An Exposition on Bayesian Inference

John Laffoon
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Part of the Applied Mathematics Commons, and the Mathematics Commons

Recommended Citation
Laffoon, John, "An Exposition on Bayesian Inference" (1967). All Graduate Theses and Dissertations. 6813.
https://digitalcommons.usu.edu/etd/6813

This Thesis is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
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

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>SCHOOLS OF STATISTICAL THOUGHT</td>
<td>5</td>
</tr>
<tr>
<td>History of Bayesianism</td>
<td>5</td>
</tr>
<tr>
<td>Philosophical Views of Probability</td>
<td>12</td>
</tr>
<tr>
<td>On Prior Distributions</td>
<td>18</td>
</tr>
<tr>
<td>APPLICATIONS</td>
<td>26</td>
</tr>
<tr>
<td>Coin Flipping</td>
<td>26</td>
</tr>
<tr>
<td>Authorship Assignment</td>
<td>27</td>
</tr>
<tr>
<td>Selection</td>
<td>30</td>
</tr>
<tr>
<td>Behrens via Bayes</td>
<td>31</td>
</tr>
<tr>
<td>Bayesian Regression</td>
<td>34</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>44</td>
</tr>
<tr>
<td>Literature Survey</td>
<td>44</td>
</tr>
<tr>
<td>Literature Cited</td>
<td>70</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>73</td>
</tr>
<tr>
<td>Future Trends</td>
<td>73</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>74</td>
</tr>
<tr>
<td>VITA</td>
<td>77</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Regression Data</td>
<td>39</td>
</tr>
<tr>
<td>2.</td>
<td>Computational Factors</td>
<td>40</td>
</tr>
<tr>
<td>3.</td>
<td>Regression Coefficients</td>
<td>43</td>
</tr>
<tr>
<td>4.</td>
<td>Classification System</td>
<td>69</td>
</tr>
</tbody>
</table>
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, potential 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 ex-
perience 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.
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:

\[
\frac{n!}{a!(n-a)!} \int_0^1 \binom{n}{a} (1-p)^{n-a} f(p) \, dp
\]

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 \( \theta \), then

\[
P(\theta | X) = \frac{P(X | \theta) P(\theta)}{P(X)}
\]

will predict the value of \( \theta \) once the sample \( X \) is taken. This is the product of a predetermined (prior) probability of \( \theta \), \( P(\theta) \), and an estimate of \( X \) given \( \theta \), \( P(X | \theta) \). \( P(X) \) 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 $-\infty < x < \infty$, 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, $\theta$, 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, $\mu$, which could produce the $\bar{x}$ with a probability

$$\alpha = \Pr\left[\frac{\bar{x} - \mu}{s}\right]$$

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 \( \theta \) 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 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).

**Bayesian Revival**

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

\begin{itemize}
  \item Axiom 1 \quad 0 \leq p(A|B) \leq 1 \quad \text{and} \quad p(A|A) = 1
  \item Axiom 2 \quad \text{If the events in } A_\infty \text{ are exclusive given } B \text{ then } p(\sum A_n | B) = \sum p(A_n | B)
  \item Axiom 3 \quad p(C | A, B) p(A | B) = p(A, C | B)
\end{itemize}

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 \( p(A | B) = \frac{p(B | A) p(A)}{p(B)} \) 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 $\theta$. 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 $\theta$ and choosing a region such that $p(\theta) \cdot P(x \mid \theta)$ is valid, thus producing a "Bayes solution." By agreeing to place more emphasis on some values of $\theta$ 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 $\theta$ 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 $\sigma^2$ is known, and the prior density of $\theta$ be $N(\mu_o, \sigma_o^2)$. Then the posterior density of $\theta$ is

$$N(\mu_1, \sigma_1^{-2})$$

where

$$\mu_1 = \frac{x/\sigma^2 + \mu_o/\sigma_o^2}{1/\sigma^2 + 1/\sigma_o^2}$$

and

$$\sigma_1^{-2} = \sigma^{-2} + \sigma_o^{-2}$$

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
   \[
P(x=0) = P(x=1) = P(x=2) = \ldots = P(x=5) = \frac{1}{6}
   \]
   For this prior, the posterior probability
   \[
P(x<1 | \text{prior}) = 0.73
   \]

2. Constant Density
   \[
P(x) = x \text{ gives posterior } P(x<1 | \text{prior}) = 0.9975
   \]

3. Previous count
   - \(x = 0\) in 3086 cases
   - \(x = 1\) in 279 cases
   - \(x = 2\) in 32 cases
   - \(x = 3\) in 15 cases
   - \(x = 4\) in 5 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. Frequentists 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, 

\[
\frac{m-\theta}{h} \text{ where } m = \frac{y_{max} + y_{min}}{2}, \quad h = \frac{y_{max} - y_{min}}{2}
\]

and \( y_{max} \) and \( y_{min} \) are respectively the largest and the smallest of the observations—jointly sufficient statistics for \((\theta, \sigma)\) 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 explicit inclusion of 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 setup 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, \( \sigma \), in our problem. Using the uniform prior distribution for \( h \), however, does not provide the same answer as using the uniform prior of \( \sigma \) (i.e., \( d\sigma \leq dh \)). This lack of invariance is troublesome because of its effect on the user.

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

The same invariance argument holds for the power of \( \sigma \). If \( P(\sigma | H) \propto d\sigma \) then \( P(\sigma^n | H) \propto d\sigma^n \propto \sigma^{n-1}d\sigma \) but \( d\sigma \propto \sigma^{-1}d\sigma \). 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 \( P(\sigma | H) \propto d\sigma/\sigma \) as a statement that \( \sigma \) may have any value between 0 and \( \infty \) and compare probabilities for finite ranges of \( \sigma \), we must use \( \infty \) instead of 1 to denote certainty. However, with certainty being \( \infty \), the probability that \( \sigma \) is less than any finite number, \( P(\sigma \leq k) \), is zero. This is inconsistent with our assumption of no knowledge of \( \sigma \). \[ \frac{1}{\infty} \left[ \int_0^k d\sigma + \int_k^\infty d\sigma \right] = \frac{k}{\infty} + \frac{\infty-k}{\infty} = 0 \] .: \( P(\sigma \leq k) = 0 \).

The use of \( \infty \) 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 \( \sigma \) is defined as the limit of \( \int_a^b P(\sigma | H) \) as \( a \rightarrow 0 \), \( b \rightarrow \infty \). The function does not exist at the limit, but
approaches the limit as close as desired. That is, we can select any positive \( \varepsilon \) and find values of \( a \) and \( b \) such that
\[
\int_{a - \Delta a}^{b + \Delta b} d\sigma / \sigma - \int_a^b d\sigma / \sigma < \varepsilon
\]
for any \( \Delta a < a \) and finite \( \Delta b \).

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 experiment. 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, \( \theta \), 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 multiplied 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.

\[
p(y | \theta, \sigma, \phi) = \omega \exp\left\{ -\frac{1}{2} \left( \frac{y - \phi}{\sigma} \right)^{2/(1+\phi)} \right\}
\]

\[
\omega^{-1} = \left[ \frac{1}{1 + \frac{1}{2} (1+\phi)} \right]^{\frac{1}{2+\phi}(1+\phi)} 0 - \infty < y < \infty \quad 0 < \sigma < \infty \quad -\infty < \theta < \infty \quad -1 < \beta < 1
\]

In particular, we see that when \( \phi = 0 \), we have the normal distribution, when \( \phi = 1 \), the double exponential, and as \( \phi \) 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 \( \phi \) 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 $x_1, x_2$

- $x_1 = \text{fair coin}$
- $x_2 = \text{two headed coin}$

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

Which coin was drawn?

Using standard notation:

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

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

Bayes' Theorem says

$$P(x_i \mid H) = P(H \mid x_i) \frac{P(x_i)}{P(H)} \propto P(H \mid x_i) \quad P(x_i)$$

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 $x_2$. 
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:

\[ P_i = \text{probability before observations that Hypothesis } i \text{ is true,} \]

\[ p(x|H_i) = \text{probability of observing } x \text{ given } H_i \text{ is true.} \]

Since, by Bayes' Theorem

\[ p(H_1|x) = P_1 \frac{p(x|H_1)}{\left[ P_1 p(x|H_1) + P_2 p(x|H_2) \right]} \]

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

\[ \frac{P(x|H_1)}{P(x|H_2)} = \left( \frac{\mu_1}{\mu_2} \right)^x e^{-\mu_1} \left[ e^{-\mu_2} \right] \]

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 Madison's 14, so the prior odds could have been 43/14 \( \approx 3 \).
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 statisticians that when sufficient data is available, either technique will lead to a reasonable conclusion (Mosteller, 1964).
Another problem was conquered by Bayesian methods by Dorian Feldman at the University of California at Berkeley. 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 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, \( \mu_1, \mu_2 \) and variances, \( \sigma_1^2, \sigma_2^2 \).

Let \((x_1, s_1)\) be the sufficient statistic for \((\mu_1, \sigma_1)\) based on a sample size \(n_1\) and \((x_2, s_2)\) for \((\mu_2, \sigma_2)\) based on a sample size \(n_2\).

Now we can define a statistic developed by Fisher (1934) on earlier work by Behrens (1929)

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

\[
[ (\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2) ] \text{ is distributed normally about zero with a }
\]

variance \(\left( \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)\). By defining \(\psi\) by the expression

\[
\tan \psi = \frac{s_2}{\bar{x}_2 - \mu_2} / \frac{s_1}{\bar{x}_1 - \mu_1}
\]

we have \(\mathcal{E}\) as

\[
\mathcal{E} = t_1 \cos \psi - t_2 \sin \psi
\]

But \(\frac{x_1 - \mu_1}{s_1 / \sqrt{n_1}}\) is distributed as a t. Therefore

\[
\mathcal{E} = t_1 \cos \psi - t_2 \sin \psi
\]

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

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

Bayesian Derivation

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

\[
p(x_i | \mu, \sigma^2) = \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \right\}
\]

(Kendall, 1959).

Jeffreys introduces a modification of Bayes' postulate when a parameter, \(\theta\), can extend to infinity in only one direction, \(dF_{\theta} d\theta / \theta\)

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, \( d\mu \propto d\theta \). Applying these to our problem,

\[
P(\mu, \sigma | H) = \mu \sigma \sigma'
\]

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

\[
P(\mu_1, \mu_2, \sigma_1, \sigma_2 | x_1, x_2, \sigma) = \exp\left\{ \frac{-\sum(x_i - \mu)^2}{2\sigma^2} \right\} d\mu_1 d\mu_2 d\sigma_1 d\sigma_2
\]

Integrating out \( \sigma_1, \sigma_2 \) we can get a joint distribution for \( t \).

Consider either term of the form

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

Since \( t^2 = \frac{(\bar{x} - \mu)^2 \sigma^2}{n} \) we have

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

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

so \(-\frac{d\mu}{2} \int_0^\infty e^{-ux} dx = \frac{d\mu}{2} \Gamma(\frac{n}{2}) = -\frac{s}{2\sqrt{n}} \Gamma(\frac{n}{2}) dt \frac{n}{2} \)

When integration is carried out for both, \( \sigma_1 \) and \( \sigma_2 \) we get

\[
P(\sigma_1, \sigma_2 | x_1, x_2, \sigma) \propto \frac{dt_1}{(1+\frac{t_1^2}{n-1})^{n/2}} \frac{dt_2}{(1+\frac{t_2^2}{n-2})^{n_2/2}}
\]

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 \(\mu\) is an
unknown parameter.

The definition of a t-variate, however, does not discern
between a conventional frequency test for \(x\) or a fiducial test on \(\mu\).

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 unbiased estimate of its variance.
   \[
   t = \left( \frac{\bar{x} - \mu}{s} \right) / \sqrt{n}
   \]

2. \(t^2\) is the ratio between the square of a variate
   \(\frac{\bar{x} - \mu}{\sqrt{n}}\) which is \(N(0, 1)\) and a variate \(\frac{s^2}{\sigma^2}\)
   which is distributed as \(\chi^2_\nu\) with \(\nu\) degrees of
   freedom.

The second property is taken as the definition of \(t\).

A random sample \(x_i, i = 1, \ldots, n\), is taken from a population
with mean, \(\mu\), and variance, \(\sigma^2\). \(\bar{x}\) and \(s^2\) are jointly sufficient
statistics obtained from the sample for \(\mu\) and \(\sigma^2\). Define
\[
s' = s / \sqrt{n}
\]

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

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 \( \mathbf{b} = (b_1, b_2, \ldots, b_p) \) can be written \( Y = \mathbf{X} \mathbf{b} + \epsilon \) where \( Y \) is a Tx1 vector of observations, \( \mathbf{X} \) is a Txp matrix of fixed elements with rank \( p \), and \( \epsilon \) is a Tx1 vector of random errors. We assume that the elements of \( \epsilon \) are normally and independently distributed, each with mean zero and unknown variance \( \sigma^2 \). Under these assumptions the likelihood function is

\[
L(b, \sigma^2|y) = \frac{1}{(2\pi)^{T/2} \sigma^T} \exp \left\{ -\frac{(y-\mathbf{X}b)'(y-\mathbf{X}b)}{2 \sigma^2} \right\}
\]

Using Bayes' Theorem, the likelihood function is combined with a prior distribution \( p(b, \sigma^2) \) to obtain a joint posterior distribution \( p(b, \sigma^2|y) \) for the parameters \( b \) and \( \sigma^2 \).

\[
p(b, \sigma^2|y) = \frac{L(b, \sigma^2|y) p(b, \sigma^2)}{\int_{\mathbb{R}} L(b, \sigma^2|y) p(b, \sigma^2) \, db \, d\sigma^2}
\]

In situations where little is known about \( b \) and \( \sigma^2 \), Jeffreys (1961) and Savage (1962) suggested that the prior distributions of \( b \) and \( \log \sigma^2 \) should be taken as locally independent and uniform. That is

\[
p(b) \propto K_1 \quad p(\log \sigma^2) \propto K_2 \quad p(\sigma) \propto 1/\sigma
\]

Thus

\[
p(b, \sigma^2|y) \propto \frac{1}{2\sigma^2} \exp \left\{ -\frac{1}{2\sigma^2} [(\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{y}) - (\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{y})] \right\}
\]

The marginal posterior distribution of \( b \) is obtained by integrating the
joint posterior density function over $\sigma$

$$p(b,\gamma) \propto \left\{ 1 + \left[ b - (x'x)^{-1}(x'y) \right] (x'x) \left[ b - (x'x)^{-1}(x'y) \right] \right\}^{-\frac{\tau}{2}}$$

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 $p(b,\gamma | y_1)$ which in turn serves as the prior to obtain the posterior $p(b,\gamma | y_2)$.

As shown by Tiao and Zellner (1964) the final marginal distribution of $b$ depends on the relationship between $\sigma_1$ and $\sigma_2$. We are interested in the situation where $\sigma_1$ and $\sigma_2$ 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 $\sigma_1 = \kappa \sigma_2$ 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

$$p(b,\sigma | y_1, y_2) \propto \sigma^{-(\tau_1 + \tau_2 + 2)} \exp \left\{ \frac{\tau_1}{\sigma^2} \left[ (\tau_1 - p) s_1^2 + (\tau_2 - p) s_2^2 + [b - (x'y)(x'y)](x'x)[b - (x'y)(x'y)] \right] \right\}$$

where $x'x = x'_i x'_i + x'_z x'_z$ and $x'y = x'_i y_i + x'_z y_z$ and the marginal distribution of $b$ is

$$p(b | y_1, y_2) \propto \left\{ 1 + \frac{b - (x'y)(x'y)}{(\tau_1 - p) s_1^2 + (\tau_2 - p) s_2^2} \right\}^{-\frac{\tau_1 + \tau_2}{2}}$$

which is the same form as that given previously.

**One Variance Known.** Next assume $\sigma_1$ known but $b$ and $\sigma_2$ unknown. Again assuming locally uniform priors, the posterior distribution of $b$ is given by
\[ p(b|y_1, y_2) \propto \exp \left\{ -\frac{1}{2\sigma_2^2} \left[ b - \left( \frac{x'_1 x_2}{\sigma_1^2} \right) \right] \left[ b - \left( \frac{x'_1 x_2}{\sigma_1^2} \right) \right] \right\} \]

\[
\left\{ 1 + \frac{\left[ b - \left( \frac{x'_1 x_2}{\sigma_1^2} \right) \right] \left( x'_2 x_2 \right) \left[ b - \left( \frac{x'_1 x_2}{\sigma_1^2} \right) \right]}{(T_2 - p) \sigma_2^2} \right\}^{-T_2/2}
\]

Since \( \sigma_1 \), 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, \( \sigma_1 \), but integrating over all values of the still random variable \( \sigma_2 \).

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.

\[ B = \left[ \frac{1}{\sigma_1^2} (x'_1 x_1) + \frac{1}{\sigma_2^2} (x'_2 x_2) \right]^{-1} \left[ \frac{1}{\sigma_1^2} (x'_1 y_1) + \frac{1}{\sigma_2^2} (x'_2 y_2) \right] \]

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

**Independent Unknown Variances.** Now assume that \( \sigma_1 \) and \( \sigma_2 \) 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 \( \sigma_1 \) and \( \sigma_2 \). Again assuming locally uniform priors for \( b \), \( \log \sigma_1 \), and \( \log \sigma_2 \), the posterior distribution of \( b \) based on two samples is

\[
p(b|y_1, y_2) \propto \left\{ 1 + \frac{\left[ b - \left( \frac{x'_1 x_2}{\sigma_1^2} \right) \right] \left( x'_2 x_2 \right) \left[ b - \left( \frac{x'_1 x_2}{\sigma_1^2} \right) \right]}{(T_2 - p) \sigma_2^2} \right\}^{-T_2/2}
\]
The expected value of this form can be approximated by an asymptotic solution from Tiao & Zellner (1964) of the same form derived by Theil.

\[
B = \left[ \frac{1}{S_1} (x'_1 x'_1) + \frac{1}{S_2} (x'_2 x'_2) \right]^{-1} \left[ \frac{1}{S_1} (x'_1 y'_1) + \frac{1}{S_2} (x'_2 y'_2) \right]
\]

where, of course, \(S_1\) and \(S_2\) 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 $\beta$ and plug it into set 2 to get $\alpha$.
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

<table>
<thead>
<tr>
<th>Set 1</th>
<th>Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt. (lb.)</td>
<td>Cost ($)</td>
</tr>
<tr>
<td>x₁</td>
<td>y₁</td>
</tr>
<tr>
<td>17.4</td>
<td>9,000</td>
</tr>
<tr>
<td>85.6</td>
<td>26,000</td>
</tr>
<tr>
<td>102.7</td>
<td>32,000</td>
</tr>
<tr>
<td>189.0</td>
<td>29,000</td>
</tr>
<tr>
<td>312.2</td>
<td>27,000</td>
</tr>
<tr>
<td>1,441.4</td>
<td>69,000</td>
</tr>
<tr>
<td>17.4</td>
<td>1,250</td>
</tr>
<tr>
<td>85.6</td>
<td>3,200</td>
</tr>
<tr>
<td>102.7</td>
<td>6,400</td>
</tr>
<tr>
<td>189.0</td>
<td>7,500</td>
</tr>
<tr>
<td>312.2</td>
<td>31,000</td>
</tr>
<tr>
<td>1,441.4</td>
<td>82,500</td>
</tr>
<tr>
<td>17.4</td>
<td>9,364</td>
</tr>
<tr>
<td>85.6</td>
<td>11,612</td>
</tr>
<tr>
<td>102.7</td>
<td>12,460</td>
</tr>
<tr>
<td>189.0</td>
<td>15,517</td>
</tr>
<tr>
<td>312.2</td>
<td>17,923</td>
</tr>
<tr>
<td>1,441.4</td>
<td>38,195</td>
</tr>
</tbody>
</table>

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

\[ y = b x + e \]

with the following computational forms:

\[ b = \left( \begin{array}{c} \hat{\alpha} \\ \hat{\beta} \end{array} \right) \quad (X'X) = \left( \begin{array}{c} \sum x \\ \sum x^2 \end{array} \right) \quad (X'Y) = \left( \begin{array}{c} \sum y \\ \sum xy \end{array} \right) \]

The data reduce to the computational factors,

<table>
<thead>
<tr>
<th>T1</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2</td>
<td>16</td>
</tr>
<tr>
<td>p</td>
<td>2</td>
</tr>
<tr>
<td>\Sigma x</td>
<td>39.3428</td>
</tr>
<tr>
<td>\Sigma x^2</td>
<td>92.1067</td>
</tr>
<tr>
<td>\Sigma y</td>
<td>75.4562</td>
</tr>
<tr>
<td>\Sigma xy</td>
<td>168.4677</td>
</tr>
<tr>
<td>\hat{\alpha}</td>
<td>2.92585</td>
</tr>
<tr>
<td>\hat{\beta}</td>
<td>0.57929</td>
</tr>
<tr>
<td>s_1</td>
<td>0.09162</td>
</tr>
</tbody>
</table>

For method 5 use the Tiao and Zellner (1964) formulation.
Set 1

\[
\begin{align*}
(X_1', X_1) &= \begin{pmatrix} 18.39.342.8 \end{pmatrix} \\
\begin{pmatrix} 39.342.8 \\ 92.106.7 \end{pmatrix}
\end{align*}
\]

\[
(X_1', Y_1) = \begin{pmatrix} 75.456.2 \\ 16.846.77 \end{pmatrix}
\]

\[
(X_1', X_1)^{-1} = \begin{pmatrix} 0.8368 \\ -0.3575 \\
-0.3575 \\ 0.1635 \end{pmatrix}
\]

\[
(X_1', X_1)^{-1}(X_1', Y_1) = \begin{pmatrix} 2.9144 \\ 0.5689 \end{pmatrix}
\]

\[
(T_1 - p) S_1^2 = 1.465.92
\]

Set 2

\[
\begin{align*}
(X_2', X_2) &= \begin{pmatrix} 16 \\ 48.6412 \\
48.6412 \\ 153.6020 \end{pmatrix}
\end{align*}
\]

\[
(X_2', Y_2) = \begin{pmatrix} 74.318.4 \\ 22.987.32 \end{pmatrix}
\]

\[
(X_2', X_2)^{-1} = \begin{pmatrix} 1.6757 \\ -0.5306 \\
-0.5306 \\ 0.1745 \end{pmatrix}
\]

\[
(X_2', X_2)^{-1}(X_2', Y_2) = \begin{pmatrix} 2.5647 \\ 0.6795 \end{pmatrix}
\]

\[
(T_2 - p) S_2^2 = 0.111.24
\]

Marginal Posterior

\[
p(\alpha, \beta | y_1, y_2) \propto \left[ 1 + \frac{(\alpha - 2.9144)^{y_1}}{\beta - 0.5689} \right]^{18} \frac{39.342.8}{92.106.7} \frac{(\alpha - 2.9144)}{0.5689} \left[ 1 + \frac{(\alpha - 2.5647)^{y_2}}{\beta - 0.6795} \right]^{16} \frac{48.6412}{153.6020} \frac{(\alpha - 2.5647)}{0.6795} \left[ 12.1001 \alpha^2 - 100.6202 \alpha + 52.8943 \beta \right]
\]

\[
- 224.6036 \beta + 61.9163 \beta^2 + 211.5123 \right]^{-9}
\]

\[
(13.8997 \alpha^2 - 128.6678 \alpha + 84.4759 \beta \right]^{-8}
\]

\[
- 397.9205 \beta + 133.3814 \beta^2 + 301.1867 \right)\]
The most probable values for the Bayesian estimates of $\alpha$ and $\beta$ were obtained by maximizing the joint posterior distribution $p(\alpha, \beta \mid y_1, y_2)$ with respect to $\alpha$ and $\beta$ jointly by use of an IBM 360 program. Brinton and Garner (1966).

The expected value of method 5 is approximated by

$$
\begin{pmatrix}
\alpha \\
\beta
\end{pmatrix} = \left[ \frac{1}{.09162} \begin{pmatrix} 18 & 39.3428 \\ 39.3428 & 92.1067 \end{pmatrix} + \frac{1}{.0079517} \begin{pmatrix} 16 & 48.6412 \\ 48.6412 & 153.6020 \end{pmatrix} \right]^{-1} 
\begin{pmatrix}
\frac{1}{.09162} (75.4562) \\
\frac{1}{.09162} (168.4680)
\end{pmatrix} + \frac{1}{.0079517} \begin{pmatrix} 74.3194 \\
229.8730
\end{pmatrix}
$$

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

Method 6 involves assuming $S_1 = \sigma_1^{-1}$, 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 $\varphi_1 = S_1$.

Table 3 summarizes the coefficients from the six methods.
Table 3. Regression coefficients

<table>
<thead>
<tr>
<th>Method</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.92585</td>
<td>0.579292</td>
</tr>
<tr>
<td>2</td>
<td>2.55418</td>
<td>0.687720</td>
</tr>
<tr>
<td>3</td>
<td>2.86058</td>
<td>0.596870</td>
</tr>
<tr>
<td>4</td>
<td>2.88380</td>
<td>0.579292</td>
</tr>
</tbody>
</table>
| 5      | 2.83358     | 0.597302   | Most probable value
| 6      | 2.85203     | 0.60192    | Expected value approx.
|        | 2.85203     | 0.60192    |
Literature Survey

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:
- Berkeley 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


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


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


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


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
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

Geisser, Seymour. 1966. Estimation associated with linear discrimi-
nants. 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


Godambe, V. P., and V. Joshi. 1965a. Admissibility and Bayes
estimation in sampling finite populations. Abstract in

Godambe, V. P., and V. Joshi. 1965b. Admissibility and Bayes
estimation in sampling finite populations. I. Annals of

Godambe, V. P. 1966a. The empirical Bayes shortest confidence
intervals for estimating the mean of a finite populations.

Godambe, V. P. 1966b. A new approach to sampling from finite
populations I--sufficiency and linear estimation. Abstract


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.


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


Kraft, Charles H., and Constance V. Eeden. 1964. Bayesian bio-

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.

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.

Statistical Society. 15:30-76.


Lindley, D. V. 1957b. Binomial sampling schemes & concept of


Lindley, D. V., D. A. East, and P. Hamilton. 1960. Table for
making inferences about the variance of a normal distribution.

Lindley, D. V. 1961. The use of prior probability distributions in
statistical inference and decision. 4th Berkeley Symposium
on Mathematical Statistics & Probability Vol. Berkeley,

Ottawa meeting of the Institute of Mathematical Statistics.

Lindley, Dennis V. 1964. The Bayesian analysis of contingency tables.

Lindley, D. V. 1965. Probability and statistics from a Bayesian


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


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


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


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


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


<table>
<thead>
<tr>
<th>Col.</th>
<th>SUBJECT</th>
<th>Col.</th>
<th>SUBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General Considerations</td>
<td>29</td>
<td>Marketing, Advertising, &amp; Spares</td>
</tr>
<tr>
<td>2</td>
<td>Comparison with Classical Methods</td>
<td>30</td>
<td>Medical Research</td>
</tr>
<tr>
<td>3</td>
<td>Comparison with Fiducial Probability</td>
<td>31</td>
<td>Games of Chance and Urn Problems</td>
</tr>
<tr>
<td>4</td>
<td>Prior Distributions</td>
<td>32</td>
<td>Classification, Authorship, Archeology</td>
</tr>
<tr>
<td>5</td>
<td>Estimation and Bayes Estimator</td>
<td>33</td>
<td>Search Problems</td>
</tr>
<tr>
<td>6</td>
<td>Decision Functions</td>
<td>34</td>
<td>Agricultural Research</td>
</tr>
<tr>
<td>7</td>
<td>Confidence Intervals</td>
<td>35</td>
<td>Smoothing &amp; Noise</td>
</tr>
<tr>
<td>8</td>
<td>Sampling</td>
<td>36</td>
<td>Education</td>
</tr>
<tr>
<td>9</td>
<td>Estimation and Bayes Estimator</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Minimax &amp; Game Theory</td>
<td>38</td>
<td>Other</td>
</tr>
<tr>
<td>11</td>
<td>Regression</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Analysis of Variance</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Fisher-Behrens Test</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Robustness</td>
<td>42</td>
<td>Binomial</td>
</tr>
<tr>
<td>15</td>
<td>Invariance</td>
<td>43</td>
<td>Poisson</td>
</tr>
<tr>
<td>16</td>
<td>Opposition to Bayesian Inference</td>
<td>44</td>
<td>Normal</td>
</tr>
<tr>
<td>17</td>
<td>Significance Testing</td>
<td>45</td>
<td>Uniform (Rectangular)</td>
</tr>
<tr>
<td>18</td>
<td>Multivariate Analysis</td>
<td>46</td>
<td>Power (Exponential)</td>
</tr>
<tr>
<td>19</td>
<td>Nuisance Parameters</td>
<td>47</td>
<td>F</td>
</tr>
<tr>
<td>20</td>
<td>Dynamic Programming</td>
<td>48</td>
<td>t</td>
</tr>
<tr>
<td>21</td>
<td>Markov Process</td>
<td>49</td>
<td>Log Normal</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>50</td>
<td>Negative Exponential</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>51</td>
<td>Gamma</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>52</td>
<td>Negative Binomial</td>
</tr>
<tr>
<td>25</td>
<td>APPLICATION</td>
<td>53</td>
<td>Beta</td>
</tr>
<tr>
<td>26</td>
<td>Reliability</td>
<td>54</td>
<td>Inverse</td>
</tr>
<tr>
<td>27</td>
<td>Acceptance Sampling</td>
<td>55</td>
<td>Other</td>
</tr>
<tr>
<td>28</td>
<td>Inventory</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Preventive Maintenance &amp; Spares</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td></td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td></td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td></td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td></td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td></td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td></td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td></td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td></td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td></td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td></td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td></td>
<td>77</td>
<td>Other A-I</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>78</td>
<td>Other J-R</td>
</tr>
<tr>
<td>51</td>
<td></td>
<td>79</td>
<td>Other S-Z</td>
</tr>
<tr>
<td>52</td>
<td></td>
<td>80</td>
<td>Foreign Language</td>
</tr>
</tbody>
</table>

**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 Berkeley Symposium on Math Statistics & Probability
- 65 Institute of Math Statistics Symposium
- 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


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.


CONCLUSIONS

Future Trends

John Tukey, mathematics professor at Princeton and a member of the staff of Bell Telephone Laboratories, 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).
**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 \( \mu \) and known variance \( \sigma^2 \). Let \( f(x_n) \) be an estimate of \( \mu \). 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 $\mu$, $f(x_n+c) - c$. 
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.