5-1967

An Exposition on Bayesian Inference

John Laffoon
Utah State University

AN EXPOSITION ON BAYESIAN INFERENCE

by

John Laffoon

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Applied Statistics

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1967

ACKNOWLEDGMENTS

I would like to thank Dr. David White for proposing this subject, and making his library available for use.

Others who helped in this study included the late Mr. John Johnson whose constant questioning prevented me from flippantly glossing over many disconcerting problems, my wife Carolyn, who spent the last year of her life encouraging me on, Mrs. Roy Poole who labored on the typing from the first draft, and Miss Lila Armstrong who finished typing the final copy.

John Laffoon

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

An Exposition on Bayesian Inference

by

John Laffoon, Master of Science

Utah State University, 1967

Major Professor: Dr. David White
Department: Applied Statistics

The Bayesian approach to probability and statistics is described, a brief history of Bayesianism is related, differences between Bayesian and Frequentist schools of statistics are defined, protential applications are investigated, and a literature survey is presented in the form of a machine-sort card file.

Bayesian thought is increasing in favor among statisticians because of its ability to attack problems that are unassailable from the Frequentist approach. It should become more popular among practitioners because of the flexibility it allows experimenters and the ease with which prior knowledge can be combined with experimental data.

(82 pages)

INTRODUCTION

To the majority of the 21,000 professional statisticians in the U. S. statistics is a science, which, like any other science, succeeds only to the extent that it accurately describes the past and predicts the future. Through the years, there has been much debate as to how to describe and predict properly. Today, discussions between "Frequentists" and "Bayesians" over viewpoint and methodology, are reminiscent of the disputes in the 1930's between the followers of Sir Ronald A. Fisher and the adherents of Egon Pearson and Jerzy Neyman.

Inevitably the statistician working on a practical problem has to rely to a certain extent on human judgment, either his own or that of an outside expert. In sequential analysis, for example, someone has to estimate the loss entailed in making an erroneous decision. Once such a judgment has been made, classical frequentist statisticians try to apply their techniques with utter impartiality.

Bayesian statisticians scorn such impartiality and treat hunches and educated guesses as if they were part of the experimental information. Personal conviction is an integral part of probability to the Bayesians.

A simple example of the Bayesian approach can be shown in testing a coin. A Bayesian would first look at the coin and make an estimate of the probability that it would land heads up and then modify this prior judgment with statistical evidence gained by tossing the coin a number of times. Thus the Bayesian has a prior idea of the results. It is in the use of a mathematical process of combining prior judgment with experimental data developed by Thomas Bayes, an 18th century cleric, that Bayesians get their name.

For a more practical example, assume that a factory has two machines that produce screws. Experience has shown the proportion of defective screws that each machine is likely to turn out, p_1 and p_2. If a batch of screws from the factory happens to include, say five percent defectives, what is the probability that it came from the first machine? This is the sort of question to which Bayes's formula is applicable, and no one quarrels seriously with its use in such situations.

This example provides an opportunity to differentiate between the normal statement of frequentist and Bayesian statistical problems.

> Frequentist Problem: A machine is known, from long experience to produce a fraction P of imperfect products. What is the probability that a fraction p of the next n products will be imperfect. (Berkson, 1930, p 42)
>
> Bayesian Problem: The performance of a machine is unknown or known with uncertainty. n products are examined and fraction p found to be imperfect. What is the probability the machine generally turns out a fraction P imperfect products. (Berkson, 1930, p 42)

In the Frequentist problem P is treated as a fixed value; in the Bayesian problem, as a random variable.

The Bayesian statistician has a great deal of leeway in designing experiments and gathering data. If prior opinion is strong and the first few experimental results seem to support it, a Bayesian may call off further experimentation. In contrast, classicists insist that the whole experiment must be planned in advance and rigidly adhered to. If, for example, a geneticist is trying to estimate the proportion of fruit flies that have red eyes, he may observe 130 flies and find that twenty are red-eyed. The frequentist wants to know: Did he plan in advance to examine 130 flies? Did he intend to examine flies until he had seen twenty with red eyes? Or did he catch flies until he could find no more? The answer to these questions would determine the conclusion that classical statistics would deliver. But to the Bayesian, the original plan of the experimenter is not important and does not affect the conclusion. He can stop whenever he is satisfied with the strength of his convictions, and even before that he can quote odds on the outcome. This flexibility has attracted adherents and created foes (Boehm, 1964).

In addition, Bayesian approaches permit solution of problems which are difficult or unsolved from a frequentist approach, provide insight into the axiomatic basis of statistics, and clarify some of the interpretations that bother students. Because of these reasons, it seems propitious to develop a base for investigating Bayesianism. This thesis relates the history of Bayesian thought, investigates problems with Bayesian approaches, and gives a comprehensive bibliography of Bayesian writings.

SCHOOLS OF STATISTICAL THOUGHT

History of Bayesianism

Thomas Bayes

Any history of Bayesian statistics seems to begin with Thomas Bayes, an 18th century English cleric. In a paper written in 1763, but published after his death, Bayes developed an equation for combining prior knowledge with experimental data to make probability statements. Bayes, himself, was not a Bayesian, but his Theorem is central to the Bayesian approach and his Postulate became a center of dispute with the frequentists. His theorem is best described by reconstructing his example.

Using binomial data--success or failure-- Bayes assumes that prior to and independent of the experimental observations, the unknown probability of success, p, is a random variable of known distribution f(p). If the probability of p falling in the range dp is f(p)dp, then the probability of the event p combined with the outcome of the observation of a successes in n trials is:

$$\frac{n!}{a!(n-a)!} p^a (1-p)^{n-a} f(p)\, dp$$

This expression divided by its integral from 0 to 1 (all possible values of p) we call "posterior" probability of p. Direct integration of this posterior probability between any limits, p_1, p_2, allows the test of a compound hypothesis.

Bayes' example was an idealized billiard table. As Bayes stated:

> The square table ABCD be so made and levelled that if either of two balls O or W be thrown upon it, there shall be the same probability that it rests upon any one equal part of the plane as another, and that it must necessarily rest somewhere upon it.
>
> I suppose that the ball W be first thrown and through the point where it rests a line shall be drawn parallel to the ends of the table and that afterwards the ball O shall be thrown N times, and that its resting beyond the line should constitute a success. (Fisher, 1959, p 27)

The first ball, W, was thrown to select an observation from the prior distribution. If p is the proportion of the table beyond the line drawn it is inferred that

$$Pr(p<p_1) = p_1$$

for all values of p_1, and therefore that the probability that p should lie within the range dp is equal to dp, so the f(p) in the previous analysis may be equated to unity (Fisher, 1959).

Bayes' Theorem led to an apparently harmless probability equation used by Bayesians. If we let X be a variate from a population with a parameter θ, then

$$P(\theta \mid X) = P(X \mid \theta) \; P(\theta) / P(X)$$

will predict the value of θ once the sample X is taken. This is the product of a predetermined (prior) probability of θ, P (θ), and an estimate of X given θ, $P(X \mid \theta)$. P(X) is the normalizing term.

The mechanics of this equation are accepted by all statisticians, however, the way of obtaining the prior probability is a point of contention. Bayes stated a postulate known as "Bayes postulate", or the "indifference principle", that in the absence of better knowledge, any allowable value is equally probable, and thus a locally uniform distribution can be used.

In many cases, e.g. estimating the mean of a normal distribution where $-\infty \leq X \leq \infty$, a uniform prior is not defined by standard methods. In the early days of statistics, this did not seem to bother the practitioners. Fisher (1936) referred to the fact that Laplace was thoroughly Bayesian and used Bayes postulate in his Principle of Succession. Poisson used both Bayesian and frequentist methods. Even Karl Pearson did not fully discredit Bayes' inverse probability approach.

The Demise of Bayesian Thought

It is unfortunate that Bayes' postulate created disagreement in the statistical world since Bayes believed that prior distributions should be obtained from the real world -- auxiliary experiments or previous experience. He used the uniform distribution only when prior knowledge was absent or vague. Laplace (Fisher, 1959) believed prior probabilities were axiomatic and led later followers of Bayes to misuse his postulate by insisting on using uniform prior distributions. Uniform priors were so entrenched that when Carl Gauss attempted to use them and went awry, he discarded the entire use of priors rather than define a more reasonable prior density (Fisher, 1936). Other statisticians and mathematicians of the period including Venn, Chrystal, and Boole, following Gauss' lead, destroyed confidence in the Bayesian approach as well as the source of prior distributions. For years, the frequentist school reigned supreme.

Fiducial Inference

In the early 1930's however, discrimination--assigning an individual or object to a category--became a major problem for statisticians. A classical discrimination method was developed by Fisher that was, in essence, Bayesian.

Fiducial Inference says basically that, given a population with an unknown parameter, θ, it is possible to make an inference about the parameter by a sufficient statistic drawn from the population. The inference describes the range of values of the parameter which could produce the sample statistic with a given probability. That is, given a sample x_i with mean $\bar{x}$ and variance s^2 there is some range for the population mean, μ , which could produce the $\bar{x}$ with a probability $\alpha = P\left[(\bar{x}-\mu)/s\right]$

The fiducial argument sounds much like Bayesian inference. This similarity has led men such as Jeffreys (1957) to solve fiducial problems from a Bayesian point of view. There are, however, differences between the two methods and theoretically, situations in which one, but not the other method is possible.

Two instances were given by Fisher (1935, p 392-393)

1. In fiducial inference, the distribution of the parameter θ does not necessarily exist nor does it have a prior distribution. In Bayesian thinking, the form of the distribution must be assumed.

2. Sufficient estimators[1] or sets of estimators which together retain all of the information are required to apply the fiducial method. The fiducial method, according to Fisher, is always based on a sufficient statistic and more immediately on a pivotal quantity[1] to avoid different inferences based on the same data. Bayesian arguments can be applied without sufficient or ancillary statistics.

Despite his enumeration of differences between Fiducial and Bayesian approaches, Fisher himself stated that Fiducial probability "is entirely identical with the classical probability of the early writers, such as Bayes." (Fisher, 1956, p 186) This has been supported by Lindley (1958), Sprott (1960), and others. In addition, most statisticians refer to both methods as "Inverse Probability" without making distinction between them.

> If a Bayesian probability a priori is available we shall use the method of Bayes. If no prior probability of the form needed for Bayes' theorem is available we shall apply a fiducial argument (Fisher, 1959, p 25-26).

<u>Bayesian Revival</u>

An outstanding Bayesian of the late 1930's was Harold Jeffreys. Others, though not admitting it, let Bayesian thinking color their work.

---

[1] See Appendix for Fisher's definition of these terms.

Wald, who decries the Bayesian school, nevertheless used a Bayesian approach in his sequential sampling and optimum stopping rules by his willingness to let the data change his experimental procedure, his refusal to be bound by completely preplanned experiments, and his considering the unknown parameter as a random variable. He avoided uniform priors by using the results of one sample as the prior density for the next sample in sequence. Wald, thus removed the stigma of Bayes' Postulate and prepared the road for a Bayesian revival.

The full revival of Bayesian thought occurred with a publication by Leonard J. Savage (1954) which developed an axiomatic basis for statistics. As usual, however, Savage was not original in his development, but, according to Lindley (1965), derived his work from that of Ramsey (1931), deFinetti (1937) and the appendix to von Neumann (1947). Many axiomatic bases have since been developed. One passed on by Lindley (1965) is

Axiom 1 $0 \leq p(A|B) \leq 1$ and $p(A|A) = 1$

Axiom 2 If the events in A_n are exclusive given B then $p(\sum_n A_n|B) = \sum_n p(A_n|B)$

Axiom 3 $p(C|A,B)\,p(A|B) = p(A,C|B)$

These three axioms combined with a definition of independence are used to prove a basic set of theorems by Lindley (1965) including the equation commonly referred to as Bayes Theorem: "If p(B) does not vanish then $p(A|B) = p(B|A)\,p(A)\,/\,p(B)$" or more generally, "If A_n is a sequence of events and B is any other event with $p(B) \neq 0$ then $p(A_n|B) \propto p(B|A_n)\;p(A_n)$"

Practical applications waited until 1963. Two outstanding examples, Mosteller and Wallace's investigation of authorship of the disputed Federalist papers and Feldman's solution to the "Two Armed Bandit Problem" will be mentioned later. Workers such as Lindley, Box, Tiao, Zellner, and even such classicists as Wolfowitz and Fraser are using Bayesian reasoning to approach previously unsolved or computationally difficult problems.

Also, to the history of Bayesian thought belongs every statistician who has used a weighting factor which is a function of the parameter in question or has let the data influence the procedure of his experiment.

Philosophical Views of Probability

Prevailing Views

Probability has several interpretations--probability as relative frequency, probability as logical necessity, and probability as degree of belief. These are referred to as the Frequentist, Necessary, and Personal views of probability.

Frequentist view holds that some repetitive events, such as tosses of a coin, prove to be in reasonable agreement with the mathematical concept of independently repeated random events, all with the same probability. The magnitude of the probability can be obtained by observing repetitions of the events, and from no other source whatsoever. Thus, probabilities are independent of your knowledge.

The difficulty in the frequentist position is that probabilities can apply only to repetitive events. It is meaningless to talk about the probability that a given proposition is true (this probability can only be 1 or 0 according as the proposition is in fact true or false). One cannot then consider expected outcome or pursue a statistical analysis to maximize the expected outcome (Savage, 1954).

In the same vein a frequentist claims that you can make probability statements about events which have not occurred, but not about events which have occurred but about which you know nothing. For example, assuming a fair coin, for which the probability of heads is 1/2, any statistician will claim the probability of flipping a head is 1/2. Once the coin is flipped, but the result is unknown, some persons balk at saying the probability of it being a head is 1/2. The coin either landed heads or tails so no probability statement can be made (or is trivial when applied to coin flipping, but the same reasoning causes stumbling when it comes to making probability statements about propositions.

Carl Gauss made probability statements about features of the real world which can be ascertained only with some uncertainty (e. g., the distance from earth to sun). $Pr(x < x_0) = P$ could be computed for all values $0 \leq P \leq 1$. At any instant there is only one distance to the sun, so P is zero if $x \geq x_0$ and, is one if $x < x_0$, so probability statements about the real world are meaningless.

Necessary view holds that probability measures the extent to which one set of propositions, out of logical necessity and apart from human opinion, confirms the truth of another. That is, probability is a quantitative expression and extension of logical relationships.

The necessary view has a deficiency described by Wolfowitz in that it requires the determination of "logic" separate from human opinion. Human opinion is, however, inevitably involved in the logical correlation of propositions. Thus, this interpretation is really Personal probability and will not be considered separately.

Personal view holds that probability measures the degree of belief that an individual has in the truth of a particular proposition.

Three weaknesses exist in the Personal view--all related to the ability to specify a unique, quantitative prior probability in the absence of symmetry or long-run relative frequencies on which to base opinion. Lindley claims you can set a numerical value by demanding that any action based on the opinion be rational and consistent. This raises the second problem in which two reasonable individuals, faced with the same evidence, may have different degrees of confidence in the truth of the same proposition, i.e., different prior probabilities.

Wolfowitz believes that the personal view has a weakness in that the experimenter must have an opinion about every event. In effect, he must order all alternatives. Wolfowitz likens this to the situation where a young man getting married must not only select the girl, but also order all other female acquaintances in order of preference. This is a needless and meaningless restriction to put on the experimenter.

Position of Bayesians

Bayesians are followers of the Personal view. Most applied statisticians have a foot in both Personal and Frequentist camps. Wald, Mises, Wolfowitz and others are very comfortable accepting elements of both views and avoiding the conflicts of the opposing camps.

Modern Bayesians have answered some of the objections to their view of statistics and resolve some of the difference between themselves and frequentists.

Lindley (1965) gives an example of why, though different in nature, relative frequency and degree of belief can be expressed by the same terms.

Given two rods, A & B, their lengths (L), placed end to end are:

$$L(A + B) = L(A) + L(B)$$

Their masses (M) are:

$$M(A + B) = M(A) + M(B)$$

Hence length and mass have the same mathematical properties. This coincidence of mathematical properties happens with frequency limits and degrees of belief and they are both called probabilities. The elements A and B, usually represent events when dealing with frequencies and propositions or hypotheses when dealing with degrees of belief. When both interpretations of probability are possible the two values agree. For a fair coin our degree of belief that it will fall heads in a single toss is 1/2, as is the frequency probability.

To illustrate the Bayesian answer to the objection of two men setting different prior probabilities to the same event, carry the coin tossing example one step further and consider the coin flipped. One observer looks at the coin, a second observer does not. Is the probability of heads now different for each of the two observers? This points up a common oversight, not a deficiency of statistics. All probability statements are conditional--conditional on the knowledge of the observer. The true probabilities are not different, but the experimenters' evidence is.

Prior probabilities rely on past experience, and it is impossible for two persons to have identical backgrounds. For this reason it must be realized that every probability is conditional and it is dangerous to omit the conditioning event when specifying the probability. The Bayesian approach highlights (and solves) this problem by including the conditioning statement at all times.

The frequentist has difficulty in discrimination problems. He has two approaches. He generally estimates the parameter with a statistic which is not a function of θ. This requires a pivotal statistic. For example, in the well-known case where $p(x \mid \theta) \sim N(\mu, \sigma)$ he uses $(\bar{x}, s)$, the sample mean and standard deviation, and so gets away from $\theta = (\mu, \sigma)$.

In cases where no minimal sufficient statistic exists, it is difficult to see on what principle, other than that of expediency, the frequentist could choose his pivotal statistic. He is then reduced to introducing a weighting factor $p(\theta)$ associated with each θ and choosing a region such that $p(\theta)\ P(x \mid \theta)$ is valid, thus producing a "Bayes solution." By agreeing to place more emphasis on some values of θ than others he is moving towards a Bayesian outlook.

A Bayesian can also reduce the argument of the frequentist against making a probability statement about a proposition or past event by phrasing his proposition "What is the probability that I will guess the outcome correctly." Thus a Bayesian can take a frequentist approach in terms of right guesses.

Most Bayesians consider modern statistics to be perfectly sound in practice but done for the wrong reason. Lindley contends that intuition has saved the statistician from error, and the Bayesian view justifies what the statistician has been doing and develops new methods that the orthodox approach lacks.

On Prior Distributions

Importance of Priors

Information about θ comes from two sources, the data and the prior knowledge. It should be obvious that the choice of the prior knowledge affects the inference drawn, especially when data is scarce and the prior distribution carries great importance in the posterior probability. The more precise the data, the greater is the weight attached to it; the more precise the prior knowledge the greater is the weight attached to it.

Lindley (1965) shows quantitatively the influence of priors are reduced as the data increases

THM 1: Let x be $N(\theta, \sigma^2)$, where σ^2 is known, and the prior density of θ be $N(\mu_0, \sigma_0^2)$. Then the posterior density of θ is $N(\mu_1, \sigma_1^2)$ where

$$\mu_1 = \frac{x/\sigma^2 + \mu_0/\sigma_0^2}{1/\sigma^2 + 1/\sigma_0^2} \qquad \text{and} \qquad \sigma_1^{-2} = \sigma^{-2} + \sigma_0^{-2}$$

Posterior precision equals the data precision plus the prior precision. The posterior mean equals the mean of the data and prior mean, weighted with their precision.

The changes in knowledge take place according to Baye's theorem, which says that the posterior probability is proportional to the product of the likelihood (the probability density of the random variables forming the sample) and the prior probability.

Example of Prior Effect

An example of the effect of prior distributions can be seen in the following example by V. Mises (1942). Bacilli counts in a water supply were made. From this count an estimate was made of the probability of a user getting no bacillus in a standard sized sample. Three prior distributions were used.

(1) Equal Frequency
$P(x=0) = P(x=1) = P(x-2) = \ldots = P(x=5) = 1/6$
For this prior, the posterior probability
$P(x<1 \mid \text{prior}) = 0.73$

(2) Constant Density
$P(x) = x$ gives posterior $P(x<1 \mid \text{prior}) = 0.9975$

(3) Previous count

x = 0 in 3086 cases	x = 3 in 15 cases
x = 1 in 279 cases	x = 4 in 5 cases
x = 2 in 32 cases	x = 5 in 3 cases

The posterior probability in this case is 0.99915. (If the prior probability is 0, then, no matter what value x is observed, the posterior probability is also 0. This is an example of the general principle that if some event is regarded as virtually impossible, then no evidence whatsoever can lend it credibility.)

Robust Priors

A Bayesian approach seems necessary if one is to recognize the uncertainty in the assumptions which are built into many statistical procedures. Since the prior distribution is known only vaguely, Bayesians attempt to select a robust prior in preference to most other considerations. Frequentist oppose this because it may lead to wrong conclusions. They have been doing the same thing, however, in assuming specific parent distributions, then treating such assumptions as if they were axiomatic when, in fact, they are conjectures--subjective probability distributions.

Once having assumed the form of the parent distribution, frequentists derive appropriate criterion, and proceed with an "objective" analysis, by pretending to knowledge they do not have and ignoring what the sample has to tell about the distribution. For example, assuming normality for the comparison of two means they would derive the t-statistic then justify its use by showing that the distribution is robust under non-normality. However, this argument ignores the fact that if the parent distribution really differed from the normal, the appropriate criterion would no longer be the t-statistic.

We illustrate this using Darwin's paired data on the heights of self-and cross-fertilized plants quoted by Fisher (1935b). An appropriately scaled t-distribution centered about the sample mean gives a significance level for the hypothesis that $\theta = 0$ against the alternate $\theta > 0$ of 2.485 percent.

Now instead of assuming normality for the parent distribution we assume it to be uniform over some finite range the significance level is 2.388 percent rather than 2.485 percent. The test of the hypothesis that the true difference is zero using the t-criterion is thus very little affected by this major departure from normality.

If the parent distribution were assumed uniform when it were normal, we should not consider the "t" at all. We should use the function, $(m-\theta)/h$ where $m = (y_{max} + y_{min})/2$, $h = (y_{max} - y_{min})/2$ and y_{max} and y_{min} are respectively the largest and the smallest of the observations--jointly sufficient statistics for (θ, σ) on the uniform assumption. The significance is now 23.215 percent instead of 2.485 percent.

Thus, the conclusions we can draw assuming a uniform parent distribution are very different from those assuming normality, even though the t-criterion itself is very little affected.

Uncertainty in inferences we can make concerning the parameter can be resolved by explicity including the knowledge that we have about the parent distribution into our formulation. This knowledge (from the sample itself and from prior knowledge of the physical set up appropriate to the problem) is taken into account in Bayesian formulation.

Source & Nature of Priors

Bayes believed that sources for prior distributions were auxiliary experiments. In Bayes' billiard table problem, the auxiliary experiment is the throw of the first ball. The alternative outcomes for the first ball with their probabilities serve as the prior distribution for p. This particular experiment resulted in a uniform prior. Bayes stated a postulate known as the "indifference principle" that in the absence of prior knowledge a locally uniform prior distribution can be used. The uniform distribution sets all probabilities equal and is a precise way of saying we have no ground for choosing between the alternatives. Where something is known of the distribution, we should incorporate that knowledge. For example, in genetics $1/2^n$ and with dice 1/3, 1/6, 1/36, are common functions.

Karl Pearson apparently made the first serious effort to modify the prior distribution. He espoused the use of frequency arguments to provide the prior distribution. This approach was derided until recently when thinking began to swing toward "listening to what the data tells you"--even to the extent of using the sample data to modify the prior distribution.

Uniform Priors

The uniform distribution of the prior probability was first applied to the standard error by Gauss, who found it unsatisfactory. This section shows the problem of finding a valid prior distribution for the standard error. According to Jeffrey's uniform estimates of prior distributions are not invariant over a semi-infinite range.[2] The invariance argument led Jeffreys to propose $(\log\sigma)$ for a prior over a semi-infinite range. This function, however, is not a density function over the entire range. Traditionally this problem was avoided by redefining the conventions of probability or defining the distribution over an extremely large finite range. This latter approach satisfies practical situations in which the likelihood function is significant only in a finite range.

---

[2] See Appendix for explanatory example of invariance.

Invariant Estimate. We can use the precision, h, in place of the standard deviation, σ, in our problem. Using ths uniform prior distribution for h, however, does not provide the same answer as using the uniform prior of σ. (i.e., $d\sigma \not\propto dh$). This lack of invariance is troublesome because of its effect on the user.

However, since $h\sigma$=(constant) then $\frac{dh}{h} + \frac{d\sigma}{\sigma} = 0$. The prior of σ, $0<\sigma<\infty$ is taken proportional to $d\sigma/\sigma$ i.e., $P(\sigma \mid H) = \frac{d\sigma}{\sigma}$ then $0<h<\infty$ and its prior is proportional to dh/h. These expressions are consistent.

The same invariance argument holds for the power of σ. If $P(\sigma \mid H) \propto d\sigma$ then $P(\sigma^n \mid H) \propto d\sigma^n \propto \sigma_1^{n-1} d\sigma$ but $d\sigma \not\propto \sigma_1^{n-1} d\sigma$. However, $d\sigma/\sigma \propto d\sigma^n/\sigma^n$. We now have a distribution which ranges over the positive real line.

Density Estimate. One difficulty with this distribution is that if we take $P(\sigma \mid H) \propto d\sigma/\sigma$ as a statement that σ may have any value between 0 and ∞ and compare probabilities for finite ranges of σ, we must use ∞ instead of 1 to denote certainty. However, with certainty being ∞, the probability that σ is less than any finite number, $P(\sigma \leq K)$, is zero. This is inconsistent with our assumption of no knowledge of σ. $\frac{1}{\infty}\left[\int_0^K d\sigma + \int_K^\infty d\sigma\right] = \frac{K}{\infty} + \frac{\infty - K}{\infty}$ $\quad \therefore P(\sigma \leq K) = 0$

The use of ∞ as certainty will give us more trouble in integrating the posterior distribution, so a second, more satisfying, but esthetically messier solution is used.

The prior distribution of σ is defined as the limit of $\int_a^b P(\sigma \mid H)$ as $a \to 0$, $b \to \infty$. The function does not exist at the limit, but approaches the limit as close as desired. That is, we can select any positive ϵ and find values of a and b such that $\int_{a-\Delta a}^{b+\Delta b} d\sigma/\sigma - \int_{a}^{b} d\sigma/\sigma < \epsilon$ for any $\Delta a < a$ and finite Δb.

The method creates little problem since an intermediate range contributes most of the values and the limits make a negligible contribution. Also, we are ultimately concerned with the posterior distribution which is the product of the prior distribution and the likelihood function. If the nature of the prior distribution is such that the standard deviation is zero, we have defined the location and there is no sense of performing the experiement. If the standard deviation is infinite, we are in a range where the likelihood is insignificant.

Actual experimental situations often permit us to assume that the prior distribution of the location parameter, θ, is locally uniform if we can say that the prior is fairly constant in the region in which the likelihood is appreciable and at no other point is it of sufficient magnitude to become appreciable when multipled by the likelihood. Most actual experiments will be conducted only when it is expected that the likelihood will exert a much stronger influence in the final result than will the prior distribution; otherwise, there is little point in doing the experiment.

General Priors

A more satisfying approach is to assume our distribution is a well-behaved distribution, defined in a manner to approach uniformity. For example, a normal distribution, itself with a very large standard deviation, approximates a uniform distribution over a reasonable range. A more general approach is to assume the prior to be a member of a class of symmetric power distributions which include the normal, together with other more leptokurtic and more platykurtic distributions.

$$p(y \mid \theta, \sigma; \beta) = \omega \exp\left\{-\frac{1}{2}\left|\frac{y-\theta}{\sigma}\right|^{2/(1+\beta)}\right\}$$

$$\omega^{-1} = \Gamma\left[1+\tfrac{1}{2}(1+\beta)\right] 2^{[1+\frac{1}{2}(1+\beta)]}\sigma$$

$$-\infty < y < \infty \qquad 0 < \sigma < \infty \qquad -\infty < \theta < \infty \qquad -1 < \beta < 1$$

In particular, we see that when $\beta = 0$, we have the normal distribution, when $\beta = 1$, the double exponential, and as β tends to -1, our distribution tends to the uniform distribution.

Box and Tiao (1962) develop this form of priors for normal, uniform and double exponential distributions and show how information concerning β coming from the sample is included in the formulation and eliminates the influence of unlikely parent distributions.

APPLICATIONS

Coin Flipping

The application that revived Bayes' Theorem was discrimination--assigning an object or individual to a category. To illustrate the nature of the problem and the Bayesian approach we will resort to a noncontroversial coin tossing example using well defined priors.

You have two coins x_1, x_2

x_1 = fair coin

x_2 = two headed coin

You draw a coin, flip it once and get a head H.

Which coin was drawn?

Using standard notation:

$$p(x_1) = 1/2 \qquad p(x_2) = 1/2$$

$$p(H \mid x_1) = 1/2 \qquad p(H \mid x_2) = 1$$

Bayes' Theorem says

$$P(x_i \mid H) = P(H \mid x_i)\, P(x_i)/P(H) \propto P(H \mid x_i) \quad P(x_i)$$

Thus the odds are two to one in favor of having drawn x_2, rather than x_1

$$\phi\,(x_2 \mid H) = \frac{P(x_2 \mid H)}{P(x_1 \mid H)} = 2 > 1$$

The coin is thus assigned to the category x_2.

Authorship Assignment

Assignment problems apply to such varied disciplines as archeology, medicine, authorship, education, etc.

A Bayesian approach was used to determine the authorship of twelve of the eighty-five Federalist papers-which of them were written by James Madison and which by Alexander Hamilton? (Mosteller, 1964) The frequency of key words in the disputed papers was compared with the frequency of the same words in papers known to have been written by each of the two authors. The analysis was first made by a Bayesian method with the prior assumption that Madison and Hamilton were equally likely to have authored each of the disputed papers. Later the work was redone with classical techniques that ruled out any prior assumption. The Bayesian conclusion was that all twelve papers almost surely came from Madison with the most questionable paper at odds of eight to one on Madison's authorship. The next weakest had odds of 800 to one. Classical techniques indicated strongly that ten of the papers were written by Madison, but favored Hamilton slightly on one and Madison slightly on the other.

The basic approach was to work with odds so that

$$\text{final odds} = (\text{initial odds}) \times (\text{likelihood ratio})$$

Define:

P_i = probability before observations that Hypothesis i is true,

$p(x|H_i)$ = probability of observing x given H_i is true.

Since, by Bayes' Theorem

$$p(H_1|x) = P_1 \, p(x|H_1) / \left[P_1 \, p(x_1|H_1) + P_2 \, p(x_1|H_2) \right]$$

we have

$$\text{Odds } (H_1 x) = \frac{p(H_1|x)}{p(H_2|x)} = \frac{P_1 \, p(x|H_1)}{P_2 \, p(x|H_2)}$$

For a Poisson Distribution, the likelihood ratio

$$\frac{P(x|H_1)}{P(x|H_2)} = \left(\frac{\mu_1}{\mu_2}\right)^x \exp\left\{-(\mu_1 - \mu_2)\right\}$$

The log likelihood ratio is thus

$$\lambda(x) = x \log \frac{\mu_1}{\mu_2} - (\mu_1 - \mu_2)$$

The two parameters used to run the tests were

$$K = \mu_1 + \mu_2$$

$$T = \mu_1 / (\mu_1 + \mu_2)$$

Where $K/2$ measures the average frequency of the word and T measures the ability to discriminate.

The difficulty, like that of all Bayesian studies was setting initial odds P_1 and P_2. They assumed an equal chance for each author, but could have argued that Hamilton wrote 43 of the undisputed papers to Madisons 14, so the prior odds could have been $43/14 \cong 3$.

To illustrate the effect of priors, the log odds increased on the order of 50 - 150 percent when a negative binomial prior replaced the Poisson distribution.

The over-all similarity of the results confirmed the belief of many statsticians that when sufficient data is available, either technique will lead to a reasonable conclusion (Mosteller, 1964).

Selection

Another problem was conquered by Bayesian methods by Dorian Feldman at the University of California at Berkely. A gambler is faced with an imaginary slot machine equipped with two handles. He knows that one of the handles pays off 80 percent of the time and the other pays off 60 percent of the time, but he does not know which is the 80 percent handle. What strategy should he adopt to make the most money?

He could perform a "classical" experiment by pulling each of the handles a given number of times to determine which paid off more more frequently. But Feldman proved that the gambler could do better with a Bayesian procedure. His solution; start by pulling one of the handles, selected at random, and if it pays off, pull it again. When it stops paying off, switch to the other handle. As the game progresses for any given play he pulls the handle with the greatest chance of paying off (Boehm, 1964).

David Blackwell, who suggested the problem to Feldman thinks a similar strategy might be adopted to such problems as physicians who have a choice of two new drugs to administer to patients with a given disease.

Behrens Via Bayes

The Behrens-Fisher problem is to test the difference in the means of two normal populations.

Suppose we sample from two normal populations with unknown and possibly unequal means, μ_1, μ_2 and variances, σ_1^2, σ_2^2. Let (x_1, s_1) be the sufficient statistic for (μ_1, σ_1) based on a sample size n_1 and (x_2, s_2) for (μ_2, σ_2) based on a sample size n_2. Now we can define a statistic developed by Fisher (1934) on earlier work by Behrens (1929)

$$\epsilon = \left[(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)\right] \Big/ \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

$\left[(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)\right]$ is distributed normally about zero with a variance $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$. By defining ψ by the expression $\tan\psi = \frac{s_2}{\sqrt{n_2}} \Big/ \frac{s_1}{\sqrt{n_1}}$ we have ϵ as $\frac{(\bar{x}_1 - \mu_1)}{s_1/\sqrt{n_1}} \cos\psi - \frac{(\bar{x}_2 - \mu_2)}{s_2/\sqrt{n_2}} \sin\psi$

But $\frac{\bar{x} - \mu}{s/\sqrt{n}}$ is distributed as a t. Therefore

$$\epsilon = t_1 \cos\psi - t_2 \sin\psi$$

where the joint distribution of t_1, t_2, is given by:

$$d(t_1, t_2) \propto \left\{ dt_1 \Big/ \left(1 + \frac{t_1^2}{n_1 - 1}\right)^{n_1/2} \right\} \left\{ dt_2 \Big/ \left(1 + \frac{t_2^2}{n_2 - 1}\right)^{n_2/2} \right\}$$

Bayesian Derivation

To derive this same joint distribution from a Bayesian approach consider a likelihood function of drawing samples

$$p(x_i \mid \mu, \sigma) = \frac{1}{\sigma^n} \exp\left\{ -\Sigma (x_i - \mu)^2 / 2\sigma^2 \right\} \qquad \text{(Kendall, 1959).}$$

Jeffreys introduces a modification of Bayes' postulate when a parameter, θ, can extend to infinity in only one direction, $dF \propto d\theta/\theta$

Bayes provided us information about the form of the prior distribution when the parameter is symmetrical about zero and can extend to infinity in both directions, $dF \propto d\theta$. Applying these to our problem,

$$P(d\mu, d\sigma | H) = d\mu \; d\sigma / \sigma$$

Note: H is what Jeffreys (1957) calls general datum that is included in all experience, e.g., H might be rules of pure mathematics. Statisticians normally omit H but Lindley (1965) warns against this omission.

Now by Bayes Theorem, the joint posterior distribution

$$P(d\mu_1, d\mu_2, d\sigma_1, d\sigma_2 | x_1, x_2, H) = \frac{\exp\left\{-\frac{\Sigma(x_{i_1}-\mu)^2}{2\sigma_1^2} - \frac{\Sigma(x_{i_2}-\mu)^2}{2\sigma_2^2}\right\}}{\sigma_1^{n_1+1}\,\sigma_2^{n_2+1}} \, d\mu_1 \, d\mu_2 \, d\sigma_1 \, d\sigma_2$$

Integrating out σ_1, σ_2 we can get a joint distribution for t_1, t_2.

Consider either term of the form

$$\frac{1}{\sigma^{n+1}} \exp\{-\Sigma(x_i-\mu)^2/2\sigma^2\} \, d\mu \, d\sigma$$

Now:

$$\Sigma(x_i-\mu)^2 = \Sigma(x_i-\bar{x}+\bar{x}-\mu)^2 = \Sigma(x_i-\bar{x})^2 + 2\Sigma(x_i-\bar{x})(\bar{x}-\mu) + \Sigma(\bar{x}-\mu)^2$$

$$= (n-1)s^2 + n(\bar{x}-\mu)^2$$

Integrating over σ

$$d\mu \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{1}{2\sigma^2}\left[n(\bar{x}-\mu)^2 + (n-1)s^2\right]\right\} d\sigma$$

Since $t^2 = (\bar{x}-\mu)^2 n / s^2$ we have

$$d\mu \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{1}{2\sigma^2}\left[s^2 t^2 + (n-1)s^2\right]\right\} d\sigma$$

Let $u = \frac{n-1}{2} s^2 \left(1 + \frac{t^2}{n-1}\right)$ and $x = 1/\sigma^2$

so $-\frac{d\mu}{2}\int_\infty^0 x^{\frac{n}{2}-1} e^{-ux} dx = \frac{d\mu}{2} \frac{\Gamma(n/2)}{2u^{n/2}} = -\frac{s\,\Gamma(n/2)\,dt\,2^{n/2}}{2\sqrt{n}\,[(n-1)s]^{n/2}\left(1+\frac{t^2}{n-1}\right)^{n/2}}$

When integration is carried out for both, σ_1 and σ_2 we get

$$P(dt_1, dt_2 | x_1, x_2, H) \propto \frac{dt_1}{\left(1+\frac{t_1}{n_1-1}\right)^{n_1/2}} \frac{dt_2}{\left(1+\frac{t_2}{n_2-1}\right)^{n_2/2}}$$

Thus Jeffreys uses a Bayesian derivation to show that the joint fiducial distribution of the means of two populations is distributed as a joint distribution of two t's. This is identical to the distribution used in Behrens integral and is the basis for tables of Behrens' - Fisher distribution computed by Sukhatme (1938)

Fiducial t

It is not always obvious why, in a Fiducial test $(\bar{x} - \mu)/s'$ has a t-distribution, since $\bar{x}$ and s' are fixed samples and μ is an unknown parameter.

The definition of a t-variate, however, does not discern between a conventional frequency test for $\bar{x}$ or a fiducial test on μ. Two properties of t, according to Anderson and Bancroft (1952) are:

1. t is the ratio between a normal deviate and the square root of an ubiased estimate of its variance.

$$t = (\bar{x}-\mu)\Big/\frac{s}{\sqrt{n}}$$

2. t^2 is the ratio between the square of a variate $(\bar{x} - \mu)\Big/\frac{s}{\sqrt{n}}$ which is N (0, 1) and a variate s^2/σ^2 which is distributed as χ^2_ν with ν degrees of freedom.

The second property is taken as the definition of t.

A random sample x_i, i =1, . . . , n, is taken from a population with mean, μ, and variance, σ^2. $\bar{x}$ and s^2 are jointly sufficient statistics obtained from the sample for μ and σ^2. Define

$s' = s/\sqrt{n}$ Now: $$t^2 = \frac{(\bar{x}-\mu)^2}{\sigma^2/n}\Big/\frac{s^2}{\sigma^2} = \frac{(\bar{x}-\mu)^2}{(s/\sqrt{n})^2} = \left(\frac{\bar{x}-\mu}{s'}\right)^2$$

So the distribution satisfies the definition of t.

Bayesian Regression

A regression problem was approached from a Bayesian view based on work by Tiao and Zellner (1964). The first part of this section is a short development of equations; the second part shows the numerical results for an industrial problem.

Development

The usual regression model with coefficient vector $b=(b_1, b_2, \ldots, b_p)$ can be written $Y=Xb+e$ where Y is a Txl vector of observations, X is a Txp matrix of fixed elements with rank p, and e is a Txl vector of random errors. We assume that the elements of e are normally and independently distributed, each with mean zero and unknown variance σ^2. Under these assumptions the likelihood function is

$$L(b,\sigma|y) = (\sigma\sqrt{2\pi})^{-T} \exp\{-(Y-Xb)'(Y-Xb)/2\sigma^2\}$$

Using Bayes' Theorem, the likelihood function is combined with a prior distribution $p(b,\sigma)$ to obtain a joint posterior distribution $p(b,\sigma|y)$ for the parameters b and σ.

$$p(b,\sigma|y) = p(b,\sigma)\,L(b,\sigma|y) \Big/ \int_R p(b,\sigma)\,L(b,\sigma|y)\,db\,d\sigma$$

In situations where little is known about b and σ, Jeffreys (1961) and Savage (1962) suggested that the prior distributions of b and $\log\sigma$ should be taken as locally independent and uniform. That is

$$p(b)\propto K_1 \qquad p(\log\sigma)\propto K_2 \quad \text{or} \quad p(\sigma)\propto 1/\sigma$$

Thus $p(b,\sigma|y) = (2\pi)^{-T/2}\sigma^{-(T+1)}\exp\left\{-\frac{1}{2\sigma^2}\left[(T-p)s^2+[b-(X'X)^{-1}(X'Y)]'(X'X)[b-(X'X)^{-1}(X'Y)]\right]\right\}$

The marginal posterior distribution of b is obtained by integrating the joint posterior density function over σ,

$$p(b,y) \propto \left\{1 + [b-(X'X)^{-1}(X'Y)]'(X'X)[b-(X'X)^{-1}(X'Y)] / (T-p)s^2\right\}^{-T/2}$$

This is a multivariate t-distribution first derived by Savage (1961).

If information is available about the parameters, it should affect the priors. For example, if the prior information is obtained from previous (or concurrent) experiments one can use this prior information to obtain the posterior distribution $p(b,\sigma|y_1)$ which in turn serves as the prior to obtain the posterior $p(b,\sigma|y_2)$.

As shown by Tiao and Zellner (1964) the final marginal distribution of b depends on the relationship between σ_1 and σ_2. We are interested in the situation where σ_1 and σ_2 are independent and unknown, but will lead up to that problem by considering two more restrictive situations.

Equal Variances. Raiffa and Schlaifer (1961) considered the case where $\sigma_1 = K\sigma_2$ with K known. Without loss of generality, let us assume for this example that $K = 1$ so that $\sigma_1 = \sigma_2 = \sigma$ (This condition is often encountered in controlled biological experiments.) For this case

$$p(b,\sigma|y_1,y_2) \propto \sigma^{-(T_1+T_2+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[(T_1-p)s_1^2+(T_2-p)s_2^2+[b-(X'X)^{-1}(X'Y)]'(X'X)[b-(X'X)^{-1}(X'Y)]\right]\right\}$$

where $X'X = X_1'X_1 + X_2'X_2$ and $X'Y = X_1'Y_1 + X_2'Y_2$

and the marginal distribution of b is

$$p(b|y_1,y_2) \propto \left\{1 + \frac{[b-(X'X)^{-1}(X'Y)]'(X'X)[b-(X'X)^{-1}(X'Y)]}{(T_1-p)s_1^2+(T_2-p)s_2^2}\right\}^{-\frac{T_1+T_2}{2}}$$

which is the same form as that given previously.

One Variance Known. Next assume σ_1 known but b and σ_2 unknown. Again assuming locally uniform priors, the posterior distribution of b is given by

$$p(b|y_1,y_2) \propto \exp\left\{\frac{-1}{2\sigma_1^2}\left[b-(X_1'X_1)^{-1}(X_1'Y_1)\right]'(X_1'X_1)\left[b-(X_1'X_1)^{-1}(X_1'Y_1)\right]\right\}$$
$$\left\{1+\frac{\left[b-(X_2'X_2)^{-1}(X_2'Y_2)\right]'(X_2'X_2)\left[b-(X_2'X_2)^{-1}(X_2'Y_2)\right]}{(T_2-p)\,S_2^2}\right\}^{-T_2/2}$$

Since σ_1, is known and no longer a random variable, it can be used directly in the formulation and not approximated by the statistic s_1. The marginal distribution of b is then obtained by directly evaluating the posterior with the fixed value, σ_1, but integrating over all values of the still random variable σ_2.

Theil (1963) considered this case within classical sampling theory rather than a Bayesian approach and obtained an estimator for b which incorporates information from both samples.

$$B = \left[\frac{1}{\sigma_1^2}(X_1'X_1) + \frac{1}{S_2^2}(X_2'X_2)\right]^{-1}\left[\frac{1}{\sigma_1^2}(X_1'Y_1) + \frac{1}{S_2^2}(X_2'Y_2)\right]$$

B is the limiting mean of the Bayesian distribution as $(T_2 - p) \rightarrow \infty$

Independent Unknown Variances. Now assume that σ_1 and σ_2 are independent and unknown parameters. This condition is valid when data are collected under different conditions and there is no basis for assuming any prior relationship between the unknown parameters σ_1 and σ_2. Again assuming locally uniform priors for b, log σ_1, and log σ_2, the posterior distribution of b based on two samples is

$$p(b|y_1,y_2) \propto \left\{1+\frac{\left[b-(X_1'X_1)^{-1}(X_1'Y_1)\right]'(X_1'X_1)\left[b-(X_1'X_1)^{-1}(X_1'Y_1)\right]}{(T_1-p)\,S_1^2}\right\}^{-T_1/2}$$
$$\left\{1+\frac{\left[b-(X_2'X_2)^{-1}(X_2'Y_2)\right]'(X_2'X_2)\left[b-(X_2'X_2)^{-1}(X_2'Y_2)\right]}{(T_2-p)\,S_2^2}\right\}^{-T_2/2}$$

The expected value of this form can be approximated by an asymptotic solution from Tiao & Zellner (1964) of the same form derived by Theil.

$$B = \left[\frac{1}{S_1^2}(X_1'X_1) + \frac{1}{S_2^2}(X_2'X_2)\right]^{-1}\left[\frac{1}{S_1^2}(X_1'Y_1) + \frac{1}{S_2^2}(X_2'Y_2)\right]$$

where, of course, σ_1 and σ_2 are both represented by the statistics S_1 and S_2.

Numerical Results

The example used to illustrate this technique is taken from an actual industrial problem. The exact nature of the data is proprietary but it represents an attempt to properly define a log-linear relationship between weight and cost of a solid rocket motor component.

$$\log \text{cost} = \alpha + \beta \log \text{weight}$$

Data. We have two sets of data, but do not have complete confidence in the validity of either set. The first set resulted from a theoretical study. Costs were generated by three companies, knowing that no business would result. Such cost studies are consistent within themselves but are generally poor estimates. Though each company priced the same six designs, costs varied ridiculously between companies. It was not possible to eliminate the company effect or to select one "best" cost for each design, since a company may be giving intentionally low costs to particular designs to influence cost-effectiveness studies. When actual components are built, the higher costs may result. Therefore all data from this set are equally-weighted. The slope of the curve developed on the basis of this set is probably close to "correct," but the magnitude of the costs may be wrong.

The second set of data is taken from actual and proposed costs. The magnitude of the curve developed from these data may be good, but the data represent dissimilar situations--some are R&D components, some demonstration equipment, and others are production systems. Some points represent total component costs, others are only partial costs in which the vendor took a loss to get the business and be in a better technical position to get future business. Therefore, this set of data is not consistent.

The two sets of data are given in Table 1.

Data Reduction. There are several ways to reduce the data to obtain an estimating relationship.

1. Use set 1.
2. Use set 2.
3. Combine both sets of data into one.
4. Use set 1 to get β and plug it into set 2 to get α.
5. Adopt the Bayesian approach assuming locally uniform priors for both sets.
6. Use a Bayesian sequential approach with the posterior of the original data serving as the prior value for the actual data.

Table 1. Regression Data

Set 1		Set 2	
Wt. (lb.) x_1	Cost ($) y_1	Wt. (lb.) x_2	Cost ($) y_2
17.4	9,000	1,707	57,500
85.6	26,000	1,707	47,000
102.7	32,000	301	21,000
189.0	29,000	360	22,600
312.2	27,000	784	21,550
1,441.4	69,000	784	32,800
17.4	1,250	9,596	250,000
85.6	3,200	1,640	46,000
102.7	6,400	1,640	63,200
189.0	7,500	204	15,400
312.2	31,000	204	16,400
1,441.4	82,500	5,134	112,850
17.4	9,364	295	17,200
85.6	11,612	9,129	246,068
102.7	12,460	8,581	184,416
189.0	15,517	177	14,200
312.2	17,923		
1,441.4	38,195		

The drawback of the first two methods is that neither set, by itself, inspires confidence in the users. Method 3 is the one presently used by industry in developing cost estimating functions. The two sets, however, do not form a consistent set and the resulting confidence interval is too large.

The fourth method is satisfactory in practice, but there is no theoretical basis for this procedure. The resulting equation does not yield itself to further statistical analysis. Methods 5 and 6 are developed in the following portion of this section. The coefficients of all six methods are given in Table 3.

The regression equation of interest is of the form

$$y = \alpha + \beta x + \epsilon$$

or, in matrix notation

$$Y = bX + e$$

with the following computational forms:

$$b = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \qquad (X'X) = \begin{pmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{pmatrix} \qquad (X'Y) = \begin{pmatrix} \Sigma Y \\ \Sigma XY \end{pmatrix}$$

The data reduce to the computational factors,

Table2. Computational factors

T_1	18	T_2	16
P	2	P	2
Σx_1	39.3428	Σx_2	48.6412
Σx_1^2	92.1067	Σx_2^2	153.6016
Σy_1	75.4562	Σy_2	74.3184
Σy_1^2	168.4677	Σy_2^2	229.8732
$\hat{\alpha}_1$	2.92585	$\hat{\alpha}_2$	2.55418
$\hat{\beta}_1$	0.57929	$\hat{\beta}_2$	0.68772
s_1^2	0.09162	s_2^2	0.0079517

For method 5 use the Tiao and Zellner (1964) formulation.

Set 1

$$(X_1'X_1) = \begin{pmatrix} 18. & 39.3428 \\ 39.3428 & 92.1067 \end{pmatrix} \qquad (X_1'Y_1) = \begin{pmatrix} 75.4562 \\ 168.4677 \end{pmatrix}$$

$$(X_1'X_1)^{-1} = \begin{pmatrix} 0.8368 & -0.3575 \\ -0.3575 & 0.1635 \end{pmatrix} \qquad (X_1'X_1)^{-1}(X_1'Y_1) = \begin{pmatrix} 2.9144 \\ 0.5689 \end{pmatrix}$$

$$(T_1 - p)\, s_1^2 = 1.46592$$

Set 2

$$(X_2'X_2) = \begin{pmatrix} 16 & 48.6412 \\ 48.6412 & 153.6020 \end{pmatrix} \qquad (X_2'Y_2) = \begin{pmatrix} 74.3184 \\ 229.8732 \end{pmatrix}$$

$$(X_2'X_2)^{-1} = \begin{pmatrix} 1.6757 & -0.5306 \\ -0.5306 & 0.1745 \end{pmatrix} \qquad (X_2'X_2)^{-1}(X_2'Y_2) = \begin{pmatrix} 2.5647 \\ 0.6795 \end{pmatrix}$$

$$(T_2 - p)\, s_2^2 = 0.111324$$

Marginal Posterior

$$p(\alpha, \beta \mid y_1, y_2) = \left[1 + \begin{pmatrix} \alpha - 2.9144 \\ \beta - 0.5689 \end{pmatrix}' \begin{pmatrix} 18 & 39.3428 \\ 39.3428 & 92.1067 \end{pmatrix} \begin{pmatrix} \alpha - 2.9144 \\ \beta - 0.5689 \end{pmatrix} \Big/ 1.46592 \right]^{-9}$$

$$\left[1 + \begin{pmatrix} \alpha - 2.5647 \\ \beta - 0.6795 \end{pmatrix}' \begin{pmatrix} 16 & 48.6412 \\ 48.6412 & 153.6020 \end{pmatrix} \begin{pmatrix} \alpha - 2.5647 \\ \beta - 0.6795 \end{pmatrix} \Big/ 0.111324 \right]^{-8}$$

$$= (12.1001\alpha^2 - 100.6202\alpha + 52.8943\,\alpha\beta - 224.6036\beta + 61.9163\beta^2 + 211.5123)^{-9}$$

$$(13.8937\alpha^2 - 128.6678\alpha + 84.4759\,\alpha\beta - 397.9205\beta + 133.3814\beta^2 + 301.1867)^{-8}$$

The most probable values for the Bayesian estimates of α and β were obtained by maximizing the joint posterior distribution $p(\alpha, \beta \mid y_1, y_2)$ with respect to α and β jointly by use of an IBM 360 program. Brinton and Garner (1966).

The expected value of method 5 is approximated by

$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \left[\frac{1}{.09162}\begin{pmatrix}18 & 39.3428\\ 39.3428 & 92.1067\end{pmatrix} + \frac{1}{.0079517}\begin{pmatrix}16 & 48.6412\\ 48.6412 & 153.6020\end{pmatrix}\right]^{-1}$$

$$\left[\frac{1}{.09162}\begin{pmatrix}75.4562\\ 168.4680\end{pmatrix} + \frac{1}{.0079517}\begin{pmatrix}74.3184\\ 229.8730\end{pmatrix}\right]$$

$$= \begin{pmatrix}2.85203\\ 0.60192\end{pmatrix}$$

Method 6 involves assuming $S_1 = \sigma_1$, which uses the posterior value of the original data as the prior s for the actual data. Instead of computing that value directly, the limit was computed as defined by Theil (1963) and is identical to the expected value approximation of method 5 since we assume $\sigma_1 = S_1$.

Table 3 summarizes the coefficients from the six methods.

Table 3. Regression coefficients

Method	α	β	
1	2.92585	0.579292	
2	2.55418	0.687720	
3	2.86058	0.596870	
4	2.88380	0.579292	
5	2.83358	0.597302	Most probable value
	2.85203	0.60192	Expected value approx.
6	2.85203	0.60192	

BIBLIOGRAPHY

Literature Survey

Literature Reviewed

A comprehensive literature survey was made of Bayesian articles. The following journals, symposia and book authors proved most fruitful for articles pertaining to the theory or application of Bayesian statistics:

Journals:

- Biometrika
- Annals of Mathematical Statistics
- Journal of American Statistical Association
- Journal of the Royal Statistical Society
- Technometrics
- Operations Research
- International Abstracts in Operations Research

Symposia:

- Berkely Symposia on Mathematical Statistics & Probability
- Operations Research Society of America National Meetings
- Regional Meetings of the Institute of Mathematical Statistics

Books by:

- Fisher
- Jeffreys
- Savage
- Lindley

Other sources were reviewed when referenced but no systematic coverage was made of them. Some sources were not covered because of lack of availability. The most glaring deficiency is Econometrica, followed by the Management Science Journal and works in Industrial Engineering. Most of the important work has been covered in the documents reviewed. Very little work was done previous to the last eight years that has not been improved upon or at least reviewed in later publications.

Some listed sources were not actually read. This is especially true of government reports, papers given to the Institute of Mathematical Statistics, and foreign language articles. In these cases, the published abstracts were used to classify the subject matter.

Reference File

To facilitate use of the reference material, both machine-sort and key-sort card files of this material are provided. Classifications are set up as to subject matter, application, and statistical distribution involved. A second classification is provided as to source, availability, and author. These classifications are shown in Table 4.

The following listing of the articles in the card file is alphabetized by author. No attempt is made in this list to indicate the subject matter or cross-reference the material.

AD reports can be purchased from the Clearinghouse, Springfield, Virginia.

Literature Listing

Aggarwal, O.P. 1959. Bayes and minimax procedures in sampling from finite and infinite populations. Annals of Mathematical Statistics. 30:206-218.

Aggarwal, O.P. 1965. Bayes and minimax procedures for estimating mean of a population with two-stage sampling. Philadelphia meeting of the Institute of Mathematical Statistics.

Aggarwal, O.P. 1966. Ratio and regression estimators as minimax procedures for estimating the mean of a population. Upton, New York meeting of the Institute of Mathematical Statistics.

Aitchison, J. 1964. Bayesian tolerance regions. Journal of the Royal Statistical Society. 26:164-175.

Amster, Sigmund J. 1962. A modified Bayes stopping rule. Chapel Hill meeting of the Institute of Mathematical Statistics.

Amster, Sigmund J. 1963. A modified Bayes stopping rule. Annals of Mathematical Statistics. 34:1404-1413.

Anderson, T.W. 1964. On Bayes procedures for a problem with choice of observations. Annals of Mathematical Statistics. 35:1128-1135.

Anscombe, F.J. 1957. Dependence of the fiducial argument on the sampling rule. Biometrika. 44:464-469.

Anscombe, F.J., and R. J. Aumann. 1963. A definition of subjective probability. Annals of Mathematical Statistics. 34:199-205.

Antelman, G.R. 1965. Insensitivity to non-optimal design in Bayesian decision theory. Journal of the American Statistical Association. 60:584-601.

Arrow, K.J., M.A. Blackwell, and M.A. Girshick. 1948. Bayes and minimax solutions of sequential decision problems. AD603-804.

Baranchik, A.J. 1965. Admissible minimax estimation of the mean of multivariate normal random variable. Philadelphia meeting of the Institute of Mathematical Statistics.

Barnard. 1937. Sequential tests in industrial statistics. Journal of Royal Statistical Society. 8:1.

Barnard, G. A. 1950. On the Fisher-Behrens test. Biometrika. 37:203-207.

Bartholomew, D. 1965. A comparison of some Bayesian and frequentist inferences. Biometrika. 52:19-35.

Bartholomew, D. J., and E. E. Bassett. 1966. A comparison of some Bayesian and frequentist inferences II. Biometrika. 53:262-264.

Bartlett, M. S. 1957. A comment on D. V. Lindley's statistical paradox. Biometrika. 44:533-534.

Bayes, T. 1763. An essay toward solving a problem in the doctrine of chances. Phil. Trans. of the Royal Statistical Society of London. 53:370-418.

Behara, Minaketan, and Gunter Menges. 1962. Bayes risk of sequential sampling decisions. Statistische Hefte, Germany. 3(1):39-61.

Behrens, W. V. 1929. Ein beitrag sur fehlerberechnung bei wenigen beobachtungen (A contribution to the error calculation with few observations). Landwirtschaftliche Jahrbucher. 18:807-837.

Berk, Robert H. 1966. Limiting behavior of posterior distribution when the model is incorrect. Annals of Mathematical Statistics. 37:51-58.

Berkson, Joseph. 1930. Bayes theorem. Annals of Mathematical Statistics. 1:42-56.

Bhat, B. R. 1964. Bayes solution of sequential decision problem for Markov dependent observations. Annals of Mathematical Statistics. 35:1656-1662.

Bickel, Peter, and Joseph Yahav. 1966 a. Asymptotically optimal sequential Bayes estimates. Abstract in Annals of Mathematical Statistics.

Bickel, Peter, and Joseph Yahav. 1966b. Asymptotic theory of Bayes solutions. Abstract in Annals of Mathematical Statistics. 37:550.

Birnbaum, A. 1962. On the foundations of statistical inference. Journal of the American Statistical Association. 57:296-306.

Blake, Archie. 1948. Transformations induced by series approximation of prior probability amplitude. Seattle meeting of the Institute of Mathematical Statistics.

Bland, Richard Park. A minimum average risk solution for the problem of choosing the largest mean. AD-260-717.

Bloch, D.A., and G. S. Watson. 1966. A Bayesian study of the multinomial distribution. AD-641-800.

Bloch, D.A., and G.S. Watson. 1966b. A Bayesian study of the multinomial distribution. Upton, New York meeting of the Institute of Mathematical Statistics.

Boehm, Geo. A. 1964. The science of being almost certain. Fortune. 69(2):104-107.

Bohrer, Robert. 1965. Bayes sequential design in the two action case. Tallahassee, Florida, meeting of the Institute of Mathematical Statistics.

Bohrer, Robert. 1966. On Bayes sequential design with two random variables. Biometrika. 53:469-476.

Borch, Karl. 1965. The economics of uncertainty XIII. AD-618-344.

Box, G.E.P., and G.C. Tiao. 1962. A further look at robustness via Bayes's theorem. Biometrika. 49:419-433.

Box, G.E.P., and G.C. Tiao. 1963. A Bayesian approach to the importance of assumptions applied to the comparison of variances. Eugene, Oregon, meeting of the Institute of Mathematical Statistics.

Box, G.E.P., and G.C. Tiao. 1964a. A Bayesian approach to the importance of assumptions applied to the comparison of variances. Biometrika. 51:153-168.

Box, G.E.P., and G.C. Tiao. 1964b. A note on criterion robustness and inference robustness. Biometrika. 51:169-173.

Box, G.E.P., and N.R. Draper. 1965a. The Bayesian estimation of common parameters from several responses. Biometrika. 52:355-365.

Box, G. E. P., and G. C. Tiao. 1965b. Multiparameter problems from a Bayesian point of view. Annals of Mathematical Statistics. 36:1468-1482.

Bracken, Jerome. 1966. Percentage points of the beta distribution for use in Bayesian analysis of Bernoulli processes. Technometrics. 8:687-694.

Braga-Illa, Alvise. 1964. A simple approach to the Bayes choice criterion: The method of critical probabilities. The Journal of the American Statistical Association. 59:1227-1230.

Breakwell, John V. 1960. Bayes approach to control of fraction defective. Seattle meeting of the Institute of Mathematical Statistics.

Breakwell, J. V. 1961. Minimax tests for the rate of a poisson process and the bias rate of a normal process. Sankhya (Series A, India) 23(2):161-182.

Brinton, B. C., and C. E. Garner. 1966. IBM 360 computer program for input/output, optimization and coding interpretation for user supplied routines. Thiokol Chemical Corporation, Wasatch Division. Program No. 3081.

Bross, I. 1950. Fiducial intervals for variance components. Biometrika. 37:136-144.

Brown, Morton B. 1965. A secondarily Bayes approach to the two-means problem. Philadelphia meeting of the Institute of Mathematical Statistics.

Buehler, Robert J. 1958. Use of prior knowledge in finding the maximum response. Cambridge, Mass. meeting of the Institute of Mathematical Statistics.

Buehler, R. J. 1959. Some validity criteria for statistical inferences. Annals of Mathematical Statistics. 30:845-863.

Buehler, Robert J. 1963. A new test of fiducial consistency. Madison, Wisc. meeting of the Institute of Mathematical Statistics.

Cam, Lucien Le. 1952. On asymptotic properties of estimates. Eugene, Oregon, meeting of Institute of Mathematical Statistics.

Camp, Burton H. 1942. Some recent advances in mathematical statistics. Annals of Mathematical Statistics. 13:62-73.

Chand, Uttam. 1950. Distributions related to comparison of two means and two regression coefficients. Annals of Mathematical Statistics. 21:507.

Chernoff, Herman. 1959. Sequential design of experiments. Annals of Mathematical Statistics. 30:755-770.

Chernoff, Herman, and S.N. Ray. 1964. A Bayes sequential sampling inspection plan. Amherst, Mass. meeting of the Institute of Mathematical Statistics.

Chernoff, Herman, and S.N. Ray. 1965. A Bayes sequential sampling inspection plan. Annals of Mathematical Statistics. 36:1387-1407.

Clutton-Brock, M. 1965. Using the observations to estimate the prior distribution. Journal of the Royal Statistical Society. 27:17-27.

Court, L.M. 1945. The minimax character of the Neyman-Pearson critical region. Iowa City meeting of the Institute of Mathematical Statistics.

Cox, D.R. 1960. Serial sampling acceptance schemes derived from Bayes's theorem. Technometrics. 2:353-360.

Dane, C.W. 1965. Statistical decision theory and its application to forest engineering. Journal of Forestry. 63(4):276-279.

Davis, R.C. 1951. The asymptotic properties of Bayes estimates. Santa Monica meeting of the Institute of Mathematical Statistics.

Deeley, John J. 1966. Non-parametric empirical Bayes procedures for selecting the best of k populations. Lafayette, Indiana, meeting of the Institute of Mathematical Statistics.

Deffenbaugh, Robert M. 1964. Investigation of the statistical decision process for anti-submarine warfare tactical decisions. AD-481-269.

DeGroot, M.H. 1962. Uncertainty, information, and sequential experiments. Annals of Mathematical Statistics. 33:404-419.

DeGroot, M.H., and M.M. Rao. 1963. Bayes estimation with convex loss. Annals of Mathematical Statistics. 34:839-846.

Dempster, A.P. 1963a. On a paradox concerning inference about a covariance matrix. Annals of Mathematical Statistics. 34:1414-1418.

Dempster, A.P. 1963b. On direct probabilities. Journal of the Royal Statistical Society. 25:100-110.

Dempster, A.P. 1966. New methods for reasoning towards posterior distributions based on sample data. Annals of Mathematical Statistics. 37:355-374.

Dickey, J.M. 1966. On a multivariate generalization of the Behrens-Fisher distributions. Upton, New York meeting of the Institute of Mathematical Statistics.

Draper, Norman R., and Irwin Guttman. 1966a. Some Bayesian statified two-stage sampling results. London meeting of the Institute of Mathematical Statistics.

Draper, Norman R., and Irwin Guttman. 1966b. Unequal group variances in the fixed-effects one-way analysis of variance: A Bayesian sidelight. Biometrika. 53:27-35.

Duncan, David B. 1961. Bayes rules for a common multiple comparisons problem and related student-t problems. The Annals of Mathematical Statistics. (U.S.) December. 32:1013-1033.

Duncan, D.B. 1965. A Bayesian approach to multiple comparisons. Technometrics. 7:171-222.

Dvoretzky, A., J. Kiefer, and J. Wolfowitz. 1953. Sequential decision problems for processes with continuous time parameter problems of estimation. Annals of Mathematical Statistics. 24:403-415.

Earnest, Charles M. 1966. Estimating reliability after corrective action: A Bayesian viewpoint. AD-487-411.

Edwards, Ward, Harold Lindman, and Leonard J. Savage. 1962. Bayesian statistical inference for psychological research. AD-420-287.

Edwards. (no date) Probabilistic information processing by men, machines and man-machines systems. AD-428-727.

Eisenberg, H. B. 1960. Multi-stage Bayesian lot-by-lot sampling inspection. Eastern Regional Meeting of the Institute of Mathematical Statistics.

Eisenhart, C. 1939. A note on a priori information. Annals of Mathematical Statistics. 10:390-393.

Evans, I. G. 1964. Bayesian estimation of the variance of a normal distribution. Journal of the Royal Statistical Society. 26:63-68.

Evans, I. G. 1965. Bayesian estimation of parameters of a multivariate normal distribution. Journal of the Royal Statistical Society. 27:279-283.

Fabius, J. 1964. Asymptotic behavior of Bayes' estimates. Annals of Mathematical Statistics. 35:846-856.

Feeney, G. J., and C. Sherbrooke. 1965a. An objective Bayes approach for inventory decisions. AD-613-470.

Feeney, G. J., and C. Sherbrooke. 1965b. A system approach to basic stockage of recoverable items. AD-627-644.

Feldman, Dorian. 1962. Contributions to the "Two-armed bandit" problem. Annals of Mathematical Statistics. 33:847-856.

Finetti de, B. 1937. La Prevision: Ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincare. 7:1-68.

Fisher, R. A. 1930. Inverse probability. Proceedings of the Cambridge Philosophical Society. 26:528-535.

Fisher, R. A. 1934. The asymptotic approach to Behrens' integral with further tables for the d-test of significance. Annals of Eugenics. 2:141.

Fisher, R. A. 1935a. Design of experiments. Hafner Publishing Company, New York.

Fisher, R. A. 1935b. The fiducial argument in statistical inference. Annals of Eugenics. 6(4):391-398.

Fisher, R.A. 1936. Uncertain inference. Proceedings of the American Academy of Arts & Sciences. 71(4):245-258.

Fisher, R.A. 1939. The comparison of samples with possibly unequal variances. Annals of Eugenics 9(2):174-180.

Fisher, R.A. 1950. Contributions to mathematical statistics. John Wiley and Sons. New York.

Fisher, R.A. 1953. Dispersion on a sphere. Proceedings of the Royal Society. 217:295-305.

Fisher, R.A. 1956. Statistical methods and scientific inference. Oliver and Boyd, Edinburgh.

Fisher, R.A. 1959. Mathematical probability in the natural sciences. Technometrics. 1:21-29.

Fisher, R.A. 1962. Some examples of Bayes' method of the experimental determination of probabilities a priori. Journal of the Royal Statistical Society. 24:118-124.

Fraser, D.A.S. 1958. Statistics: An introduction. John Wiley and Sons, New York, 243 pages.

Fraser, D.A.S. 1961. The fiducial method and invariance. Biometrika. 48:261-280.

Fraser, D.A.S. 1962. On the consistency of the fiducial method. Journal of the Royal Statistical Society. 24:425-434.

Fraser, D.A.S. 1964. Fiducial inference for location and scale parameters. Biometrika. 51:17-24.

Fraser, D.A.S. 1965. Fiducial consistency and group structure. Biometrika. 52:55-66.

Freedman, David A. 1963. On the asymptotic behavior of Bayes' estimates in the discrete case. Annals of Mathematical Statistics. 34:1386-1403.

Freedman, David A. 1965. On the asymptotic behavior of Bayes estimates in the discrete case II. Annals of Mathematical Statistics. 36:454-456.

Freedman, David A. 1966. A note on mutual singularity of priors. Annals of Mathematical Statistics. 37:375-381.

Geisser, Seymour. 1963. Posterior odds for multivariate normal classifications. Eugene, Oregon, meeting of the Institute of Mathematical Statistics.

Geisser, Seymour. 1964. Posterior odds for multivariate normal classifications. Journal of the Royal Statistical Society. 26:69-76.

Geisser, Seymour. 1965a. A Bayes approach for combining correlated estimates. Tallahassee, Florida, meeting of the Institute of Mathematical Statistics.

Geisser, Seymour. 1965b. Bayesian estimation in multivariate analysis. Annals of Mathematical Statistics. 36:150-159.

Geisser, Seymour. 1966. Estimation associated with linear discriminants. Upton, New York meeting of the Institute of Mathematical Statistics.

Girshick, M.A., and H. Rubin. 1950. A Bayes approach to a quality control model. Chicago meeting of the Institute of Mathematical Statistics.

Girshick, M.A., and L. Savage. 1951. What is Bayes postulate? Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics & Probability. p53-73.

Girshick, M.A., and H. Rubin. 1952. A Bayes approach to a quality control. Annals of Mathematical Statistics. 23:114-125.

Godambe, V.P., and V. Joshi. 1965a. Admissibility and Bayes estimation in sampling finite populations. Abstract in Annals of Mathematical Statistics. 36:364.

Godambe, V.P., and V. Joshi. 1965b. Admissibility and Bayes estimation in sampling finite poupulations. I. Annals of Mathematical Statistics. 36:1707-1722.

Godambe, V.P. 1966a. The empirical Bayes shortest confidence intervals for estimating the mean of a finite populations. Abstract in Annals of Mathematical Statistics. 37:772.

Godambe, V.P. 1966b. A new approach to sampling from finite populations I--sufficiency and linear estimation. Abstract in Annals of Mathematical Statistics. 37:552.

Godambe, V.P. 1966c. Bayesian sufficiency in survey-sampling. Abstract in Annals of Mathematical Statistics. 37:1414.

Godambe, V.P. 1966d. Statistical inference. Abstract in Annals of Mathematical Statistics. 37:1414.

Good, I.J. 1950. Probability and the weighing of evidence. Griffin Ltd., New York.

Good, I.J. 1956. The surprise index for the multivariate normal distribution. Annals of Mathematical Statistics. 27:1130-1135.

Green, Paul E. 1962. Bayesian decision theory in advertising. Journal of Advertising Research. 2(4):33-41.

Green, Paul E. 1963. Bayesian decision theory in pricing strategy. Journal of Marketing. 27(1):5-14.

Grundy, P.M. 1956. Fiducial distributions and prior distributions: An example in which the former cannot be associated with the latter. Journal Royal Statistical Society. 18:217-221.

Guthrie, D., Jr., and M.V. Johns. 1959. Bayes acceptance sampling procedures for large lots. Annals of Mathematical Statistics. 30:896-925.

Guttman, Irwin, and George C. Tiao. 1964. A Bayesian approach to some best population problems. Annals of Mathematical Statistics. 35:825-835.

Hald, Anders. 1960. The compound hypergeometric distribution and a system of single sampling inspection plans based on prior distributions and costs. Technometrics. 2:275-340.

Hald, A. 1964. Acceptance probability and minimum average costs. AD-609-993.

Hald, A. 1965. Bayesian single sampling attribute plans for discrete prior distributions. Abstract in Annals of Mathematical Statistics. 36:1082.

Hald, A. 1966. Asymptotic properties of Bayesian single sampling plans. AD-635-451.

Hall, W.J., and Melvin R. Novick. 1963. A note on classical and Bayesian prediction intervals for location, scale and regression models. Eugene, Oregon meeting of the Institute of Mathematical Statistics.

Hansel, G. , and Grouchko. 1965. Bayesian sequential estimation. Revue de Statistique Appliquee (France). 13(3):67-81.

Harris, Lawrence. 1967. A Bayesian approach to minimizing production testing requirements. 31st National Meeting of the Opeations Research Society of America.

Hartigan, J. A. 1964. Invariant prior distributions. Annals of Mathematical Statistics. 34:836-845.

Healy, M. J. R. 1963. Fiducial limits for a variance component. Journal of the Royal Statistical Society. 25:128-130.

Henderson, R. 1932. A postulate for observations. Annals of Mathematical Statistics. 3:32-39.

Hendricks, Walter A. 1965. Estimation of the probability that an observation will fall into a specified class. Journal of the American Statistical Association. 59:225-232.

Hill, Bruce M. 1963. The three parameter log normal distribution and Bayesian analysis of a point source epidemic. Journal of the American Statistical Association. 58:72-84.

Hirshleifer, Jack. 1961. The Bayesian approach to statistical decision an exposition. Journal of Business. 34:471-489.

Hoeffding, Wassily. 1960. Lower bounds for the expected sample size and the average risk of a sequential procedure. Annals of Mathematical Statistics. 31:353-368.

Hollander, Myles. 1965. Rand tests for randomized blocks when the alternatives have a prior ordering. Abstract in Annals of Mathematical Statistics. 36:1082.

Howard, Ronald A. 1963. Stochastic process models of consumer behavior. Journal of Advertising Research. 3(3):35-42.

Iglehart, Donald L. 1964. The dynamic inventory problem with unknown demand distribution. Management Science, 10(3):429-440.

Jeffreys, H. 1957. Scientific inference. Cambridge University Press, New York.

Jeffreys, H. 1961. Theory of probability. 3rd edition. Clarendon Press, Oxford.

Johns, M. V. 1955. Empirical Bayes estimation. New York meeting of the Institute of Mathematical Statistics.

Johns, M. V. Jr. 1957. Non-parametric empirical Bayes procedures. Annals of Mathematical Statistics. 28:649-669.

Joshi, V. M. 1965. Admissibility and Bayes estimation in sampling finite populations. II. Annals of Mathematical Statistics. 36:1723-1729.

Joshi, V. M. 1965b. Admissibility and Bayes estimation in sampling finite populations. III. Annals of Mathematical Statistics. 36:1730-1742.

Joshi, V.M. 1966. Admissibility and Bayes estimation in sampling finite populations. IV. Annals of Mathematical Statistics. 37:1658-1670.

Kale, B.K. 1963. Decisions opposite markov chains. Statistische Hefte (Germany). 4(2):172-177.

Kantor, Michael. 1966. Estimation of the mean of the multivariate normal distribution with an empirical Bayes application. New Brunswick, New Jersey meeting of the Institute of Mathematical Statistics.

Karlin, Samuel, and S. Johnson. 1954. A Bayes model in sequential design. AD-604-145.

Kendall, M.G. 1959. The advanced theory of statistics, 3rd edition. Hafner Publishing Co., New York, New York.

Kerridge, D. 1963. Bounds for the frequency of misleading Bayes inferences. Annals of Mathematical Statistics. 34:1109-1110.

Keynes, J.M. 1962. A treatise on probability. Harper and Row, New York.

Kiefer, J., and R. Schwartz. 1965. Admissibile Bayes character of T^2-, R^2-, and other fully invariate tests for classical multivariate normal problems. Annals of Mathematical Statistics. 36:747-770.

Kraft, Charles H., and Constance V. Eeden. 1964. Bayesian bio-assay. Annals of Mathematical Statistics. 35:886-890.

Krutchkoff, R. G., and F. Rutherford. 1965. Some parametric empirical Bayes techniques. Philadelphia meeting of the Institute of Mathematical Statistics.

Kupperman, Morton. 1958. Probabilities of hypotheses and information statistics in sampling from exponential-class populations. Annals of Mathematical Statistics. 29:575-575.

Leone, F. C. 1949. Some extensions of Bayes' theorem. New York meeting of Institute of Mathematical Statistics.

Lever, Wm. E. 1965. A chi-square decision procedure using prior information. Tallahassee, Florida, meeting of the Institute of Mathematical Statistics.

Lindley, D. V. 1953. Statistical inference. Journal of the Royal Statistical Society. 15:30-76.

Lindley, D. V. 1957. A statistical paradox. Biometrika. 44:187-192

Lindley, D. V. 1957b. Binomial sampling schemes & concept of information. Biometrika. 44:179-186.

Lindley, D. V. 1958. Fiducial distributions and Bayes' theorem. Journal of the Royal Statistical Society. 20:102-107.

Lindley, D. V., D. A. East, and P. Hamilton 1960. Table for making inferences about the variance of a normal distribution. Biometrika. 47:433-437.

Lindley, D. V. 1961. The use of prior probability distributions in statistical inference and decision. 4th Berkeley Symposium on Mathematical Statistics & Probability Vol. Berkeley, California. University of California Press. pp. 453-468.

Lindley, D. V. 1963. Bayesian inference for contingency tables. Ottawa meeting of the Institute of Mathematical Statistics.

Lindley, Dennis V. 1964. The Bayesian analysis of contingency tables. Annals of Mathematical Statistics. 35:1622-1643.

Lindley, D. V. 1965. Probability and statistics from a Bayesian viewpoint. Cambridge University Press, New York .

Lorden, Gary. 1966. Bayes risk of asymptotically Bayes sequential tests. Lafayette, Indiana, meeting of the Institute of Mathematical Statistics.

Maritz, J.S. 1966. Smooth empirical Bayes estimation for one-parameter discrete distributions. Biometrika. 53:417-429.

Martin, James J. 1965. Some Bayesian decision problems in a Markov chain. Abstract in Annals of Mathematical Statistics. 36:1613.

Mayberry, John P. 1964. Alternate payoff-functions in statistical decision theory. Proceedings of the 3rd International Conference on Operations Research, Oslo. English Universities Press, Ltd. London. pp 548-558.

Mazumdar, Mainak 1966. Optimal sequential plans based on prior distributions and costs. AD-634-341.

McGhee, R.B., and R. Walford. 1965. A Bayesian solution of the re-entry tracking problem. AD-475-450.

McGlothlin, Wm. H. 1963. Development of Bayesian parameters for spare parts demand prediction. AD-411 356.

McGlothlin, W.H., and Eloise Bean. (No date). Application of Bayes to spare parts demand prediction. AD-255-161.

Mises, R. Von. 1941. On the foundations of probability and statistics. Annals of Mathematical Statstics. 12:191-205.

Mises, R.V. 1942. On the correct use of Bayes' formula. Annals of Mathematical Statistics. 13:156-165.

Miyasawa, Koichi. 1962. Decision theory and risk. Or, Juse (Japan), 6(5):22-25.

Molina, Edward C. 1931. Bayes' theorem--an expository presentation. Annals of Mathematical Statistics. 2:23-37.

Moriguti, S., and H. Robbins. 1962. A Bayes test of $p \leq 1/2$ vs $p > 1/2$. Rep. Stat. Applied Research of the Union of Japanese Scientists & Engineers. 9:39-60.

Mosteller, Frederick, and David Wallace. 1963. Inference in an authorship problem. Journal of the American Statistical Association. 58:275-309.

Mosteller, Frederick, and David L. Wallace. 1964. Inference and disputed authorship; the federalist. Addison-Wesley Inc., Reading, Massachusetts. 287 pages.

Neumann, J. von., and Oskar Morgenstern. 1947. Theory of games and economic behavior. Princeton University Press, Princeton, N.J.

Neyman, J., and E. Pearson. 1933. On the testing of statistical hypotheses in relation to probabilities apriori. Proceedings of the Cambridge Philosophical Society. 29:492-510.

Neyman, J. 1956. An empirical Bayes approach to statistics. Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability. University of California Press.

Novick, Melvin R. 1962. A Bayesian indifference rule. Chapel Hill meeting of the Institute of Mathematical Statistics.

Novick, M. R. 1963. A Bayesian indifference procedure. Institute of Statistics Mimeograph Series #345. University of North Carolina.

Novick, M.R. 1964. On the choice of prior distribution. Research Bulletin 64-23. Educational Testing Service. Princeton, N.J.

Novick, M.R., and J.E. Girzzle. 1965. A Bayesian approach to the analysis of data from clinical trials. Journal of the American Statistical Association. 60:81-96.

Novick, M.R. 1965. Bayesian indifference procedure. Journal of the American Statistical Association. 60:1104-1117.

Novick, Melvin R. 1966. Multiparameter consistency and Bayesian indifference specification. London meeting of the Institute of Mathematical Statistics.

Parzen, Emanuel. A new approach to the synthesis of optimal smoothing and prediction systems. AD-240-426.

Patil. 1965. Approximation to the Behrens-Fisher distributions. Biometrika. 52:267-271.

Pearson, E. 1925. Bayes' theorem examined in the light of experimental sampling. Biometrika. 17:388-442.

Pearson, E.S. 1962. Some thoughts on statistical inference. Annals of Mathematical Statistics. 33:394-403.

Peers, H.W. 1965. On confidence points and Bayesian probability points in the case of several parameters. Journal Royal Statistical Society. 27:9-16.

Pfanzagl, J. 1963. Sampling procedures based on prior distributions and costs. Technometrics. 5:47-61.

Posten, H.O. 1963. Robustness of uniform Bayesian encoding. Technometrics. 5:122-125.

Pratt, J.W., H. Raiffa, and R. Schlaifer. 1964. The foundations of decision under uncertainty: An elementary exposition. Journal of the American Statistical Association. 59:353-375.

Pratt, John W. 1965. Bayesian interpretation of standard inference statements. Journal Royal Statistical Society. 27:169-192.

Pratt, J.W. 1966. The outer needle of some Bayes sequential continuation regions. Biometrika. 53:455-467.

Pugachev, V.S. (No date). An effective method of obtaining Bayes' solution. Document 62-10140.

Pugachev, V.S. (No date). A method of determination of the optimum system from the general Bayes criterion. Document 61-27534.

Raiffa, H., and R. Schlaifer. 1961. Applied statistical decision theory. Harvard Business School, Boston, Mass.

Ramachandran, K.V., and C.G. Khatri. 1957. On a decision procedure based on the Tukey statistic. Annals of Mathematical Statistics. 28:802-806.

Ramsey, F.P. 1931. The foundations of mathematics and other logical essays. Routledge and Kegan Paul Ltd., London.

Ray, Sudhindra Narayan. 1963. Bayes sequential procedures for some binomial problems. Cambridge, Mass. meeting of the Institute of Mathematical Statistics.

Ray, S.N. 1965. Bounds on the maximum sample size of a Bayes sequential procedure. Annals of Mathematical Statistics. 36:859-878.

Richardson, Wyman. 1966. Decision-theoretic principles of the design of verification systems. AD-481-519.

Robbins, Herbert. 1956. An empirical Bayes approach to statistics. Proceedings of the 3rd. Berkeley symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, California. 157-164.

Robbins, H. 1964. The empirical Bayes approach to statistical decision problems. Annals of Mathematical Statistics. 35:1-20.

Roberts, Harry V. 1960. The new business statistics. Journal of Business. 33:21-30.

Roberts, Harry V. 1963. Bayesian statistics in marketing. Journal of Marketing. 27(1):1-4.

Rubel, Ronald A. 1967. Decision analysis and the treatment of a sore throat. Operations Research Society of America meeting.

Rubin, Herman, and Sethuraman. 1965a. Bayes risk efficiency. Berkeley, California meeting of the Institute of Mathematical Statistics.

Rubin, Herman, and Sethuraman. 1965b. Some further results in Bayes risk efficiency. Philadelphia meeting of the Institute of Mathematical Statistics.

Rutherford, J.R., and R.G. Krutchkoff. 1965. Empirical Bayes estimation of prior and posterior distribution. Philadelphia meeting of the Institute of Mathematical Statistics.

Rutherford, J.R. 1966a. Monte Carlo analysis of the rate of convergence of some empirical Bayes point estimators. Upton, N.Y. meeting of the Institute of Mathematical Statistics.

Rutherford, J.R. 1966b. The empirical Bayes approach: Estimation of posterior quantities. New Brunswick, N.J. meeting of the Institute of Mathematical Statistics.

Ryzin, J. Van. 1966. Bayes risk consistency of classification procedures using density estimation. Lafayette, Indiana, meeting of the Institute of Mathematical Statistics.

Sacks, Jerome. 1958. Generalized Bayes solutions in estimation problems. Cambridge, Mass. meeting of Institute of Mathematical Statistics.

Sacks, Jerome. 1963. Generalized Bayes solutions in estimation problems. Annals of Mathematical Statistics. 34:751-768.

Sakaguchi, Minoru. 1961. Dynamic programming of some sequential sampling design. Keiei-Kagaku (Japan), 4(3):170-182.

Samuel, Ester. 1963. An empirical Bayes approach to the testing of certain parametric hypotheses. Annals of Mathematical Statistics. 34:1370-1385.

Sarkadi, K. 1957. On the distribution of the number of exceedances. Annals of Mathematical Statistics. 28:1021-1023.

Sardal, Carl-Erik. 1965. Derivation of a class of frequency distributions via Bayes's theorem. Journal Royal Statistical Society. 27:290-300.

Savage, Leonard J. 1954. Foundations of statistics. John Wiley & Sons Inc., New York.

Savage, Leonard J. 1961a. Decision and information processes. Macmillan & Co., New York.

Savage, L. J. 1961b. The foundations of statistics reconsidered. Proceedings of the 4th Berkeley Symposium on Probability & Statistics. University of California Press. 575-586.

Savage, Leonard J. 1961c. The subjective basis of statistical practice. University of Michigan Manuscript.

Savage, Leonard J. 1962a. Bayesian statistics in decision and information processes. Macmillan & Co., New York.

Savage, Leonard J. 1962b. Foundations of statistical inference. John Wiley & Sons, New York.

Saxena, K. M. Lal. 1966. Some nonparametric Bayesian decision problems. Lafayette, Indiana, meeting of the Institute of Mathematical Statistics.

Scarf, Herbert. 1959. Bayes solutions of the statistical inventory problem. Annals of Mathematical Statistics. 30:490-508.

Scarf, Herbert, and J. VanderVeer. (no date). A comparative analysis of the Bayes inventory policy. AD 408-523.

Schaefer, M. B. 1948. The employment of marked members in the estimation of animal populations. Seattle Meeting of Institute of Mathematical Statistics.

Schlaifer, Robert. 1959. Probability and statistics for business decisions. McGraw Hill, New York.

Schotta, Charles Jr., and R. Hoffman. 1967. A priori decision functions for educational evaluation. Operations Research Society of America.

Schwartz, Sideon. (no date). Asymptotic shapes of optimal sampling regions in sequential testing. AD 245-786.

Shen, Ching Lai. 1936. Fundamentals of the theory of inverse sampling. Annals of Mathematical Statistics. 7:62-112.

Skibinsky, Morris. 1960. Some properties of a class of Bayes two-stage tests. Annals of Mathematical Statistics. 31:332-351.

Skibinsky, Morris, and Louis J. Cote. 1962. On the inadmissibility of some standard estimates in the presence of prior information. Minneapolis meeting of the Institute of Mathematical Statistics.

Skibinsky, M., and L. J. Cote. 1963. On the inadmissibility of some standard estimates in the presence of prior information. Annals of Mathematical Statistics. 34:539-548.

Skibinsky, M. 1963. Bio-assay with prior information. Madison, Wisconsin meeting of the Institute of Mathematical Statistics.

Skibinsky, Morris. 1965. Minimax prediction of random probabilities. Berkeley, California, meeting of the Institute of Mathematical Statistics.

Smith, Armand V., Jr. 1964. Comparison of two drugs in multistage sampling using Bayesian decision theory. Manhattan, Kansas, meeting of the Institute of Mathematical Statistics.

Smith, W. E. 1966. An a posteriori probability method for solving an overdetermined system of equations. Technometrics. 8:675-686.

Soland, Richard M. 1966a. Bayesian anaysis of the Weibull process with unknown scale parameter. AD-643-816.

Soland, Richard M. 1966b. Optimal Bayesian stratified sampling by nonlinear programming. RAC-TP-221. Research Analysis Corporation.

Springer, M. D., and W. Thompson. 1966a. Bayesian confidence limits for the product of N binomial parameters. Biometrika. 53:611-613.

Springer, Melvin D., and William Thompson. 1966b. Bayesian confidence limits for the reliability of cascaded exponential subsystems. Los Angeles meeting of the Institute of Mathematical Statistics.

Springer, M. D., and W. Thompson. 1966c. On Bayesian confidence limits for the reliability of redundant subsystems having exponential distribution of life when subsystem tests are terminated at first failure. General Motors Defense Research Laboratories.

Sprott, D. A. 1960. Necessary restrictions for distributions a posteriori. Journal of the Royal Statistical Society. 22:312-318.

Steinhaus, H. 1957. The problem of estimation. Annals of Mathematical Statistics. 28:633-648.

Sterne, T. E. 1954. Some remarks on confidence or fiducial limits. Biometrika. 41:275-278.

Stone, M. 1959. Application of a measure of information to the design and comparison of regression experiments. Annals of Mathematical Statistics. 30:55-70.

Stone, M. 1963. The posterior distribution. Annals of Mathematical Statistics. 34:568-573.

Stone, M. 1964. Comments on a posterior distribution of Geisser and Cornfield. Journal of Royal Statistical Society. 26:274-276.

Stone, M., and B. G. Springer. 1965. A paradox involving quasi prior distributions. Biometrika. 52:623-627.

Stone. 1954. Right Haar measure for convergence in probability to quasi posterior distributions. Journal of the American Mathematical Society. 36:440-453.

Sukhatme, P. V. 1938. On Fisher & Behrens' test of significance for the difference in means of two normal samples. Sankhya. 4:39-48.

Swain, Donald D. 1965. Bounds and rates of convergence for the extended compound estimation problem in the sequence case. Philadelphia meeting of the Institute of Mathematical Statistics.

Tada, Daxuo. 1964. On the distribution of search effort. Keviei Kagaku (Japan), 7(2): 81-86.

Thatcher, A. R. 1964. Relationships between Bayesian and Confidence limits for predictions. Journal of the Royal Statistical Society. 26:164-175.

Theil, H. 1963. On the use of incomplete prior information in regression analysis. Journal of the American Statistical Association. 58:401-414.

Thompson, Wm. R. 1936. On confidence ranges for the median and other expectation distributions for populations of unknown distribution forms. Annals of Mathematical Statistics. 7:122-128.

Tiao, George C., and A. Zellner. 1964a. Bayes' Theorem and the use of prior knowledge in regression analysis. Biometrika. 51:219-221.

Tiao, G., and A. Zellner. 1964b. On the Bayesian estimation of multivariate regression. Journal of the Royal Statistical Society. 26:277-285.

Tiao, G. C., and W. Tan. 1965. Bayesian analysis of variance. I. posterior distribution of variance-componets. Biometrika. 52:37-53.

Tiao, George C. 1965. Bayesian comparison of means of a mixed model with application to regression analysis. Philadelphia meeting of the Institute of Mathematical Statistics.

Tiao, G. C., and W. Tan. 1966. Bayesian analyses of random effect models in the analysis of variance II: Effect of autocorrelated errors. Biometrika. 53:477.

Tiao, George C. 1966a. Bayesian analysis of hierarchical design model. Upton, N. Y. meeting of the Institute of Mathematical Statistics.

Tiao, G. C. 1966b. Bayesian comparison of means of a mixed model with application to regression analysis. Biometrika 53:11-25.

Torrens-Ibern, J. 1964. The application of Bayes theorem to the adaptive control processes. Bern, Switzerland, meeting of the Institute of Mathematical Statistics.

Trybula, Stanislaw. 1958. Some problems of simultaneous minimax estimation. Annals of Mathematical Statistics. 29:245-253.

Wald, Abraham. 1939. Contributions to the theory of statistical estimation and testing hypotheses. 10: 299-326.

Wald, A. 1950. Statistical decision functions. John Wiley & Sons, New York.

Wald, A. 1955. Selected papers in statistics and probability. Stanford University Press, Stanford, California.

Walsh, John E. 1949. On the power function of the "best" t-test solution of the Behrens'-Fisher problem. Annals of Mathematical Statistics. 20:616-618.

Weiler, H. 1965. The use of incomplete beta functions for prior distributions in binomial sampling. Technometrics 7:335-347.

Weir, W. T. 1966. Proceeding of missile and space division seminar on Bayes' theorem and its application to reliability measurement. AD-481-645.

Welch, B. L. and H. W. Peers. 1963. On formulae for confidence points based on integrals on weighted likelihoods. Journal of the Royal Statistical Society. 25:318-329.

Wertheimer, Albert. 1939. A note on confidence intervals and inverse probability. Annals of Mathematical Statistics. 10:74-76.

Wetherill, G. B. 1960. Some remarks on the Bayesian solution of the single sample inspection scheme. Technometrics. 2:341-352.

Wetherill, G. B. 1961. Bayesian sequential analysis. Biometrika. 48:281-292.

Williams, E. J. 1963. A comparison of direct and fiducial arguments in the estimation of a parameter. Journal of the Royal Statistical Society. 25:95-99.

Wolfowitz, J. 1952. Maximum likelihood estimators and a posteriori distributions. East Lansing meeting of Institute of Mathematical Statistics.

Wolfowitz, J. 1962. Bayesian inference and axioms of consistent decision. Econometrica, 30(3):470-479.

Woodbury, Max A. 1955. The Bayesian inference problem in stochastic systems. New York meeting of the Institute of Mathematical Statistics.

Yahav, Joseph A. 1964. On optimal stopping. Monterey, California meeting. IMS committee for the symposium on system and control optimization.

Yates, F. 1939. An apparent inconsistency arising from test of significance based on fiducial distributions of unknown parameters. Proceedings of the Cambridge Philosophical Society. 35:579-591.

Zacks, Shelemyahu. 1966. A Bayes sequential strategy for crossing a field containing absorption points. Los Angeles meeting of the Institute of Mathematical Statistics.

Zellner, A., and G. Tiao. 1964. Bayesian analysis of the regression model with autocorrelated errors. Journal of the American Statistical Association. 59:763-778.

Table 4. Classification System

Col.	SUBJECT
1	General Considerations
2	Comparison with Classical Methods
3	Comparison with Fiducial Probability
4	Prior Distributions
5	Estimation and Bayes Estimator
6	Decision Functions
7	Confidence Intervals
8	Sampling
9	
10	Minimax & Game Theory
11	Regression
12	Analysis of Variance
13	Fisher-Behrens Test
14	Robustness
15	Invariance
16	Opposition to Bayesian Inference
17	Significance Testing
18	Multivariate Analysis
19	Nuisance Parameters
20	Dynamic Programming
21	Markov Process
22	
23	
24	
	APPLICATION
25	Reliability
26	Acceptance Sampling
27	Inventory
28	Preventive Maintenance & Spares

Col.	
29	Marketing, Advertising, & Spares
30	Medical Research
31	Games of Chance and Urn Problems
32	Classification, Authorship, Archeology
33	Search Problems
34	Agricultural Research
35	Smoothing & Noise
36	Education
37	
38	
39	Other
40	
41	
	DISTRIBUTION
42	Binomial
43	Poisson
44	Normal
45	Uniform (Rectangular)
46	Power (Exponential)
47	F
48	t
49	Log Normal
50	Negative Exponential
51	Gamma
52	Negative Binomial
53	Beta
54	Inverse
55	Other

Col.	
	SOURCE
56	Biometrika
57	Annals of Mathematical Statistics
58	Journal of American Statistical Association
59	Journal of Royal Statistical Society
60	Proceedings of Cambridge Philosophical Society
61	Technometrics
62	Other Journals and Periodicals
63	IFORS Listing
64	Berekely Symposia on Math Statistics & Probability
65	Institute of Math Statistics Symposia
66	Other Symposia
67	Documents and Reports
68	Books
	AVAILABILITY
69	Reprint
	AUTHOR
70	Box
71	Tiao
72	Savage
73	Jeffrey
74	Lindley
75	Fraser
76	Fisher
77	Other A-I
78	Other J-R
79	Other S-Z
80	Foreign Language

Literature Cited

Anderson, R. L., and T. A. Bancroft. 1952. Statistical theory in research. McGraw Hill, Inc., New York.

Behrens, W. V. 1929. Ein beitraz sur fehlerberechnung bei wenigen beobachtungen. Landw. Jb. 68:807 ff. (Original not seen; referenced in Kendall (1959).)

Berkson, Joseph. 1930. Bayes theorem. Annals of Mathematical Statistics. 1:42-56.

Boehm, George A. W. 1964. The science of being almost certain. Fortune. 69(2):104-107.

Box, G. E. P., and G. C. Tiao. 1962. A further look at robustness via Bayes's theorem. Biometrika. 49:419-433.

Brinton, B. C., and C. E. Garner. 1966. IBM 360 computer program for input/output, optimization and coding interpretation for user supplied routinges. Thiokol Chemical Corporation, Wasatch Division, Program No. 3081.

Finetti, B. de. 1937. La prevision: ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincare. 7:1-68.

Fisher, R. A. 1934. The asymptotic approach to Behrens' integral with further tables for the d-test of significance. Annals of Eugenics. 2:141.

Fisher, R. A. 1935a. The fiducial argument in statistical inference. Annals of Eugenics. 6(4):391-398.

Fisher, R. A. 1935b. Design of experiments. Hafner Publishing Co., New York.

Fisher, R. A. 1936. Uncertain inference. Proceedings of the American Academy of Arts and Sciences. 71(4):245-258.

Fisher, R. A. 1956. Statistical methods and scientific inference. Oliver and Boyd, Edinburgh.

Fisher, R. A. 1959. Mathematical probability in the natural sciences. Technometrics. 1:21-29.

Fraser, D. A. S. 1958. Statistics: an introduction. John Wiley and Sons, New York. p 242-243.

Jeffreys, H. 1957. Scientific inference. Cambridge University Press, Cambridge.

Jeffreys, H. 1961. Theory of probability. 3rd edition. Clarendon Press, Oxford.

Kendall, M. G. 1959. Vol II, Chapter 20--Fiducial inference. The advanced theory of statistics, 3rd edition. Hafner Publishing Co., New York.

Lindley, D. V. 1958. Fiducial distributions and Bayes' theorem. Journal of the Royal Statistical Society. 20:102-107.

Lindley, D. V. 1965. Introduction to probability and statistics from a Bayesian viewpoint. University Press, Cambridge.

Mises, R. V. 1942. On the correct use of Bayes' formula. Annals of Mathematical Statistics. 13:156-165.

Mosteller, Frederick, and D. L. Wallace. 1964. Inference and disputed authorship; The federalist. Addison-Westley, Inc., Reading, Mass. p 287.

Neumann, J. von, and O. Morgenstern. 1947. Theory of games and economic behaviour. Princeton University Press, Princeton, New Jersey.

Raiffa and Shlaifer. 1961. Applied statistical decision theory. Harvard Business School, Boston, Mass.

Ramsey, F. P. 1931. Truth and probability. The foundations of mathematics and other logical essays. Routledge and Kegan Paul Ltd., London. pp 156-198.

Savage, L. J. 1954. The foundations of statistics. John Wiley & Sons, Inc., New York.

Savage, L. J. 1961. The subjective basis of statistical practice. Manuscript, University of Michigan (Original not seen; results given in Tiao and Zellner, 1964).

Savage, L. J. 1962. Bayesian statistics in decision and information processes. Macmillan and Co., New York.

Sprott, D. A. 1960. Necessary restrictions for distributions a posteriori. Journal of the Royal Statistical Society. 22:312-318.

Sukhatme, P. V. 1938. On Fisher's and Behrens' test of significance for the difference in means of two normal samples. Sankhya. 4:39-48.

Theil, H. 1963. On the use of incomplete prior information in regression analysis. Journal of the American Statistical Association. 58:401-404. (Original read, but reference taken from Tiao and Zellner, 1964).

Tiao, G. C. and Arnold Zellner. 1964. Bayes's theorem and regression analysis. Biometrika. 51:220-221.

CONCLUSIONS

Future Trends

John Tukey, mathematics professor at Princeton and a member of the staff of Bell Telephone Laboratores, foresees the development of new methods of statistical analysis that will make more use of human judgement and intuition and turn data analysis into more of a creative art, in which the statistician can "listen to what the data is trying to tell him". He urges his fellow statisticians to keep an open mind and let preliminary results feed back into the analysis; to approach data with some definite questions in mind but expand and revise the questions if the study suggests such action (Boehm, 1964).

APPENDIX

Sufficient statistic is a statistic whose conditional distribution is independent of the parameters. In general,

$$p(x|\theta) = p(t(x)|\theta)\ p(x|t(x), \theta).$$

If also,

$$p(x|\theta) = p(t(x)|\theta)\ p(x|t(x))$$

then for any prior distribution the posterior distributions given $t(x)$ and given x are the same. Neyman's factorization theorem: An NSC for $t(x)$ to be sufficient for $p(x|\theta)$ is that

$$p(x|\theta) = f(t(x), \theta)\ g(x).$$

No information is lost if you replace a sample by a sufficient statistic. (Lindley, 1965).

Pivotal statistic is a random variable such that

1. its dependence on the outcome is by means of a sufficient statistic $(\bar{x}, s^2)$.
2. it depends on the parameter for which a confidence interval is wanted and no other parameter.
3. it has a fixed distribution regardless of the values of the parameters.

Invariance requires that any two estimates for a given parameter should be equal. Given a sample $\{x_n\}$ from a normal distribution with unknown mean μ and known variance σ^2. Let $f\{x_n\}$ be an estimate of μ. Now given the same sample coded by a constant, c, $\{x_n+c\}$, an estimate of $\mu+c$ is $f\{x_n+c\}$. When uncoded it yields a second estimate for the quantity μ, $f\{x_n+c\} - c$.

VITA

John Donald Laffoon

Candidate for the Degree of

Master of Science

Thesis: An Exposition on Bayesian Statistics

Major Field: Applied Statistics

Biographical Information:

Personal Data: Born in Tulsa, Oklahoma, 1 August 1934, son of John A. & Iola Laffoon; married Carolyn L. Carson 17 December 1960; two children, Jeanette and Mark.

Education: Received Bachelor of Arts degree from the Rice Institute in 1956 with a major in physics; did graduate work in mathematics at the University of Southern California, 1956-1958; took undergraduate work in business, engineering, and operations research at UCLA, 1956-1960; completed requirements for Master of Science degree in Applied Statistics at Utah State University in 1967.

Professional Experience: Employed by Thiokol Chemical Corporation as a Contract Specialist since 1965, and as a Product Research Analyst and Long Range Planning analyst, 1964-1965; Senior Engineer -- Operations analysis for North American Aviation, 1956-1964.