptive, as is a NCBD.

4.5 Minimum Necessary Sample

We shall take a second idea from Lindley (1961) which is best introduced by the following quotation:

"Some clues on the nature of a prior distribution can be obtained from statements that are commonly made ... it is said that such and such an observation gives no information about a parameter. For example it is often said that a sample of size one from a normal distribution of unknown mean and variance gives no information about the variance To a Bayesian this presumably means that the prior and posterior distributions are identical."

Lindley indicates his inability to do anything substantial with this principle. We shall employ this kind of idea though for present purposes we shall modify the statement somewhat. Let us replace the last sentence with the following:

To an information-theoretically-oriented-Bayesian this presumably means that the Shannon information in the prior and posterior densities are identical (or perhaps in certain contexts, that the prior and posterior densities are both improper).

By careful statement and application this principle will be seen to be workable for important examples when used in conjunction with the other principles developed in this chapter.

In Chapter V we shall examine the normal models in detail; however in order to introduce our method for utilizing the above general idea, it will be convenient to consider the location parameter normal model here. A NCBD for the normal model (4.1) is the class of normal densities indexed by the parameters (m, \bar{z}) . The expectation of this density is \bar{z} and the variance is $(m+1)/m$. Using this as a prior Bayes density the posterior density is indexed by the parameters $(m+n, \frac{\sum z_i + \sum x_i}{m+n})$. Clearly then, the entire class of Bayes densities can be characterized by the improper and proper densities

$$b_m(\mu) = \int \left(\frac{m+1}{m}\right)^{-1/2} \frac{1}{\sqrt{2\pi}} \exp \left[-(\mu - \bar{z})^2 / 2 \left(\frac{m+1}{m}\right) \right],$$

for $m > 0$ and $-\infty < \bar{z} < \infty$ and by $b_0(\mu) = C$. Now $b_m(\mu)$ will be proper for $m = \epsilon > 0$, however small. For $m = 0$ we have an improper uniform density. It is not unreasonable to think of this as implying that with the prior improper uniform density, the posterior density will be proper after any amount of data, even an epsilon of an observation! Similarly in the two parameter normal model we find that the posterior joint Bayes density will be improper after one observation, but proper (with probability one) after $1 + \epsilon$ observations when the prior density is joint improper uniform and the parameters are (μ, λ) . If we were to add a skewness parameter, presumably we would want to have an improper joint Bayes density after two observations but a proper one after $2 + \epsilon$ observations. We shall not attempt to make explicit now just how an epsilon of an observation may be obtained; however it would seem reasonable that two dependent observations would on the average contain

more information than one observation provided that the dependence was not complete and that these two observations would contain less information, on the average, than two independent observations. Also an epsilon of an observation may be thought of in terms of a procedure which takes an observation with probability ϵ and fails to take an observation with probability $1 - \epsilon$.

For most models studied we shall require that the joint Bayes density be minimally-Shannon-informative after $K-1$ observations, where K is the number of parameters in the model and more than minimally informative after $K-1+\epsilon$ observations. The multinomial and bivariate-normal models will not follow this pattern but will be seen to be consistent with "statements that are commonly made". We specify m_0 to be the largest sample number (always an integer) for which the joint Bayes density is to be minimally Shannon informative. The Bayesian analysis presented in Chapter VII gives some insights into why, for the multinomial model, $m_0 = 0$, regardless of the number of categories.

We shall refer to this requirement that the posterior Bayes density become more than minimally (Shannon) informative at just the proper time as the principle of the Minimum Necessary Sample (MNS). In multi-parameter problems it will be most convenient to apply this principle to appropriate marginal and conditional densities. For example, for the two parameter normal model we will require that the conditional density of μ given λ be proper for $m = \epsilon$, the conditional density of λ given μ be proper for $m = \epsilon$, while the marginal density of λ be proper only after $m=1+\epsilon$.

We shall not attempt to give general rules for determining m_0 , freely admitting that this involves a real weakness in this principle. However, no difficulties arise in many important examples, including those treated herein. The heuristic idea is that m_0 should be one less than the number of whole observations needed to "say something about the unknown parameter."

For the improper uniform density the Shannon information is $-\infty$, zero, or $+\infty$ as c is < 1 , $=1$ or >1 . The Shannon information for the density x^{-1} does not even have a definition as an extended Lebesgue integral as

$$\int_0^1 \frac{\log x}{x} dx = -\infty \quad \text{and} \quad \int_1^{\infty} \frac{\log x}{x} dx = +\infty .$$

We shall be dealing with such improper densities; and, in order to make the proposed principles more precise, it will be necessary to define the information in these densities in some consistent manner. We have no difficulty in doing this as these densities are obtained as limiting densities of a class of proper densities. It will be convenient then to define the information in the limiting density as the limit of the information in the sequence of densities. Thus for the cases discussed in this paper we shall define the information in these densities to be $-\infty$, this being the above limiting information.

4.6 Minimum Fiducial Information

The fourth general principle to be reviewed is that due to Novick (1962). This principle is based upon the idea that the central subject of study should properly be taken to be the random variable itself and

not some arbitrary parameter. (A similar suggestion has since been made by Fraser (1962).) It was suggested that within the context of the parametric structuring defined above a prior Bayes density should be selected so as to minimize the information $I(f)$ in the fiducial density of X . Having specified a particular minimally informative f the integral equation

$$f(x) = \int m_{\theta}(x) b(\theta) d\theta \quad (4.11)$$

is considered, where $f(x)$ and $m_{\theta}(x)$ are presumed known and $b(\theta)$ is presumed to be unknown. The problem of lack of invariance under reparametrization then disappears. It is replaced, however, by two other problems. The first is the uniqueness of solution of the integral equation (4.11) (assuming a solution exists). For the example (4.1) a factorization of the density of a single observation

$$m_{\mu}(x) = (2\pi)^{-\frac{1}{2}} \exp - (x-\mu)^2/2 = e^{x\mu} \left((2\pi)^{-\frac{1}{2}} \exp - \left(\frac{x^2 + \mu^2}{2} \right) \right) = e^{x\mu} \mathcal{L}(\mu),$$

say, yields the equation

$$f(x) = \int_{-\infty}^{\infty} e^{x\mu} \left\{ \mathcal{L}(\mu) b(\mu) \right\} d\mu .$$

Hence $\mathcal{L}(\mu) b(\mu)$ and hence $b(\mu)$ are unique by the general uniqueness of bilateral Laplace transforms. For the Poisson model we take

$$m_{\lambda}(x) = \frac{e^{-\lambda t} (\lambda t)^x}{x!} \quad t > 0, \lambda > 0 \quad x = 0, 1, 2, \dots .$$

The equation (4.10) becomes

$$f(x) = \int_0^{\infty} \lambda^x g(\lambda) d\lambda, \text{ say.}$$

The requirement that this equation hold for $x = 0, 1, 2, \dots$ requires that all solutions $g(\lambda)$ have the same moments and hence, since the moment theorem is applicable the solution is unique.

The choice of a minimally informative fiducial density is more difficult than that of the choice of a minimally informative Bayes density. At times it will be possible to obtain a solution to (4.10) when the fiducial density is specified to be uniform. If a uniform fiducial density can be obtained from some uniform Bayes density it will be required, otherwise "we will attempt to get as close as possible to a uniform fiducial density". This principle will be called the principle of Minimum Fiducial Information (MFI). Some difficulties arise in applying the principle on its own, particularly in obtaining unique solutions to (4.11) and in specifying the correct fiducial density. We shall find, however, that its application in conjunction with the other principles proposed in this chapter will be fruitful.

Since a uniform fiducial density is often unobtainable, we will often need to determine which of a class of improper densities is least informative. Ideally we would want some more general information measure than the Shannon information measure in order to make this discrimination. Unfortunately we know of none currently available. If we were to attempt to make such a discrimination over an arbitrary class of improper densities, we would now find that task impossible. Within the framework discussed in this paper, however, the problem seems less diffi-

cult. We shall find that we need only make such a discrimination among classes of densities which are rather obviously ordered in this respect.

To illustrate the logic we shall employ consider the densities (4.10). Each of these functions is a proper density though with Shannon information equal to minus infinity. However, from an intuitive point of view there is no difficulty in making the needed discrimination. For θ near one, almost all of the probability is concentrated near θ , and we must say that this density is very informative. As θ becomes large the density approaches a uniform density on (θ, ∞) and hence becomes progressively less informative over its spectrum. If we reduce the class by restricting θ to $1 < \theta \leq \theta_0$, then it seems most reasonable to say that m_{θ_0} is least informative.

Along this same line of reasoning, when comparing the improper densities $m_{\theta}(x) = x^{-\theta}$, ($x > 0$), $0 < \theta_0 \leq \theta$, we would say that m_{θ_0} was minimally informative. Also we say that $m_1(x) \propto x^{-1} - ax^{-2}$ ($x > 0$) is more informative than $m_2(x) \propto x^{-1}$ or more generally that $m_2(x) \propto x^{-1}$ is minimally informative in the class $m_3(x) \propto x^{-1} - ax^{-r-1}$, for $a, r > 0$ (see 5.2).

We may note that the fiducial density is proper if and only if the corresponding Bayes density is proper, for, formally,

$$\int_{\mathcal{X}} f_m(x) dx = \int_{\mathcal{X}} \int_{\Theta} m_{\theta}(x) b(\theta) d\theta dx = \int_{\Theta} \int_{\mathcal{X}} m_{\theta}(x) dx b(\theta) d\theta = \int_{\Theta} b_m(\theta) d\theta .$$

The stated result is then a consequence of Fubini's theorem for positive functions. Additionally we may note that

$$\int_{\mathcal{X}^n} f_n(\underline{x}) \frac{d\underline{x}}{d\underline{x}_n} = \int_{\mathcal{X}^n} \prod_{i=1}^n m_{\theta}(x_i) b(\theta) d\theta$$

exists if and only if $\int_{\mathcal{X}^n} b(\theta) d\theta$ exists, where the first integration is over the n-dimensional X-spectrum. Also Lindley has shown that

$$E_{\mathcal{X}} I(f_n) \geq I(f)$$

where the expectation is with respect to f .

4.7 The Natural Bayesian Structuring

In this chapter, we have introduced four principles which might be considered in the selection of a parametrization and a prior distribution thereon. In summary these four principles are:

(1) Minimum Prior Data (MPD) - the prior density shall be determined as a limiting NCBD, i.e. by taking null parameters to reflect the idea of no prior data.

(2) Minimum Bayes Information (MBI) - the prior density should be uniform over the (joint) spectrum of all parameters.

(3) Minimum Necessary Sample (MNS) - the Bayes and fiducial densities should be minimally (Shannon) informative when the number of observations $n = 0, 1, 2, \dots, m_0$ is less than some postulated value m_0 but more than minimally informative for all values $n > m_0$.

(4) Minimum Fiducial Information (MFI) - the prior fiducial density should be minimally informative, subject to the restrictions imposed by the model.

We propose that each of the four principles defined above is a reasonable requirement for the specification of a prior Bayes density under indifference in the context of a Bayesian parametric structuring. A Bayesian parametric structuring which satisfies all of these principles will be called, for convenience of reference, a Natural Bayesian Structuring (NBS) for the model. We propose to show that a NBS can be obtained for important statistical models and that this structuring is essentially unique in these cases. By essentially unique we mean unique up to a linear transformation on the parameter. For some models the uniqueness will not depend on all four principles, though the structuring determined by any lesser number will, in these cases, satisfy all four principles.

We further propose that the results of the Bayesian analysis be represented by the densities $b_n(\theta)$ and $f_n(x)$, the posterior Bayes and posterior fiducial densities. The posterior Bayes density is to be used, for $n > m_0$, to make logical probability statements about θ , where θ is the parameter which bears a uniform Bayes density a priori under the NBS. The posterior fiducial density of X , for $m > m_0$ is to be used in a predictive sense, that is, it is to be used to replace the unknown true density $m(x)$ to predict values of the $(n+1)$ -st observation.

The procedure for determining and verifying the NBS follows a typical pattern which will be demonstrated in detail for the first two models studied. If the "correct" parameter is not known, the analysis will follow the somewhat more complex pattern illustrated in 5.2;

however for brevity at this point we assume that we are fortunate enough to begin with the parameter which proves to be "correct". We then postulate the value m_0 and define the NCBD for $m > m_0$. This definition is then extended for values $m_0 = 0, 1, 2, \dots, m_0$ and the corresponding fiducial densities are defined.

The prior density of the parameter is then obtained as an extended NCBD for null values of the parameters thus satisfying the MPD principle. It is then shown that this prior density is uniform on the joint spectrum of the parameters, thus satisfying the MBI principle. It is then shown that the posterior Bayes densities became more than minimally (Shannon) informative at the proper time, thus satisfying the MNS principle. In multi-parameter problems the application of this principle will be made in terms of certain marginal and conditional densities. It is then shown that the prior fiducial density is minimally informative, in the sense discussed above. Finally we show that there cannot exist any other NBS. This is accomplished by means of the kinds of completeness arguments discussed in 4.6 or by extending the class of model densities so that it is complete.

At this point we wish to affirm that the four Bayesian indifference principles were developed and refined inductively through consideration of successive models; the task being to find some set of intuitively reasonable principles which would be adequate to specify uniquely the prior Bayes density under indifference for a reasonably wide range of models, and to furnish a model which permitted the kinds of semantic

interpretations which seemed desirable for the purposes of scientific inference. There have been many previous attempts to do this. None of these previous proposals have presented an internally consistent and widely applicable methodology.

CHAPTER V
NORMAL MODELS

In this chapter the principles developed in the preceding chapter are applied to obtain unique specifications of prior Bayes densities under indifference for each of the three basic univariate normal models and for three bivariate normal models.

5.1 Univariate Normal Model - Location Parameter Unknown

We assume the model density

$$m_{\mu}(x) = (2\pi)^{-\frac{1}{2}} \exp \left[-\frac{(x-\mu)^2}{2} \right], \quad -\infty < x < \infty, \quad -\infty < \mu < \infty. \quad (5.1)$$

The joint likelihood of n observations is then

$$m_{\mu}(\underline{x}) = (2\pi)^{-n/2} \exp \left[-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2} \right].$$

We define the NCBD as

$$\begin{aligned} b_m(\mu) &= \frac{\exp \left(\mu \sum z_i - \frac{m \mu^2}{2} \right)}{\int_{-\infty}^{\infty} \exp \left(\mu \sum z_i - \frac{m \mu^2}{2} \right) d\mu} \\ &= \left(\frac{m}{2\pi} \right)^{1/2} \exp \left(-\frac{m(\mu - \bar{z})^2}{2} \right) \end{aligned} \quad (5.2)$$

for $m > 0$, and all $\bar{z} = \sum z_i / m$. We interpret the parameters (m, \bar{z}) of the Bayes density as referring to m prior observations with mean \bar{z} . The fiducial density, for $m > 0$, is

$$f_m(x) = \left(\frac{n}{2\pi(m+1)} \right)^{1/2} \exp - \left(\left(\frac{m}{m+1} \right) \left(\frac{x-\bar{z}}{2} \right)^2 \right) \quad (5.3)$$

The MPD principle requires that the prior Bayes density be found as an extension of $b_m(\mu)$ for null values of $(m, \Sigma z_i)$. The ratio $b_m(\mu_1)/b_m(\mu_2)$ for arbitrary values (μ_1, μ_2) is

$$\frac{b_m(\mu_1)}{b_m(\mu_2)} = \frac{\exp \left[-m\mu_1^2/2 + \mu_1 \Sigma z_i \right]}{\exp \left[-m\mu_2^2/2 + \mu_2 \Sigma z_i \right]} \longrightarrow 1, \text{ as}$$

$(m, \Sigma z_i) \rightarrow (0, 0)$. Hence we define $b(\mu) = c$. Then by direct integration we obtain $f(x) = c$.

For $m = 0$ both (5.3) and (5.4) are improper uniform densities, and for $m > 0$ both are proper densities with finite Shannon information. Thus postulating the value $m_0 = 0$ we see that the MNS principle is satisfied and we have already shown that the MPD, MBI and MFI principles are satisfied. Hence the specification $b(\mu) = c$ defines a NBS. Since a uniform fiducial density is obtainable, it is required by the MFI principle. Hence the essential uniqueness of the NBS follows from the uniqueness of the bilateral Laplace transform. This is perhaps the only problem about which there has been no controversy concerning the "correct" choice of prior density.

Under the NBS the posterior Bayes and fiducial densities are given by (5.2) and (5.3) with (m, \bar{z}, x) replaced by (n, \bar{x}, \hat{x}) , where \hat{x} is the $(n+1)$ st observation of X . The Shannon information in these densities is

$$I(b_n) = -\log \sqrt{2\pi e/n} \quad \text{and} \quad I(f_n) = -\log \sqrt{2\pi e(n+1)/n} \quad (5.4)$$

and depends only upon the sample size. This result is quite atypical. We note also that both the Bayes and fiducial densities are minimally informative densities, for fixed variance; also the model density is a minimally informative density, for variance one. The limiting values, as $n \rightarrow 0$, of the functions (5.4) are $-\infty$, i.e. minimally informative on the real line. In this case the limiting information of both the sequences of Bayes and fiducial densities is equal to the information of the limiting density. In this case and in others employing this derivation the requirement, stipulated in Chapter II, concerning the power series expansion of the transformation is not required.

5.2 Univariate Normal - Scale Parameter Unknown

We assume the model density

$$m_{\sigma}(x) = \sigma^{-1}(2\pi)^{-1/2} \exp - x^2/2\sigma^2, \quad -\infty < x < \infty, \quad \sigma > 0. \quad (5.5)$$

The joint likelihood of n observations is

$$m_{\sigma}(x) = \sigma^{-n}(2\pi)^{-n/2} \exp(- \sum_{i=1}^n x_i^2/2\sigma^2). \quad (5.6)$$

If σ were the correct parameter we would define the NCED as

$$\begin{aligned} b_m(\sigma) &= \frac{\sigma^{-m} \exp(- \sum_{i=1}^m z_i^2/2\sigma^2)}{\int_0^{\infty} \sigma^{-m} \exp(- \sum_{i=1}^m z_i^2/2\sigma^2) d\sigma} \\ &= \frac{2 \left(\frac{\sum z_i^2}{2} \right)^{m-\frac{1}{2}} \sigma^{-m} e^{-\sum z_i^2/2\sigma^2}}{\Gamma(m - \frac{1}{2})} \end{aligned} \quad (5.7)$$

Postulating the value $m_0 = 0$ we see that (5.7) is not acceptable since $b_{1/2}(\sigma)$ is improper, thus contradicting the MNS principle. We therefore require a more suitable parameter. Let $\lambda = \varphi(\sigma)$ be an allowable transformation and suppose φ' is of order K . The corresponding integral to be considered is then

$$\int \frac{\sigma^{-(m-K)}}{\Gamma(m-K-1)} \exp\left(-\sum_{i=1}^m z_i^2/2\sigma^2\right) d\sigma$$

which will diverge for null values and converge otherwise for $K \leq -1$. Hence the MNS principle requires that the derivative be of order less than or equal to σ^{-1} . The corresponding fiducial density will be of maximum order, i.e. of order x^{-1} , when $\varphi'(\sigma) = \sigma^{-1}$. Since φ is allowable we also must have $\varphi'(\sigma) > 0$. Considering any other φ (other than $\varphi(\sigma) = \log \sigma$ with $\varphi'(\sigma) = \sigma^{-1}$) satisfying these conditions, e.g. $\varphi(\sigma) = \log \sigma - a\sigma^{-r}$, $r > 0$, $a > 0$, we find that for any such parameter the MPD fiducial density is of order x^{-1} but in each case the addition of terms of lower order in the Jacobian results in the fiducial density tending to infinity more quickly as $x \rightarrow 0$ than the function x^{-1} . The series expansion requirement for φ' is necessary to employ the logic of section 4.6 and to define the improper density x^{-1} as the minimally informative fiducial density and hence to conclude that $\lambda = \log \sigma$ is (essentially) uniquely specified as the correct parameter of the NBS. Without the power series requirement on φ' we would have difficulty in ruling out $\lambda = (\log \sigma)^{2n+1}$, $n = 1, 2, 3, \dots$. We may also see that there is no solution $K < -1$ to the equation

$$x^\alpha = \int \frac{1}{\sigma/\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \sigma^K d\sigma$$

$$= \frac{2 \left(\frac{x^2}{2}\right)^{2K} \Gamma(-2K)}{\sqrt{2\pi}}$$

for $-1 < \alpha \leq 0$ and hence we see again that x^{-1} is the required MFI density. All future examples employing this mode of derivation are dependent upon the power series requirement on φ' while those employing the derivation of 5.1 do not.

Taking λ as the correct parameter we have

$$b_m(\lambda) = \frac{2 \left(\frac{\sum z_i^2}{2}\right)^{m/2} \exp - \left[m\lambda + \left(\frac{\sum z_i^2}{2}\right) \right] \exp(-2\lambda)}{\Gamma\left(\frac{m}{2}\right)} \quad (5.8)$$

for $m > 0$. For $m = 0$, $b_m(\lambda) = c$, which may be demonstrated in the usual manner. The fiducial density is the "t" density

$$f_m(x) = \frac{\Gamma\left(\frac{m+1}{2}\right) (\sum z_i^2)^{m/2}}{\Gamma\left(\frac{m}{2}\right) \sqrt{2\pi} \left(\frac{x^2 + \sum z_i^2}{2}\right)^{\frac{m+1}{2}}} \quad (5.9)$$

for $m > 0$. The posterior Bayes and fiducial densities are given by (5.8) and (5.9) with $(x, \sum z_i^2, m)$ replaced by $(\hat{x}, \sum x_i^2, n)$.

The results obtained in this case agrees with Jeffreys' general specification for a model indexed by a parameter taking values $(0, \infty)$, for the law $b(\lambda) = c$ is equivalent to the law $b(\sigma) = \sigma^{-1}$. This result also agrees with that obtained from Fisher's fiducial theory. Finally we note that the limiting information, as $m \rightarrow 0$, $\sum z_i^2 \rightarrow 0$,

in the density (5.9) is $-\infty$ hence the definition $I(f) = -\infty$ is consistent.

5.3 Univariate Normal - Location and Scale Parameters Unknown.

We assume the model density

$$b_{\mu, \lambda}^m(x) = (2\pi)^{-1/2} (\exp - \lambda) \left[\exp\left(-\frac{(x-\mu)^2}{2}\right) \exp - 2\lambda \right] \quad (5.10)$$

for $-\infty < x, \mu, \lambda < \infty$, which is the usual two parameter normal model with $\lambda = \log \sigma$. The choice of parameters here has been suggested by cases 1 and 2. For this parametrization we define the NCBD as

$$b_m(\mu, \lambda) = \frac{e^{-m\lambda} \exp\left[-\frac{e^{-2\lambda}}{2} \left(\sum_{i=1}^m (z_i - \mu)^2\right)\right]}{\iint e^{-m\lambda} \exp\left[-\frac{e^{-2\lambda}}{2} \left(\sum_{i=1}^m (z_i - \mu)^2\right)\right] d\mu d\lambda}$$

$$= \left(\frac{2 \left[\frac{\sum z_i^2 - m \bar{z}^2}{2} \right]^{\frac{m-1}{2}} e^{-(m-1)\lambda} e^{-e^{-2\lambda} \left[\frac{\sum z_i^2 - m \bar{z}^2}{2} \right]}}{\Gamma\left(\frac{m-1}{2}\right)} \right) \times$$

$$\left(\frac{1}{\sqrt{2\pi}} e^{-e^{-2\lambda} \frac{m}{2} \left[\mu - \bar{z} \right]^2} \right) \quad (5.11)$$

for $m > 1$, $-\infty < \bar{z} (= \sum z_i/n) < \infty$, $0 < V = \left(\frac{m \sum z_i^2 - (\sum z_i)^2}{m^2} \right) < \infty$.

Consistent with the ratio $b_1(\mu_1, \lambda_1)/b_1(\mu_2, \lambda_2)$ we define

$$b_1(\mu, \lambda) = \frac{\sqrt{e^{-2\lambda}}}{\sqrt{2\pi}} e^{-e^{-2\lambda}} \left(\frac{1}{2}\right) (\mu - \bar{z})^2 = b_1(\mu | \lambda)$$

$$\text{and } b_1(\lambda) = 1. \quad (5.12)$$

Now $b_0(\mu_1, \lambda_1) / b_0(\mu_2, \lambda_2) = 1$, hence define $b_0(\mu, \lambda) = c$.

Correspondingly we have

$$f(x) = c \quad \text{and} \quad f_1(x) = \frac{1}{(x-z)^2}.$$

For $m > 1$ the fiducial density is given by

$$f_m(x) = \frac{\Gamma(\frac{m}{2}) m^{1/2} \left(\frac{m \sum z_i^2 - (\sum z_i)^2}{2} \right)^{\frac{m-1}{2}}}{(\overline{m+1})^{1/2} (2\pi)^{1/2} \Gamma(\frac{m-1}{2}) \left(\frac{(m+1)(x + \sum z_i^2) - \sqrt{\sum z_i^2 + x^2}}{2(m+1)} \right)^{m/2}} \quad (5.13)$$

Since $b_0(\mu, \lambda) = c = f(x)$ the MBI and MFI principles are satisfied. The first factor of (5.11), which is the marginal density of λ , will be proper for $m > 1$. For $m = 1$ the marginal density of λ is improper. The second factor of (5.11), the conditional density of μ given λ , is proper for $m > 0$. For $m = 0$, $b_m(\mu | \lambda)$ is improper. Applying the uniqueness argument of 5.2 to the first density and the uniqueness argument of 5.1 to the second density we find that the above NBS is unique.

The posterior Bayes and fiducial densities for $m > 1$ are given by (5.11) and (5.13) with $(m, \sum z_i, \sum z_i^2, x)$ replaced by $(n, \sum x_i, \sum x_i^2, \hat{x})$. The fiducial density of X is again a generalized "t" density. The obtained prior law $b(\mu, \lambda) = c$ is equivalent to Jeffreys' law $b(\mu, \sigma) = \sigma^{-1}$ and also agrees with the results obtained by Fisher em-

ploying his fiducial method. Thus the obtained NBS implies a solution to the Behrens-Fisher problem which would be consistent with that originally proposed by Behrens and Fisher and that proposed by Jeffreys.

5.4 Bivariate Normal Model - Regression Parameter Unknown

In 5.2 we demonstrated a technique for finding the correct parameterization for the model. In the case of the dependent bivariate normal model the correct initial choice of association parameter is even more crucial to the analysis than the correct initial choice of scale parameter in 5.2, for if we were to choose the wrong parameter (yet one of the common association parameters) we would be unable to obtain a closed form expression for the NCB. For the one parameter (dependence parameter) bivariate normal model the choice of ρ , the correlation coefficient, as the dependence parameter leads to the consideration of the integral

$$\int_{-1}^1 (1-\rho^2)^{-m/2} e^{-(1-\rho^2) \left\{ \sum u_i^2 - 2\rho \sum u_i v_i + \sum v_i^2 \right\}} d\rho$$

which does not appear to exist in closed form. To solve this problem it seems necessary to replace the correlation parameter by a regression parameter. In the simplest case we have the model given by

$m_{\beta}(y, x) = m_{\beta}(y|x) m(x)$ where

$$m_{\beta}(y|x) = (2\pi)^{-1/2} \exp - \frac{1}{2} (y - \beta x)^2 \quad (5.14)$$

and where X is marginally standard normal. The joint likelihood of n observations is then proportional to

$$(2\pi)^{-n/2} e^{-\frac{\sum x_i^2}{2} \sqrt{\beta} - \frac{\sum x_i y_i}{\sum x_i^2} \sqrt{\beta}} e^{-\frac{1}{2} \sqrt{\sum y_i^2 - \frac{(\sum x_i y_i)^2}{n}}} \quad (5.15)$$

We define the NCBD as

$$b_m(\beta) = \frac{\sqrt{v}}{\sqrt{2\pi}} e^{-\frac{v}{2} \sqrt{\beta - b}^2} \quad (5.16)$$

for $m > 0$, $v = \sum u_i^2 > 0$, $-\infty < b = \frac{\sum u_i v_i}{\sum u_i^2} < \infty$. The fiducial density is

then given by $f_m(y|x) f_m(x)$ where

$$f_m(y|x) = \frac{\sqrt{v}}{\sqrt{x^2 + v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{v}{2(x^2 + v)} (y - bx)^2} \quad (5.17)$$

and where X is marginally standard normal. In the usual manner we obtain $b(\beta) = c = f(y|x)$ and infer uniqueness of the NBS. The usual substitutions in (5.16) and (5.17) yield the posterior Bayes and fiducial densities.

5.5 Bivariate Normal Model - Three Parameters Unknown

A somewhat more realistic and more frequently applicable model is

$$f_{\sigma_{2.1}, \mu_2, \beta_{2.1}}(x, y) = \frac{1}{\sigma_{2.1} \sqrt{2\pi}} e^{-\frac{1}{2\sigma_{2.1}^2} (y - \{\mu_2 + \beta_{2.1}x\})^2} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right) \quad (5.18)$$

The first factor is the conditional density of Y given X , and the second is the marginal density of X . We shall, of course, work with $\lambda_{2.1} = \log \sigma_{2.1}$, but shall henceforth express the model densities using the somewhat less cumbersome form employing $\sigma_{2.1}$. The unknown parameters are μ_2 , the mean of Y , $\beta_{2.1}$ the regression coefficient of Y on X and $\sigma_{2.1}^2$ the conditional variance of Y given X . We are assuming that the mean and variance of X are known and hence we may, without ambiguity let $\sigma = \sigma_{2.1}$, $\mu = \mu_2$, $\beta = \beta_{2.1}$, $\lambda = \lambda_{2.1}$. Written in terms of (μ, σ, β) the joint likelihood of n observations is

$$m(\underline{x}, \underline{y}) = (\sigma^{-n} (2\pi)^{-n/2}) e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\mu + \beta x_i)]^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (5.19)$$

The determination of the NCBDF for (μ, λ, β) is straightforward, though tedious. We have, for $m > 2$,

$$b_m(\mu, \lambda, \beta) = \left(\frac{e^{-\lambda/\bar{m}}}{\sqrt{2\pi}} e^{\frac{m}{2} e^{-2\lambda} [\bar{\mu} - (\bar{v} - \beta \bar{\mu})]^2} \right) \times$$

$$\left(\frac{e^{-\lambda/\bar{s}_1}}{\sqrt{2\pi}} e^{-\left(\frac{s_1}{2}\right) e^{-2\lambda} \left[\beta - \frac{s_{12}}{s_1}\right]^2} \right) \times$$

$$\left(\frac{2 \int \frac{s_2 - \frac{s_{12}^2}{s_1}}{2} e^{-(m-2)\lambda} e^{-\frac{1}{2} e^{-2\lambda} \left[s_2 - \frac{s_{12}}{s_1}\right]^2}}{\Gamma\left(\frac{m-2}{2}\right)} \right) \quad (5.20)$$

where u is to x as v is to y and

$$S_1 = \sum u_i^2 - \frac{(\sum u_i)^2}{m}, \quad S_2 = \sum v_i^2 - \frac{(\sum v_i)^2}{m}, \quad S_{12} = \sum u_i v_i - \frac{(\sum u_i)(\sum v_i)}{m}.$$

The first factor is the conditional density of μ given β and λ and is proper for $m > 0$. The second factor is the conditional density of β given λ and is proper for $m > 1$ by virtue of the definitions of $\sum u_i$ and $\sum v_i$. The last factor is the marginal density of λ . Proceeding in the usual manner we are able to define $b(\mu, \lambda, \beta) = c$ and $f(y|x) = c$ and infer the uniqueness of the NBS. The usual substitutions in (5.20) yield the posterior Bayes density.

5.6 Bivariate Normal Model - Five Parameters Unknown

No new techniques are needed to handle this case. The model density assuming that the mean and variance of X are unknown is

$$\left(\frac{\sigma_{2.1}^{-1}}{(2\pi)^{1/2}} e^{-\frac{1}{2\sigma_{2.1}^2} (y - \mu_2 + \beta_{2.1}(x - \mu_1))^2} \right)^x \left(\frac{\sigma_1^{-1}}{(2\pi)^{1/2}} e^{-\frac{1}{2\sigma_1^2} (x - \mu_1)^2} \right) \quad (5.21)$$

The NCBDF is then defined, for $m > 2$, by

$$\begin{aligned}
b_m(\mu_1, \mu_2, \lambda_1, \lambda_{2.1}, \beta_{2.1}) &= \left(\frac{\sqrt{m}}{\sqrt{2\pi}} e^{-\lambda_{2.1}} e^{-\frac{m}{2}} e^{-2\lambda_{2.1} \sqrt{\mu_2 - v}^2} \right) x \\
&\left(\frac{\sqrt{m}}{\sqrt{2\pi}} e^{-\lambda_1} e^{-\frac{m}{2}} e^{-2\lambda_1 \sqrt{\mu_1 - \bar{\mu}}^2} \right) \left(\frac{2 \left(\frac{S_1}{2}\right)^{\frac{m-1}{2}} e^{-(m-1)\lambda_1} e^{-e^{-2\lambda_1 \left(\frac{S_1}{2}\right)}}}{\Gamma\left(\frac{m-1}{2}\right)} \right) x \\
&\left(\frac{2 \left(\frac{S_2 - C_1}{2}\right)^{\frac{m-2}{2}} e^{-(m-2)\lambda_{2.1}} e^{-e^{-2\lambda_{2.1} \left(\frac{S_2 - C_1}{2}\right)}}}{\Gamma\left(\frac{m-2}{2}\right)} \right) \frac{\sqrt{S_1}}{\sqrt{2\pi}} e^{-\lambda_{2.1}} e^{-\frac{S_1}{2}} e^{-2\lambda_{2.1} \sqrt{\beta_{2.1} C_1}^2}
\end{aligned}$$

where $\bar{\mu} = \frac{\Sigma \{v_i - \beta_{2.1}(u_i - \mu_1)\}}{m}$, $u = \frac{\Sigma u_i}{m}$, $S_1 = \Sigma u_i^2 - \frac{(\Sigma u_i)^2}{m}$,

$$S_2 = \Sigma v_i^2 - \frac{(\Sigma v_i)^2}{m}, \quad C_1 = \frac{\left\{ \Sigma u_i v_i - \frac{(\Sigma u_i)(\Sigma v_i)}{m} \right\}^2}{\left\{ \Sigma u_i^2 - \frac{(\Sigma u_i)^2}{m} \right\}} \quad (5.22)$$

and u is to x as v is to y . The first factor of (5.22) is the conditional density of μ_2 given $\beta_{2.1}$ and $\lambda_{2.1}$; the second factor is the conditional density of μ_1 given λ_1 ; the third factor is the marginal density of λ_1 ; the fourth factor is the marginal density of $\lambda_{2.1}$; the fifth factor is the conditional density of $\beta_{2.1}$ given $\lambda_{2.1}$. It is easily seen that $b(\mu_1, \mu_2, \lambda_1, \lambda_{2.1}, \beta_{2.1}) = c$ and $f(x, y) = c$ and we may infer the uniqueness of the NBS as in

previous cases. The usual substitutions in (5.13) yield the posterior Bayes density for $m > 2$. We have assumed $m_0 = 2$ which seems reasonable for this model. Uniqueness may be demonstrated as in previous examples.

From the above analysis it is apparent that any number of regression parameters might be added and we would still be able to obtain a NBS. Also we note that the prior Bayes density obtained for each of the parameters of this model is consistent with the prior Bayes density for that parameter in a simpler model. This does not mean that the two marginal densities are the same, indeed they are not. What is meant is the following. If we obtain a prior density of μ in the one parameter model it will be identical with the prior conditional density of μ given λ in the two parameter model.

In the case of the location parameter μ and scale parameter λ we find (see 5.8 and the first factor of 5.11) that the difference in the posterior Bayes densities of λ may be described as a difference of "one degree of freedom". Similarly in (5.20) we see that there has been a loss of a second "degree of freedom" as the result of the introduction of the regression parameter β . Following through the details of the derivation of the NCBD it is readily seen that there will be a loss of one degree of freedom for each regression parameter introduced.

5.7 A Log - Normal Model

In order to exhibit one result of an allowable transformation on the space of the random variable we shall consider a simple case of an important model which is obtained by transformation from the location

parameter normal model. We assume the model density

$$f_{\mu}(x) = x^{-1}(2\pi)^{-1/2} \exp\left[-\frac{(\log x - \mu)^2}{2}\right] \quad (5.23)$$

for $x > 0$, $-\infty < \mu < \infty$. The joint likelihood of n observations is

$$f_{\mu}(x) = (\pi x_i)^{-1} (2\pi)^{-n/2} \exp\left[-\frac{1}{2} \sum_{i=1}^n (\log x_i - \mu)^2\right] \quad (5.24)$$

We define the NCBD as

$$\begin{aligned} b_m(\mu) &= \frac{e^{-\frac{m}{2}\left(\mu^2 - 2\mu \frac{\sum \log z_i}{m}\right)}}{\int e^{-m/2\left(\mu^2 - 2\mu \frac{\sum \log z_i}{m}\right)} d\mu} \\ &= \frac{\sqrt{m}}{\sqrt{2\pi}} e^{-m/2\left(\mu - \frac{\sum \log z_i}{m}\right)^2}, \quad \text{for } m > 0. \quad (5.25) \end{aligned}$$

The fiducial density is then

$$\begin{aligned} f_m(x) &= \int_{-\infty}^{\infty} x^{-1}(2\pi)^{-1/2} e^{-1/2(\log x - \mu)^2} \frac{\sqrt{m}}{\sqrt{2\pi}} e^{-m/2\left(\mu - \frac{\sum \log z_i}{m}\right)^2} d\mu \\ &= \sqrt{\frac{m}{m+1}} \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{m}{2(m+1)}\left(\log x - \frac{\sum \log z_i}{m}\right)^2}. \end{aligned}$$

For null values $m = 0$, $\sum \log z_i = 0$, we have $b(\mu) = c$, $f(x) = x^{-1}$.

For $m > 0$ both (5.24) and (5.25) are proper densities. The uniqueness of this NBS then follows as in 5.2. If X has the above log-normal density then $Y = \log X$ has the normal density with mean μ . The NBS is, in this case, invariant under transformation on X .

CHAPTER VI
POISSON MODELS

The Poisson density arises as a model density in two distinct contexts. In some cases the context is that of simple random sampling while in others it is in terms of the observation of a continuous Poisson process. We shall begin with a model in the former context but shall extend our results to cover more general problems.

6.1 Random Poisson Sampling

We have the model density

$$m_{\lambda}(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad , \quad \lambda > 0, \quad x = 0, 1, 2, \dots \quad . \quad (6.1)$$

The joint likelihood of n observations is then

$$m_{\lambda}(\underline{x}) = \frac{e^{-n\lambda} \lambda^{\sum x_i}}{n \prod_{i=1}^n x_i!} \quad . \quad (6.2)$$

We define the NCBD to be the gamma density

$$b_m(\lambda) = \frac{m^{1+z} \lambda^z e^{-m\lambda}}{\Gamma(1+z)} \quad , \quad \text{for } m > 0, \quad z \geq 0, \quad z = \sum z_i, \quad \lambda > 0. \quad (6.3)$$

The fiducial density is

$$f_m(x) = \frac{m^{1+z} \Gamma(x+z+1)}{(m+1)^{x+z+1} \Gamma(x+1) \Gamma(z+1)} \quad \text{for } x = 0, 1, 2, \dots \quad , \quad (6.4)$$

which is for suitable values of (x, z) , a negative binomial density. The quantities m and z are interpreted as the prior sample size and prior number of occurrences of the defining event. We have extended the domain of the NCED to include non-integral m and z .

The ratio $b_m(\lambda_1)/b_m(\lambda_2)$ is unity for $m = z = 0$, hence define $b(\lambda) = c$. The corresponding fiducial density is $f(x) = c$. Now (6.3) and (6.4) will be proper for any $m > 0$; hence postulating the value $m_0 = 0$, we see that the MNS principle is satisfied. Since the prior Bayes and fiducial densities are uniform the MBI and MFI principles are satisfied. Hence we have a NBS which is unique by virtue of the moment theorem argument given in 4.6. The posterior Bayes and fiducial densities are given by (6.3) and (6.4) with (m, z, \hat{x}) replaced by (n, x, \hat{x})

6.2 Poisson Process Model

We may generalize the above model somewhat, in the manner of Raiffa and Schlaifer. Suppose we were observing a Poisson process, say Poisson arrivals at a queue, and suppose we observed the process during random time intervals of variable length. It would be convenient to have a simple method of combining these data with data in a prior Bayes density. To accomplish this we define the model density as

$$m_{\lambda, t}(x) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}, \quad t > 0, \quad x = 0, 1, 2, \dots, \quad \lambda > 0 \quad (6.5)$$

where x is the total number of occurrences of the defining event (arrivals) and t is the total time of observation.

The NCED is then defined as

$$b_T(\lambda) = \frac{T^{z+1} \lambda^z e^{-\lambda T}}{\Gamma(z+1)}, \quad \lambda > 0, T > 0, z \geq 0. \quad (6.6)$$

The fiducial density is then

$$f_T(x) = \frac{T^{z+1} t^z \Gamma(x+z+1)}{\Gamma(x+1)\Gamma(z+1)(t+T)^{x+z+1}}. \quad (6.7)$$

The quantities (T, z) are thought of as the total time of prior observation and total number of prior occurrences of the defining event. The ratio $b_T(\lambda_1)/b_T(\lambda_2) = (\lambda_1/\lambda_2)^z e^{-T(\lambda_1-\lambda_2)}$ is unity for $z = T = 0$; hence we define $b(\lambda) = c$ and obtain $f(x) = c$. The proof that this is the unique NBS is as in 6.1. The posterior Bayes and fiducial densities are given by (6.6) and (6.7) with (T, z, t, x) replaced by (t, x, \hat{t}, \hat{x}) , where \hat{t}, \hat{x} are the total time of the next observation and the total number of new occurrences.

6.3 Discussion of the Indifference Rule for the Poisson Model

The result obtained here for the prior Bayes density differs from that proposed by Jeffreys, who would take $b(\lambda) = \lambda^{-1}$. Employing Jeffreys' rule the posterior Bayes density after n observations is

$$b_n(\lambda) = \frac{n^x \lambda^{x-1} e^{-n\lambda}}{\Gamma(x)}. \quad (6.8)$$

This density is improper and (Shannon) uninformative for $x = 0$ regardless of how large n might be. Under the model assumptions made here, i.e. that the underlying process is definitely Poisson with λ strictly greater than zero, and having observed a "great many" observations, or having observed a Poisson process for a "long period of

time" and having observed no occurrences it would seem that a small probability should be assignable to the set $\lambda > \lambda_0$ (say) for an appropriately chosen λ_0 depending on n . Jeffreys' rule does not permit this and hence we must conclude that Jeffreys' rule is inappropriate for the problem as we have formulated it.

With the model (6.5) and the prior Bayes density $b(\lambda) = \lambda^{-1}$ the posterior Bayes density will be

$$b_t(\lambda) = t e^{-t\lambda} \quad (6.9)$$

if $x = 1$. Since

$$\lim_{t \rightarrow 0} e^{-t(\lambda_1 - \lambda_2)} = 1$$

we see that the Bayes density $b(\lambda) = c$ is obtained as a posterior density when the prior Bayes density is $b(\lambda) = \lambda^{-1}$ and the defining event occurs instantaneously after we have begun to study the process. Also, given $b(\lambda) = \lambda^{-1}$ a priori and $x=1$, the posterior Bayes density is minimally informative (that is, uniform) only if $t=0$. It would appear then the Jeffreys law may be appropriate when we have not yet established that $\lambda > 0$ while the rule given in this paper is appropriate under the model assumption that the process is definitely Poisson with $\lambda > 0$. This idea is further developed in 7.4.

The negative exponential density is often derived as a waiting time distribution from a Poisson process. In this context the experimental density is determined by conditional stopping and hence should not be taken as the informational model; rather the prior distribution derived above should be employed. When the negative exponential density is the informational density the analysis should be as given in Chapter IX.

CHAPTER VII
MULTINOMIAL MODELS

7.1 Binomial Models

In this chapter we deal first with the example which was the initial subject of Bayes postulate. At least three prior Bayes densities have been proposed and remain seriously considered for the binomial proportion. We assume the model density

$$m_p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad 0 < p < 1, \quad x = 0, 1, 2, 3, \dots, n, \\ n = 1, 2, \dots \quad . \quad (7.1)$$

The statement of the model density in this form is analogous to the second treatment of the Poisson model. In each case we have written the model in terms of a sufficient statistic with respect to the more basic underlying process, and in each case it is this sufficient statistic which is usually of interest in applications.

Proceeding in this framework we define the NCBD to be a beta density of the first kind given by

$$b_m(p) = \frac{p^z (1-p)^{m-z}}{\beta(z+1, m-z+1)} \quad \text{for } 0 < p < 1, \quad m \geq z \geq 0 \quad (7.2)$$

with expectation

$$E_{m,z}(p) = \frac{z+1}{m+2} \quad . \quad (7.3)$$

The fiducial density is then

$$\begin{aligned}
 f_m(x) &= \int_0^1 \binom{n}{x} p^x (1-p)^{n-x} \frac{p^z (1-p)^{m-z} dp}{\beta(z+1, m-z+1)} \\
 &= \frac{\Gamma(m+1) \Gamma(m+2) \Gamma(x+z+1) \Gamma(m+m-z-x+1)}{\Gamma(x+1) \Gamma(n-x+1) \Gamma(z+1) \Gamma(m-z+1) \Gamma(m+n+2)} \quad (7.4)
 \end{aligned}$$

which Raiffa and Schlaifer have called the beta-binomial density.

Applying the MPD principle we take $m = z = 0$ which yields the prior Bayes density $b(p) = 1$, i.e. uniform and hence minimally informative. The prior fiducial density is then $f(x) = (n+1)^{-1}$, again uniform and minimally informative. Now for any fixed n it would be possible to reparameterize in a manner such that the MPD principle would imply a MBI density and MFI density. The only restriction would be that the new parameter be such that a uniform density on it was equivalent to a density on p which had its first n moments equal to the first n moments of the uniform density. Consider the following equations defined for $n = 2$ and an allowable parameter θ with inverse transformation $\pi(\theta)$:

$$\frac{1}{3} = \int \binom{2}{x} [\pi(\theta)]^x [1 - \pi(\theta)]^{2-x} b(\theta) d\theta, \text{ for } x=0,1,2.$$

If $b(\theta)$ is uniform then

$$\text{for } x = 0, \quad \frac{1}{3} = 1 - 2 \int \zeta(p) + \int \zeta(p^2);$$

$$\text{for } x = 1, \quad \frac{1}{3} = 2 \int \zeta(p) - \int \zeta(p^2);$$

$$\text{for } x = 2, \quad \frac{1}{3} = \int \zeta(p^2).$$

Hence $\int_0^1 (p) = \frac{1}{2}$, $\int_0^1 (p^2) = \frac{1}{3}$, and in general the restriction on moments is as given above. The densities (7.2) and (7.4) will be more than minimally informative for $m > 0$, hence the prior law $b(p) = 1$ defines a NBS, but not uniquely. Under this specification the posterior Bayes and fiducial densities are given by (7.2) and (7.4) with (m, z, n, x) replaced by (n, x, \hat{n}, \hat{x}) .

For $n = n+1$ and $x = n$ we have

$$f_n(\hat{n}) = \frac{\Gamma(n+2) \Gamma(n+2) \Gamma(2n+2) \Gamma(1)}{\Gamma(n+2) \Gamma(1) \Gamma(n+1) \Gamma(1) \Gamma(2n+3)} = \frac{(n+1)}{(2n+2)} = \frac{1}{2}. \quad (7.5)$$

This states that following n successes in n trials the fiducial probability that the next $n+1$ trials will also be successes is exactly $\frac{1}{2}$, independently of n . This statement has been termed the law of succession and has received almost universal rejection as a reasonable assessment of the probability that might be proper in this case (Jeffreys, 1961).

It is claimed that this value is too low and that having observed n successes in n trials, for large n the correct probability of the next $n+1$ being successes should be very close, but not equal to, one and that this probability should approach one as $n \rightarrow \infty$. The two other prior Bayes densities which have been proposed and remain seriously considered give substantially different values for this probability for all n (thus dispelling the notion that the choice of the prior density is unimportant when n is large). The first of these, due to Haldane, is $b(\theta) = c$ where $\theta = -\log p/1-p$. This corresponds to a density

for p given by (7.2) with $m=z=0$. This law has intuitive appeal in that it represents the limit of the class of beta densities just as the uniform was the limit of the class of proper normal densities. Haldane's law assigns probability one independently of n , for the above event. This is, of course, far too high a value, particularly for small values of n . The third law (Perks, 1947) is $b(\lambda) = c$ where $\lambda = \sin^{-1} \sqrt{p}$, which corresponds to (7.2) with $m = z = \frac{1}{2}$ and which gives intermediate values for the above event. Each of these parameters could define a NBS for $m=1$.

The problem of the unique specification of the NBS and the problem of the law of succession require careful consideration. At this point we shall present an argument suggesting that $b(p) = 1$ yields the only satisfactory NBS. Later (Section 7.4) we shall present another argument which will suggest the correctness of the evaluation $f_n(\hat{n}) = \frac{1}{2}$ for $\hat{n} = n+1$ and $x=n$. Finally we shall present another argument which shows that the law $b(p) = 1$ is the only one which is consistent with the result obtained for the Poisson model.

The density (7.1) differs from the normal examples and the initial treatment of the Poisson model in that it represents an entire family of densities dependent upon the index, n , of the model. In this sense it is related to the second Poisson model which also represents an entire family of densities dependent upon the index t . It certainly seems requisite that the choice of prior density for p be independent of the particular binomial model (index n) considered. The requirement then is that the prior Bayes density have all of its moments equal to the moments of the uniform density. Since the

moment theorem is clearly applicable the prior law $b(p)=1$ is uniquely specified. This admittedly is a somewhat ad hoc argument though, we feel, a compelling one. The class of multinomial models is the only class of models considered requiring any such argument. More generally, if the class of likelihood functions is not complete, as in this case, it may be necessary to enlarge the class in order to make it complete.

7.2 Trinomial Model

To clarify the questions arising from the binomial problem it will be necessary to study a more general model of which the following is a second special case. We assume the model density

$$m_{p_1, p_2}(x_1, x_2) = \binom{n}{x_1 \ x_2 \ n-x_1-x_2} p_1^{x_1} p_2^{x_2} (1-p_1-p_2)^{n-x_1-x_2}, \quad (7.6)$$

$x_1+x_2 \leq n$ $p_1 \leq 1-p_2$, $n = 1, 2, 3, \dots$. We define the NCBD as

$$b_m(p_1, p_2) = \frac{p_1^{z_1} p_2^{z_2} (1-p_1-p_2)^{m-z_1-z_2}}{\int_0^1 \int_0^{1-p_1} p_1^{z_1} p_2^{z_2} (1-p_1-p_2)^{m-z_1-z_2} dp_1 dp_2}$$

for $z_1 + z_2 \leq m$, $m \geq 0$. The transformation $r = p_1/1-p_2$, $s = p_2$ with Jacobian $(1-s)$ yields the bivariate beta density

$$b_m(p) = \frac{p_1^{z_1} p_2^{z_2} (1-p_1-p_2)^{m-z_1-z_2}}{\beta(z_1+1, m-z_1-z_2+1)\beta(z_2+1, m-z_2+2)} \quad (7.7)$$

For $m = z_1 = z_2 = 0$ we have

$$b(p) = 1/\beta(1,1) \cdot \beta(1,2) = 2, \quad p_1 \leq 1 - p_2. \quad (7.8)$$

Thus the joint density of (p_1, p_2) is uniform and hence the MBI principle is satisfied. A similar derivation yields the fiducial density

$$f_m(x_1, x_2) = \frac{\Gamma(x_1+z_1+1) \Gamma(x_2+z_2+1) \Gamma(m-x_1-x_2+m-z_1-z_2+1) \Gamma(m+1) \Gamma(m+3)}{\Gamma(x_1+1) \Gamma(x_2+1) \Gamma(n-x_1-x_2+1) \Gamma(z_1+1) \Gamma(z_2+1) \Gamma(n-z_1-z_2+1) \Gamma(n+m+3)} . \quad (7.9)$$

For $m = z_1 = z_2 = 0$ we have

$$f(x_1, x_2) = 2/(n+1)(n+2) \quad (7.10)$$

i.e. constant on the triangular lattice $x_1 + x_2 \leq n$, and hence the MFI principle is satisfied. Assuming the value $m_0 = 0$ we see that the MNS principle is satisfied by noting that (7.7) and (7.9) will be more than minimally informative for $m > 0$. The uniqueness of this NBS may be demonstrated following the requirement and method of 7.1. Under the NBS specification the posterior Bayes and fiducial densities are given by (7.7) and (7.9) with $(m, z_1, z_2, n, x_1, x_2)$ replaced by $(n, x_1, x_2, \hat{n}, \hat{x}_1, \hat{x}_2)$.

The marginal Bayes density of p_i is

$$b_m(p_i) = \frac{p_i^{z_i} (1-p_i)^{m-z_i+1}}{\beta(z_i+1, m-z_i+2)} , \quad i = 1, 2. \quad 0 < p_i < 1 \quad (7.11)$$

which, for $m = z_i = 0$ reduces to

$$b(p_i) = 2(1-p_i) . \quad (7.12)$$

The fiducial density of X_i corresponding to (7.12) is

$$f(x_i) = \sum_{x_j} \frac{2}{(n+2)(n+1)} (n - x_i + 1) = \frac{2}{(n+2)} \left(1 - \frac{x_i}{n+1}\right). \quad (7.13)$$

The results (7.12) and (7.13) differ from those obtained under the assumption of a binomial model for X_i , and will be found to be different from those obtained from higher order multinomial models. Referring these models to the popular examples involving balls in an urn, it seems to us that our prior density for p_i should be different for different assumptions as to the nomiality, (i.e. the number of categories) of the model. If we assume a binomial model, i.e. assume that there are exactly two types of balls in the urn, then our prior Bayes density should certainly be such as to assign more probability to higher values of p_i than if our model were such as to assume that there were 100, say, different kinds of balls in the urn. Indeed assuming a model of very high nomiality, we would need to assign a prior Bayes density to p_i which concentrated most of the probability near zero if we were to satisfy any of the symmetry ideas of Bayes postulate. We shall see that by stipulating that the prior Bayes density be uniform over the joint set of all parameters, as we have for our MBI principle, the marginal prior density of any p_i will behave properly as the nomiality of the model increases.

7.3 The General Multinomial Model

Employing generalizations of the transformation employed in 7.2 it is possible to obtain a NBS for a general multinomial model. In this section we shall omit consideration of the normalizing constants and

shall present only the kernels of the various Bayes densities. This will prove sufficient to establish the existence of the NBS and for our general discussion in 7.4.

The model density is

$$m_p(x_1, x_2, \dots, x_{s-1}) = \binom{n}{x_1, x_2, \dots, x_{s-1}, 1 - \sum_{i=1}^{s-1} x_i} p_1^{x_1} p_2^{x_2} \dots p_{s-1}^{x_{s-1}} (1 - \sum_{i=1}^{s-1} p_i)^{1 - \sum_{i=1}^{s-1} x_i} \quad (7.14)$$

where s is the nomality of the model. The NCBD, $b_m(p_1, \dots, p_{s-1})$, is then proportional to

$$p_1^{z_1} p_2^{z_2} \dots p_{s-1}^{z_{s-1}} (1 - \sum_{i=1}^{s-1} p_i)^{n - \sum_{i=1}^{s-1} z_i} \quad (7.15)$$

For $m = z_1 = \dots = z_{s-1} = 0$ we have

$$b(p_1, p_2, \dots, p_{s-1}) = c, \quad \text{for } \sum_{i=1}^{s-1} p_i < 1, \quad (7.16)$$

and

$$f(x_1, x_2, \dots, x_{s-1}) = c, \quad \text{for } \sum_{i=1}^{s-1} x_i \leq n. \quad (7.17)$$

Corresponding to (7.16) the marginal Bayes density of p_i is

$$b(p_i) = (s-1) (1-p_i)^{s-2} \quad i = 1, 2, \dots, s-1, \quad (7.18)$$

and the expectation of p_i corresponding to (7.15) is

$$E(p_i) = \frac{z_i + 1}{m + s} \quad i = 1, 2, \dots, s-1. \quad (7.19)$$

For $m = z_i = 0$ we have

$$\mathcal{E}(p_i) = s^{-1}. \quad (7.20)$$

The behavior of the marginal density (7.18) and the expectation (7.20) seems eminently reasonable in view of the differing assumptions of the models of differing nomiality.

7.4. Validation and Semantic Interpretation of the Multinomial Models

We now propose a heuristic explanation of the correctness of the law $b(p) = 1$ in the binomial case, the result obtained from it for the law of succession and the fact that we do not approach the limit of the class of beta densities to obtain the law for our indifference rule. We have continually emphasized the fact that the indifference procedure suggested in this paper is applicable only in a parametric model formulation. Part of this parametric formulation involves the definition of a class of model densities, which definition is based on prior information. Thus the model density, that is the conditional density of X given θ , and the prior Bayes density are both carriers of our prior knowledge.

The assumption of a binomial model contains certain information, namely that there are exactly two types of balls, say, in an urn. One way of having determined that there are two types is to have previously observed one of each type. Thus the prior law $b(p) = 1$ for the binomial model may be thought of as expressing knowledge equivalent to that which would be obtained from having observed just one of each type, and having these observations used to define the binomiality of the model. Conditional on having observed at least one of each type the

minimum information is present if just exactly one of each kind has been observed. Hence the assumption of the model (7.1) and the prior density $b(p) = 1$ may be considered to be equivalent to the assumption of having started with an urn about whose contents we knew nothing at all and then having observed just two balls, each of different type. The interpretation here is analogous to that of 6.3.

Formulas (7.3), (7.18) and (7.19) are consistent with and suggestive of this interpretation. Considering (7.19), if the first s observations have defined the s categories of the multinomial model, there would have been just one in the i -th category. Hence with m additional observations the numerator of (7.19) would be the total number of type i observed and the denominator would be the total number observed in all categories.

Perhaps then, an explanation for the fact that the prior beta density furnishing the NBS for the binomial model is not the limiting density of the beta class is that the model has incorporated information obtained from two observations. In general the marginal density of p_i , as given by 7.18, is dependent upon the nomiality of the model, or by this interpretation, upon the number of prior observations used to define the model. The extension of this idea to the general multinomial model is mirrored in the beta density (7.18).

This heuristic interpretation of the NBS for the multinomial model permits a reevaluation of the statement (7.5). It had been said that after observing n successes in n Bernoulli trials and having no other data, the probability of the next $n+1$ being successes would be

1/2, independent on n . If we accept the fact that the assumption of binomiality is equivalent to the assumption of the availability of one observation of each type, we may restate the above evaluation as follows:

Given an urn about which we know nothing a priori and the fact that in $n+2$ trials we have observed $n+1$ reds and one black, the probability that the next n draws will all be reds is $1/2$.

Perhaps this last statement may be accepted as a reasonable evaluation.

One final argument will be presented to support the prior law $b(p) = 1$ and to show its consistency with the prior law $b(\lambda) = c$ obtained for the Poisson model. This argument is independent of the heuristic considerations of the preceding argument. We have seen that when considering a simple Bernoulli model, i.e. a binomial model with index one, any prior law with expectation $1/2$ would furnish a NBS. We then argued that this process was simply a special case of the more general binomial model with arbitrary index, n , and any prior density selected should furnish a NBS for an arbitrary index. Carrying this argument to its limit we write

$$\begin{aligned} m_p(x) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \binom{n}{x} \cdot \left(\frac{n-1}{n}\right) \left(\frac{n-2}{n}\right) \dots \left(\frac{n-x+1}{n}\right) \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \end{aligned}$$

where $\lambda = np$. We then have the usual Poisson limit of the binomial (as $n \rightarrow \infty$),

$$m_\lambda(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Clearly then the prior laws $b(\lambda) = c$ and $b(p)=1$, both uniform, are consistent. Each indeed logically requires the other.

CHAPTER VIII

UNIFORM AND RELATED MODELS

The preceding three chapters have dealt with three of the most frequently used classes of statistical models. With this chapter we begin the study of models of somewhat lesser frequency of application. However, from a theoretical point of view it is important to see that the principles developed and applied to the most basic models are also applicable to some other models. In this chapter we shall consider some uniform models and one related model, and in the following chapter we shall consider some negative exponential and related models.

8.1 Uniform Model - Scale Parameter Unknown

We assume that the random variable X is uniformly distributed on the interval $(0, e^\mu)$, for $-\infty < \mu < \infty$. The model density is then

$$m_\mu(x) = e^{-\mu} U(e^\mu - x) U(x) \quad , \quad (8.1)$$

where $U(y)$ is the Heavyside unit function defined by

$$\begin{aligned} U(y) &= 1 && \text{for } y > 0 \\ &= 0 && \text{for } y \leq 0 . \end{aligned}$$

We have chosen this unusual parametrization as it is the one necessary to obtain a NBS. The joint likelihood of n observations is

$$m_\mu(x) = e^{-n\mu} U(e^\mu - x^*) U(x_*) \quad (8.2)$$

where $x^* = \max x_i$ and $x_* = \min x_i$. We then define the NCBED as

$$b_m(\mu) = \frac{e^{-m\mu} U(e^\mu - z^*)}{\int e^{-m\mu} U(e^\mu - z^*) d\mu}$$

$$= m z^{*m} e^{-m\mu} U(e^\mu - z^*), \quad (8.3)$$

for $m > 0$. The fiducial density of X is

$$f(x) = \int e^{-\mu} U(e^\mu - x) U(x) m z^{*m} e^{-m\mu} U(e^\mu - z^*) d\mu$$

$$= m z^{*m} U(x) \int e^{-(m+1)\mu} U(e^\mu - [x, z^*]) d\mu$$

for $m > 0$, where $[x, z^*]$ indicates the maximum of x and z^* .

Thus, for $m > 0$

$$f(x) = \left(\frac{m}{m+1}\right) \frac{z^{*m}}{([x, z^*])^{m+1}}. \quad (8.4)$$

We interpret m to be the number of prior observations and z^* to be the maximum of the prior observations. Applying the MPD principle we define $b(\mu)$ to be uniform on $(-\infty, \infty)$. The corresponding fiducial density is $f(x) = x^{-1}$. Now both (8.3) and (8.4) will be proper for $m > 0$. The completion of the proof that this constitutes a NBS is as in 5.2. Under the NBS the posterior Bayes and fiducial densities are given by (8.3) and (8.4) with (m, z^*, x) replaced by (n, x^*, \hat{x}) . The

posterior Bayes density is of exponential form on $(\log x^*, \infty)$ with increasing probability being concentrated near x^* as $n \rightarrow \infty$.

The posterior fiducial density is uniform on $(0, x^*)$ and proportional to x^{*n}/\hat{x}^{n+1} for $\hat{x} > x^*$. As $n \rightarrow \infty$ the amount of probability on (x^*, ∞) approaches zero. These forms seem very attractive intuitively.

8.2 Uniform Model - Location Parameter Unknown

We assume the model

$$m_{\theta}(x) = e^{-\mu} U(\int \theta + e^{\mu} \int - x) U(x - \theta), \quad (8.5)$$

where μ is considered known. The joint likelihood of n observations is

$$m_{\theta}(\underline{x}) = e^{-n\mu} U(\int \theta + e^{\mu} \int - x^*) U(x_* - \theta). \quad (8.6)$$

We define the NCBD by

$$\begin{aligned} b_m(\theta) &= \frac{U(\int \theta + e^{\mu} \int - z^*) U(z_* - \theta)}{\int U(\int \theta + e^{\mu} \int - z^*) U(z_* - \theta) d\theta} \\ &= \frac{U(\int \theta + e^{\mu} \int - z^*) U(z_* - \theta)}{\int_{z_*^* - e^{\mu}}^{z_*} d\theta} \\ &= \frac{U(\theta - \int z^* - e^{\mu} \int) U(z_* - \theta)}{(z_* - \int z^* - e^{\mu} \int)} \end{aligned} \quad (8.7)$$

for $m > 0$, i.e. uniform on $z^* - e^{\mu}$ to z_* . The fiducial density is

$$\begin{aligned}
f(x) &= \int e^{-\mu} \bigcup (\int \theta + e^{\mu} \int - x) \bigcup (x - \theta) \frac{\bigcup (\theta - \int z^* - e^{\mu} \int) \bigcup (z_* - \theta)}{(z_* - \int z^* - e^{\mu} \int)} d\theta \\
&= \frac{e^{-\mu}}{(z_* - \int z^* - e^{\mu} \int)} \int_{\int x, z^* \int - e^{\mu}}^{\int x, z_* \int} d\theta \\
&= \frac{e^{-\mu} (\int x, z_* \int - \int x, z^* \int + e^{\mu}) \bigcup (\int z_* + e^{\mu} \int - x) \bigcup (x - \int z^* - e^{\mu} \int)}{(z_* - \int z^* - e^{\mu} \int)}
\end{aligned} \tag{8.8}$$

i.e. uniform on (z_*, z^*) and decreasing on $(z^*, z_* + e^{\mu})$ and increasing on $(z^* - e^{\mu}, z_*)$. We interpret m and z as in case 1. Null values of (m, z_*, z^*) are $(0, \infty, -\infty)$ hence define $b(\theta) = f(x) = c$. For $m > 0$ (8.7) and (8.8) are proper densities, hence postulating $m_0 = 0$ we have a NBS. Uniqueness follows as in 5.1. The posterior Bayes and fiducial densities are given by (8.7) and (8.8) with (m, z_*, z^*, x) replaced by (n, x_*, x^*, \hat{x}) .

8.3 Uniform Model - Location and Scale Parameter Unknown

We assume the model density

$$m_{\theta, \mu}(x) = e^{-\mu} \bigcup (\int \theta + e^{\mu} \int - x) \bigcup (x - \theta). \tag{8.9}$$

The joint likelihood of n observations is

$$m_{\theta, \mu}(\underline{x}) = e^{-n\mu} \bigcup (\int \theta + e^{\mu} \int - x^*) \bigcup (x_* - \theta). \tag{8.10}$$

We define the NCBD to be

$$\begin{aligned}
b_m(\theta, \mu) &= \frac{e^{-m\mu} \Gamma(\theta + e^\mu - z^*) \Gamma(z_* - \theta)}{\iint e^{-m\mu} \Gamma(\theta + e^\mu - z^*) \Gamma(z_* - \theta) d\mu d\theta} \\
&= (m(m-1)(z^* - z_*)^{m-1} e^{-m\mu}) (\Gamma(\theta + e^\mu - z^*) \Gamma(z_* - \theta))
\end{aligned} \tag{8.11}$$

for $m > 1$. The second factor is a kernel of the conditional density of θ given μ , and the first factor is a kernel of the marginal density of μ . The fiducial density of X is defined by

$$\begin{aligned}
f_m(x) &= m(m-1)(z^* - z_*)^{m-1} \int \Gamma(x, z_* - \theta) \int_{\log(\Gamma(x, z^*) - \theta)}^{\infty} e^{-(m+1)\mu} d\mu d\theta \\
&= \frac{m(m-1)}{(m+1)} (z^* - z_*)^{m-1} \int \Gamma(x, z_* - \theta) \Gamma(x, z^*)^{-\theta} \Gamma^{-(m+1)} d\theta \\
&= \frac{(m-1)}{(m+1)} \frac{(z^* - z_*)^{m-1}}{(\Gamma(x, z^*) - \Gamma(x, z_*))^{m-1}}, \text{ for } m > 1.
\end{aligned} \tag{8.12}$$

Following the methods of 8.1 and 8.2 we define $b(\theta|\mu) = c$, $b(\theta, \mu) = c$, $b(\mu) = c$, $f(x) = c$, $f_1(x) = |x_1 - z|^{-1}$. (Note that $(z^* - z_*) = 0$ for $m = 1$.) The proof of the uniqueness of this NBS follows as in 5.3. The posterior Bayes and fiducial densities are given by (8.11) and (8.12) with (m, z_*, z^*, x) replaced by (n, x_*, x^*, \hat{x}) . It is interesting to note that the NBS involves a rather unusual parameterization.

8.4 A Positive Exponential Model

We assume the model density

$$m_{\mu}(x) = (1+p) x^p e^{-(1+p)\mu} \cup(x) \cup(e^{\mu} - x), \text{ for } p \geq 0. \quad (8.13)$$

The interesting feature of this model is that the density function is an increasing function of x throughout the spectrum of X , in contrast to previous models. For $p = 0$, (8.13) reduces to (8.1). The joint likelihood of n observations is

$$m_{\mu}(\underline{x}) = (1+p)^n (\prod x_i)^p e^{-n(1+p)\mu} \cup(x_*) \cup(e^{\mu} - x^*). \quad (8.14)$$

We define the NCBD by

$$\begin{aligned} b_m(\mu) &= \frac{e^{-m(1+p)\mu} \cup(e^{\mu} - z^*)}{\int e^{-m(1+p)\mu} \cup(e^{\mu} - z^*) d\mu} \\ &= m(1+p) z^{*m(1+p)} e^{-m(1+p)\mu} \cup(\mu - \log z^*), \end{aligned} \quad (8.15)$$

for $m > 0$, $z^* > 0$. The fiducial density is given by

$$\begin{aligned} f_m(x) &= \int m(1+p)^2 z^{*m(1+p)} x^p e^{-(m+1)(1+p)\mu} \cup(x) \cup(e^{\mu} - \underline{z}^*, x \underline{z}^*) d\mu \\ &= \binom{m}{m+1} \frac{(1+p) z^{*m(1+p)} \cup(x)}{(z^*, x)^{*(m+1)(1+p)}} x^p, \text{ for } m > 0, z^* > 0. \end{aligned} \quad (8.16)$$

For $m = z^* = 0$ we have $b(\mu) = c$, $f(x) = x^{-1}$. The proof that this is a unique NBS follows as in previous cases. The posterior Bayes and fiducial densities are given by (8.15) and (8.16) with (m, z^*, x)

replaced by (n, x^*, \hat{x}) . The posterior Bayes density has an exponential form on $(\log x^*, \infty)$. The posterior fiducial density has an interesting form, being an increasing function of X on $(0, z^*)$ and a decreasing function of X on (z^*, ∞) .

CHAPTER IX

NEGATIVE EXPONENTIAL AND RELATED MODELS

In this chapter we derive the NBS for three negative exponential models and for the Pareto and Weibull models. It is necessary to repeat here that we are assuming that the given experimental model is identical to the informational model.

9.1 Negative Exponential Model - Scale Parameter Unknown

We assume the model density

$$m_{\lambda}(x) = e^{\lambda - xe^{-1}} U(x) . \quad (9.1)$$

The joint likelihood of n observations is

$$m_{\lambda}(\underline{x}) = e^{n\lambda - e^{\lambda} \sum x_i} U(x_{*}) . \quad (9.2)$$

We then define the NCBD to be

$$b_m(\lambda) = \frac{e^{m\lambda - e^{\lambda} \sum z_i} (\sum z_i)^m}{\Gamma(m)} , \quad (9.3)$$

for $m > 0$, $\sum z_i > 0$. The fiducial density is then the "t" density

$$f_m(x) = \frac{m(\sum z_i)^m}{(x + \sum z_i)^{m+1}} , \quad (9.4)$$

for $m > 0$, $\sum z_i > 0$. For $m=z=0$ the ratio $b_m(\lambda_1)/b_m(\lambda_2)$ is unity, hence we define $b(\lambda) = c$ and obtain $f(x) = x^{-1}$. We see then that the Bayes and fiducial densities are proper for $m > 0$, $\sum z_i > 0$ and

improper for $m = 0$, $\Sigma z_i = 0$. The completion of the proof that this constitutes a unique NBS follows as in 5.2. Under the specification of the NBS the posterior Bayes and fiducial densities are given by (9.3) and (9.4) with $(m, \Sigma z_i, x)$ replaced by $(n, \Sigma x_i, \hat{x})$.

9.2 Negative Exponential Model - Location Parameter Unknown

We assume the model density

$$m_\theta(x) = e^{\lambda - (x - \theta)} e^{-\lambda} \bigcup (x - \theta). \quad (9.5)$$

The joint likelihood of n observations is then

$$m_\theta(\underline{x}) = e^{n\lambda - e^{-\lambda} \Sigma (x_i - \theta)} \bigcup (x_* - \theta). \quad (9.6)$$

We define the NCBD to be

$$b_m(\theta) = m e^{\lambda - m e^{-\lambda} (z_* - \theta)} \bigcup (z_* - \theta) \quad (9.7)$$

for $m > 0$. The fiducial density is

$$f_m(x) = \left(\frac{m}{m+1}\right) e^{\lambda - e^{-\lambda} \int x - (m+1)(x, z_*)_* + m z_*} \quad (9.8)$$

The ratio $b_m(\theta_1)/b_m(\theta_2) = 1$ for $m = 0$, $z_* = \infty$, hence define $b(\theta) = c$. The corresponding fiducial density is $f(x) = c$. The densities (9.7) and (9.8) are proper for $m > 0$, $z_* < \infty$, while $b(\theta)$ and $f(x)$ are improper. The essential uniqueness of the NBS may be demonstrated as in 5.1. Posterior Bayes and fiducial densities are given by (9.7) and (9.8) with (m, z_*, x) replaced by (n, x_*, \hat{x}) .

9.3 Negative Exponential Model - Location and Scale Parameter Unknown

We define the model density

$$m_{\lambda, \theta}(x) = e^{\lambda - (x - \theta)} e^{-\lambda} \bigcup_{(x - \theta)} . \quad (9.9)$$

The joint likelihood of n observations is

$$m_{\lambda, \theta}(\underline{x}) = e^{n\lambda - e^{-\lambda} \sum (x_i - \theta)} \bigcup_{(x_* - \theta)} . \quad (9.10)$$

We define the NCBED by

$$\begin{aligned} b_m(\lambda, \theta) &= \frac{e^{m\lambda - e^{-\lambda} \sum (z_i - \theta)} \bigcup_{(z_* - \theta)}}{\iint e^{m\lambda - e^{-\lambda} \sum (z_i - \theta)} \bigcup_{(z_* - \theta)} d\theta d\lambda} \\ &= \frac{m(\sum z_i - mz_*)^{m-1} e^{m\lambda - e^{-\lambda} (\sum z_i - m\theta)} \bigcup_{(z_* - \theta)}}{\Gamma(m-1)} \end{aligned} \quad (9.11)$$

for $m > 1$. The fiducial density is then

$$f_m(x) = \frac{m(m-1)}{(m+1)} \frac{(\sum z_i + mz_*)^{m-1}}{((s+x) - (m+1)(x, z_*)_*)^m} \quad (9.12)$$

for $m > 1$. The ratio $b_m(\lambda_1, \theta_1)/b_m(\lambda_2, \theta_2) = 1$ for null values $m = \sum z_i = 0$ and $z_* = \infty$. Thus define $b(\lambda, \theta) = c$ and obtain $f(x) = c$. The conditional density of θ given λ is given by (9.7) for $m > 0$ and $b(\theta|\lambda) = c$ for $m = 0$.

It should be noted that the prior density for the scale parameter in 9.1 and 9.9 does not appear to be consistent with the specification of the prior distribution for the Poisson parameter. The prior densities derived in this chapter should be used only when the models (9.1)

and (9.9) may be considered to be identical to the informational model. This will not generally be the case. When the basic underlying assumptions are those of a Poisson process, the experimental model densities (9.1) and (9.9) are obtained only by employing conditional stopping. Thus neither (9.1) nor (9.9) will coincide with the informational density. When the assumptions of the Poisson process are valid, the choice of prior distribution should be that determined in Chapter VI, namely, $b(\lambda) = c$, noting that the λ of this chapter is the logarithm of the λ of Chapter VI. A similar situation exists in the binomial negative binomial context where the binomial model is similarly employed.

9.4 A Pareto Model

Because of their importance in applications we shall now consider two rather special models. We deal first with a location parameter Pareto model density given by

$$m_{\theta}(x) = \frac{p \theta^p U(x-\theta)}{x^{p+1}}, \quad p > 0, \quad \theta > 0 \quad (9.14)$$

The joint likelihood of n observations is given by

$$m_{\theta}(\underline{x}) = \frac{p^n \theta^{np} U(x_* - \theta)}{(n x_i)^{p+1}}, \quad (9.15)$$

We define the NCBD as

$$b_m(\theta) = \frac{\theta^{mp} U(z_* - \theta)}{\int \theta^{mp} U(z_* - \theta) d\theta} = \frac{(mp+1)\theta^{mp} U(z_* - \theta) U(\theta)}{z_*^{mp+1}} \quad (9.16)$$

from $m > 0$. The corresponding fiducial density is

$$\begin{aligned}
 f_m(x) &= \int \frac{p \theta^p \mathcal{U}(x-\theta)}{x^{p+1}} \frac{\theta^{mp} \mathcal{U}(z_* - \theta)^{(mp+1)} \mathcal{U}(\theta)}{z_*^{mp+1}} d\theta \\
 &= \frac{p^{(mp+1)}}{z_*^{mp+1} x^{p+1}} \frac{\int_{x, z_*} \mathcal{U}^{(m+1)p+1}}{\int (m+1)p+1} . \quad (9.17)
 \end{aligned}$$

In the usual manner we obtain $b(\theta) = f(x) = c$, and infer the uniqueness of this NBS. The posterior Bayes and fiducial densities under the NBS are then given by (9.16) and (9.17) with (m, z_*, x) replaced by (n, x_*, \hat{x}) . The above model is obtainable by a logarithmic transformation on the location-parameter exponential model of section 9.2. The other one-parameter and the two-parameter Pareto models may be treated analogously.

9.5 A Weibull Model

The second special model to be considered is the one parameter Weibull model with density function

$$m_\lambda(x) = p e^{p\lambda} x^{p-1} e^{-x^p} e^{\lambda p} \mathcal{U}(x), \quad p > 0, \lambda > 0. \quad (9.18)$$

The joint likelihood of n observations is

$$m_\lambda(\underline{x}) = p^n e^{np\lambda} (\prod x_i)^{p-1} e^{-e^{\lambda p} \sum z_i^p} \mathcal{U}(\underline{x}_*). \quad (9.19)$$

We define the NCBDF to be

$$b_m(\lambda) = \frac{p(\sum z_i^p)^m e^{mp\lambda - e^{\lambda p}(\sum z_i^p)}}{\Gamma(m)}. \quad (9.20)$$

for $m > 0$. For $p=1$ this reduces to (9.3). The fiducial density is

$$f_m(x) = \int_0^\infty \frac{p e^{p\lambda} x^{p-1} e^{-x^p} e^{\lambda p} (x)^p (\sum z_i^p)^m e^{mp\lambda - e^{\lambda p} (\sum z_i^p)}}{\Gamma(m)}$$

$$= \frac{m p x^{p-1} (\sum z_i^p)^m}{(x^p + \sum z_i^p)^{m+1}}, \quad \text{for } m > 0. \quad (9.21)$$

For $p = 1$ this reduces to (9.4). The ratio $b_m(\lambda_1)/b_m(\lambda_2) =$

$$e^{mp(\lambda_1 - \lambda_2) - \sum z_i (e^{\lambda_1} - e^{\lambda_2})},$$

which is unity for null values $m = \sum z_i^p = 0$. Hence define $b(\lambda) = c$ and obtain $f(x) = x^{-1}$. The uniqueness of this NBS follows as in previous cases. The posterior Bayes and fiducial densities are given by (9.20) and (9.21) with $(m, \sum z_i, x)$ replaced by $(n, \sum x_i, \hat{x})$. The caution given in 9.3 also applies in this case.

CHAPTER X

EXTENSIONS AND REFINEMENTS

10.1 Limitations of the Procedure

We have shown that an essentially unique NBS can be obtained for a multitude of common examples. There are, however, an equal number of models, though they are not frequently employed, which do not seem amenable to this type of analysis. First consider the gamma density

$$m_K(x) = \frac{x^{K-1} e^{-x} U(x)}{\Gamma(K)}, \quad K > 0. \quad (10.1)$$

Following our usual approach we would define

$$b_m(K) = \frac{(\prod z_i)^{K-1} (\Gamma(K))^{-m}}{\int (\prod z_i)^{K-1} (\Gamma(K))^{-m} dK}. \quad (10.2)$$

We would require that the integral in the denominator be obtained in closed form, which does not seem possible. Even if we were to consider K to be restricted to non-negative integral values and consider the integral with respect to counting measure it would not appear that we can obtain a closed form expression. By analogy to the chi-square distribution we may refer to K as a "degrees of freedom" parameter. We have been unable to find a NCBD for any such parameter. We should note however that degrees of freedom parameters generally arise in derived

distributions and not in distributions of the basic random variable. Indeed we seem to be successful in obtaining a NBS in just those cases in which the model can be referred directly to some basic stochastic process.

A second example is illustrated by the model density

$$m_{\lambda, \theta}(x) = \frac{\lambda^K (x-\theta)^{K-1} e^{-\lambda(x-\theta)} \cup(x-\theta)}{\Gamma(K)} \quad (10.3)$$

with $K \neq 1$. This example fails because there does not exist a sufficient statistic of fixed dimensionality for (λ, θ) . Here again there may be some artificiality since there may be some more basic Poisson process involved. This model is also quite troublesome for the classical theory.

The procedure is not directly applicable to some useful truncated models such as the truncated normal. In this respect we must use some care in defining the model. If the basic underlying model is normal but our measurements are restricted by the experiment, then employing the ideas of Chapter II we would obtain our prior Bayes densities from the full normal model, even though our experimental model might be truncated normal. Again we must refer to the more basic underlying process.

If we drop the requirement of the MPD principle we can obtain an essentially unique structuring for any class of models which may be indexed by a location parameter. Since the integral of $m(x-\mu)$ is unity whether the integration is dx or $d\mu$, the assumption of a uniform prior density for μ implies a uniform prior density for X . Also the posterior density is proper after an epsilon of an observation and

hence the MNS principle is satisfied (postulating $m_0 = 0$). Thus the above gamma model, the Laplace (double exponential), and the Cauchy model admit of reasonable structuring even though there is no sufficient statistic of fixed dimensionality.

10.2 Evaluation of the Four Principles.

The MPD, MBI, MFI and MNS principles in their current formulation have been reasonably satisfactory in that they have clearly indicated essentially unique specifications of the prior Bayes density for a number of important examples. The MFI and MNS principles are somewhat less than elegant. The major problem with the MFI principle is that we do not have an adequate measure of information for discriminating between improper densities. While the discriminations made in this paper are unlikely to raise violent objection it would be desirable to have some more general statement. The problem with the MNS principle is that the value m_0 must be postulated for each case. Again, while we have taken this to be rather apparent for the cases studied a general theory would be preferred. Finally, it has been necessary to introduce some additional, though again persuasive, ideas to obtain uniqueness for the structuring in the multinomial cases. Future research should be directed towards further explicating these principles.

10.3 Transformations of the Random Variable.

For most applications the scale of measurement of the random variable may be considered fixed, at least up to a linear transformation. We may measure height in inches or in feet or even in meters but we are unlikely to measure it in any scale which is not a linear transformation of the scale of inches. In the discrete case it is clear that we need

not, even for theoretical purposes, consider transformations on X . In the continuous case the Lebesgue measure is dependent upon the coordinate system chosen and hence considerations of non-linear transformation on X may be of some theoretical interest. In addition, there are applications, in optics for example, in which there may be a legitimate choice among random variables which are not linearly related.

In studying the effect of transformations on X on the mode of analysis suggested in this paper we shall again restrict ourselves to allowable transformations $y = \varphi(x)$ with inverses $x = \tau(y)$ and to real-valued X . If $m_\theta(x)$ is the model density for X , then

$$t_\theta(y) = m_\theta[\tau(y)] \tau'(y) \quad (10.4)$$

is the model density for Y . Given the model density for Y , the NCBD for θ is

$$\begin{aligned} b_m(\theta) &= \frac{\prod_{i=1}^m t_\theta(w_i)}{\int_{\textcircled{a}} \prod_{i=1}^m t_\theta(w_i) dt} = \frac{\prod_{i=1}^m m_\theta(\tau(w_i)) \tau'(w_i)}{\int_{\textcircled{b}} \prod_{i=1}^m m_\theta(\tau(w_i)) \tau'(w_i) d\theta} \\ &= \frac{\prod_{i=1}^m m_\theta(z_i)}{\int_{\textcircled{c}} \prod_{i=1}^m m_\theta(z_i) d\theta}, \quad \text{where } w \text{ is to } y \text{ as } z \text{ is to } x. \end{aligned}$$

Hence the NCBD is invariant under φ . The fiducial density of Y is given by

$$g_m(y) = \int t_\theta(y) b_m(\theta) d\theta = \tau'(y) \int m_\theta(x) b_m(\theta) d\theta = \tau'(y) f_m(\tau(y)) .$$

Now, (H)

$$\int g_m(y) dy = \int \tau'(y) f_m(\tau(y)) dy = \int \tau'(y) \frac{f_m(x)}{\tau'(y)} dx = \int f_m(x) dx .$$

Hence the fiducial density of Y is proper if and only if the fiducial density of X is proper. If then, the spectra of X and Y are infinite, then the MPD, MBI and MNS principles will be jointly satisfied for X and Y . We have demonstrated several cases in which a transformation of \mathcal{X} leaves the NBS invariant; specifically the log-normal, Weibull and Pareto models may be related by transformation to the normal and exponential models. Also for the examples studied in this paper if λ is a scale parameter for X and we transform λ to τ and X to Y so that τ is a location parameter for Y then the NBS is invariant under the dual transformation. It should be noted that transformations on \mathcal{X} and on (H) are quite different in effect. A transformation on \mathcal{X} generally involves a change in the model distribution while a transformation on (H) does not. This simply reflects that the model assumptions are dependent upon the scale of measurement.

10.4 Fiducial Inference

In addition to suggesting methods for specifying the prior Bayes density under indifference and thus proposing a model for logical probability statements, it has been suggested that the fiducial density of X be used in place of the unknown true density of X to predict future observations from the process under study. It has been pointed

out that this is simply an extension of the law of succession. Very little mathematical justification has been given for this suggestion. We have alluded to a consistency property, i.e. $f_n(x) \rightarrow m(x)$ as $n \rightarrow \infty$. We have referred to Lindley's result that $\int I(f_n) \geq I(f)$ $n = 1, 2, \dots$, which says that on the average succeeding estimates become more informative, having started of course by being minimally informative. Finally we have shown that in the simplest normal example that a direct and exact predictive frequency result obtains.

We have been unable to obtain any general predictive frequency theorems. It is clear, however, that if this aspect of the proposed dual interpretation of the Bayesian model is to be justified some further investigations must be undertaken in this respect. The information inequality as given above and the consistency property seem hardly sufficient to reasonably justify such an interpretation.

REFERENCES

- Bayes, T. (1763), "Essay towards solving a problem in the doctrine of chances", Phil. Trans. Roy. Soc., 53, 370-418.
- Birnbaum, A. (1962), "On the foundations of statistical inference", J. A. statist. Soc., 57, 269-306.
- Boole, G. (1854), An Investigation of the Laws of Thought, reprinted New York: Dover.
- Carnap, R. (1953), "Two concepts of probability", Feigl, H., Readings in the Philosophy of Science, New York: Appleton-Century-Crofts, 438-455.
- Cramer, H. (1946), Mathematical Methods of Statistics, Princeton Univ. Press.
- Doob, J. L. (1941), "Probability as measure", Ann. math. stat., 12, 206-214.
- de Finetti, B. (1937), "La prevision: ses lois logiques, ses sources subjectives", Ann. Inst. H. Poincare, 7, 1-68.
- Fisher, R. A. (1930), "Inverse probability", Proc. Camb. Phil. Soc., 26, 528-535.
- Fisher, R. A. (1933), "The concepts of inverse probability and fiducial probability referring to unknown parameters", Proc. Roy. Soc. A., 139, 343-348.
- Fisher, R. A. (1956), Statistical Methods and Scientific Inference. Edinburgh: Oliver and Boyd.
- Fraser, D. A. S. (1961), "On fiducial inference", Ann. math. stat., 32, 661-676.

- Fraser, D. A. S. (1962), "On the sufficiency and likelihood principles", Stanford Technical Report, 43.
- Good, I. J. (1950), Probability and the Weighing of Evidence, London: Griffin.
- Jaynes, E. T. (1957), "Information theory and statistical mechanics", Physical Review, 106, 620-630.
- Jeffreys, H. (1948), Theory of Probability, 2nd Edition, Oxford University Press.
- Jeffreys, H. (1961), Theory of Probability, 3rd Edition, Oxford University Press.
- Kerridge, D. F. (1961), "Inaccuracy and inference", J. R. statist. Soc. B., 23, 184-194.
- Keynes, J. M. (1921), Treatise On Probability, London: MacMillan.
- Kolmogorov, A. N. (1956), Foundations of Probability, 2nd English Edition, New York: Chelsea.
- Laplace, P. S. (1820), Theorie Analytique des Probabilités, 2nd Edition, Paris: Courcier.
- LeCam, L. (1958), "Les propriétés asymptotiques des solutions de Bayes", Publ. de l'Institut de Statistique de l'Université de Paris, 7, 17-35.
- Lindley, D. V. (1956), "On a measure of information provided by an experiment", Ann. math. stat., 27, 986-1005.
- Lindley, D. V. (1957), "Binomial sampling schemes and the concept of information", Biometrika, 44, 179-186.

- Lindley, D. V. (1958), "Fiducial distributions and Bayes theorem",
J. R. statist. Soc. B., 20, 102-108.
- Lindley, D. V. (1961), "The use of prior probability distributions
in statistical inference and decisions", Proceedings of the
Fourth Berkeley Symposium on Probability and Statistics,
1, 453-468.
- von Mises, R. (1957), Probability, Statistics and Truth, New York:
MacMillan.
- Novick, M. R. (1962), "A Bayesian indifference postulate",
Institute of Statistics Mimeograph Series No. 319,
University of North Carolina.
- Perks, W. (1947), "Some observations on inverse probability including
a new indifference rule", Journ. Inst. Act., B., Part II,
285-334.
- Raiffa, H. and Schlaifer, R. (1961), Applied Statistical Decision
Theory, Boston: Harvard Business School.
- Rajski, C. (1960), "On the existence of entropy", Information Theory,
Statistical Decision Functions, Random Processes, Transactions
of the Second Prague Conference, 541-542. New York: Academic
Press.
- Ramsey, F. P. (1931), The Foundations of Mathematics and Other Logical
Essays, New York: Harcourt Bruce.
- Reichenbach, H. (1932), "Axiomatik der Wahrscheinlichkeitsrechnung",
Math. Zeitschrift, 34, 568-619.

- Savage, L. J. (1954), The Foundations of Statistics, New York: Wiley.
- Savage, L. J. (1961), "The foundations of statistics reconsidered", Proceedings of the Fourth Berkeley Symposium on Probability and Statistics, 575-586. University of California Press.
- Savage, J. L. (1962), The Foundations of Statistical Inference, London: Methuen.
- Schlaifer, R. (1959), Probability and Statistics for Business Decisions, New York: McGraw-Hill.
- Shannon, C. E. and Weaver, W. (1949), The Mathematical Theory of Communication, University of Illinois Press.