SOME RECENT DEVELOPMENTS IN BAYESIAN INFERENCE

by

DEVKI NANDAN KHANNA

B. A., Panjab University, Chandigarh (INDIA), 1957
M. A., University of Delhi, Delhi (INDIA), 1959

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1964

Approved by:

A. M. Feyerherm

Major Professor

TABLE OF CONTENTS

INTRODUCTION

Bayesian inferences or Bayesian Statistics involve an approach to Statistical inference based on the theory of subjective probability. The term 'Bayesian' arises from an elementry theorem of probability theory, named after the Rev. Thomas Bayes, who first enunciated it and proposed its use in inference. Since 1950, many statisticians have taken an active interest in this subject. Hence the term "neo-Bayesian" is sometimes used instead of 'Bayesian'.

BAYESIAN INFERENCES

Bayesian inference involves _a priori_ and _a posteriori_ probability distribution. A distribution which is assessed prior to sample evidence is known as _a priori_ distribution. The term 'posterior' means after the sample evidence.

Suppose before an experiment begins, it can be assumed that p_i is the probability that F_i is the true distribution of X. If an experiment consists of observations on $X_1, \ldots, X_n$, the _a posteriori_ probability that F is equal to F_i can be computed after the sample $x = (x_1, \ldots, x_n)$, has been drawn if the _a priori_ probability distribution is known or can be assumed. The _a posteriori_ probability, denoted by p_{ix}, is the conditional probability that F_i given the observed values $x_1, \ldots, x_n$,. If P is discrete, the _a posteriori_ probability function is given by

$$(1) \qquad p_{ix} = \frac{p_i f(x|F_i)}{\sum_{\Omega} p_i (x|F_i)}$$

and for the continuous case

(2) $$p_{ix} = \frac{p_i f(x|F_i)}{\int_\Omega p_i f(x|F_i)dp}$$

For any element F of Ω, $f(x|F)$ in the above equations, denotes the probability density function of X. The expression for p_{ix} is also known as Bayes' formula.

Consider the problem of making inferences about a Bernoulli process with parameter 'p'. Suppose that no direct sample evidence from the process has been obtained. Based on experience with similar processes, general knowledge etc; one may be willing to translate judgments about the process into probabilistic terms. As such the probability distribution for $\hat{p}$ ('^' indicates that parameter p is considered a random variable) may be considered to be "subjective". Suppose the a priori distribution of $\hat{p}$ is uniform in the interval (0, 1). The probability that $\hat{p}$ lies in a subinterval is that subintervals length, no matter where the subinterval is located between 0 and 1. The probability of observing a sample such as head, head, and tail on three tosses of a coin, given that the probability of observing a head is p, is $p^2(1-p)$. This function is known as a likelihood function. Through use of Bayes' theorem one can obtain the a posteriori distribution of $\hat{p}$ using the likelihood function and the a priori distribution of $\hat{p}$. In terms of inferences about $\hat{p}$, Bayes' theorem is written as

(3) $$f(\hat{p}|r,n) = \frac{f(\hat{p})\ \hat{p}^r\ (1-\hat{p})^{n-r}}{\int_0^1 f(\hat{p})\ \hat{p}^r\ (1-\hat{p})^{n-r}\ d\hat{p}}$$

where $f(\hat{p})$ = a priori density of $\hat{p}$.

$\hat{p}^r(1-\hat{p})^{n-r}$ = likelihood if r heads are observed in n trials.

$f(\hat{p}|r,n)$ = a posteriori density of $\hat{p}$ given the sample evidence.

The integral in the denominator can be regarded as a normalizing factor so that $f(\hat{p}|r,n)$ will be a density function. It is also the probability, in the light of the a priori distribution, of obtaining the sample actually observed.

In the above example

$$f(\hat{p}) = 1; \qquad (0 \leq \hat{p} \leq 1),$$

$r = 2$, $n = 3$, and

$$\int_0^1 f(\hat{p})\hat{p}^r(1-\hat{p})^{n-r}\, d\hat{p} = \int_0^1 \hat{p}^2(1-\hat{p})\, d\hat{p} = 1/12$$

$$\text{so } f(\hat{p}|r = 2, n = 3) = 12\hat{p}^2(1-\hat{p}) \qquad 0 \leq \hat{p} \leq 1$$

$$= 0 \qquad \text{elsewhere}$$

The best Bayesian point estimate can be shown to be the mean of the a posteriori distribution. In this given example, this would be

$$\int_0^1 12\hat{p}.\hat{p}^2(1-\hat{p})\, d\hat{p} = 3/5 = .6$$

It may be noticed that the a posteriori probability that the coin is "biased" in favor of heads is 2/3.

THE LIKELIHOOD PRINCIPLE

The only input needed for a Bayesian analysis are the likelihood function and the a priori distribution. Thus the import of the sample evidence is fully reflected in the likelihood function; a principle that is known as the likelihood principle. If one wants to perform his own Bayesian analysis, he needs the likelihood function. He need not be content with the distribution based on someone else's a priori, nor traditional analysis such as a significance tests, from which it may be difficult or impossible to recover the likelihood function.

PROBABILISTIC PREDICTION

The idea of calculating the probability of a sample in the light of different prior distributions has important consequences. For example, the denominator in the right hand side of the Bayes' formula (3) for Bernoulli sampling can be interpreted as the probability of obtaining the particular sample actually observed, given the a priori distribution of $\hat{p}$. While a person's subjective probability distribution of $\hat{p}$ cannot be said to be "right" or "wrong", there are better and worse subjective distributions, and the decision criterion might be predictive accuracy. Thus if A and B each has a distribution for $\hat{p}$ and a new sample is then observed, one can calculate the probability of the sample in the light of each a priori distribution. The ratio of these probabilities, technically a marginal likelihood ratio, measures the extent to which the data favors A over B or vice-versa.

MULTIVARIATE INFERENCE AND NUISANCE PARAMETERS

Consider inferences about the mean μ of a normal distribution with unknown variance σ^2. In this case, begin with a joint prior distribution for $\hat{\mu}$ and $\hat{\sigma}^2$. The likelihood function is now a function of two variables μ and σ^2. If interest centers only on $\hat{\mu}$, then σ^2 is said to be a nuisance parameter. In principle, it is easy to deal with a nuisance parameter. Simply integrate it out of the a posteriori distribution. This means that one must find the marginal distribution of $\hat{\mu}$ from the joint a posteriori distribution of $\hat{\mu}$ and $\hat{\sigma}^2$. Multivariate problems and nuisance parameters can be dealt with by such an approach.

DESIGN OF EXPERIMENTS AND SURVEYS

In the above discussion, attention centered on the analysis of samples, without concern about the kind and magnitude of sample evidence, that should be obtained. This problem is called the design problem. The Bayesian solution of a design problem requires that one looks beyond the a priori distribution to the ultimate decisions that will be made in the light of this distribution. The question of best design depends on the purposes to be served by collecting the data. Given the specific purpose and the principle of maximization of expected utility, it is possible to calculate the expected utility of the best act for any particular sample outcome. This experiment is repeated for each possible sample outcome for a given sample design. Next, one can weigh all these utilities by the probability of that outcome in the light of the a priori distribution. This gives an overall expected utility for any proposed design. Finally, one picks the sample design with the highest expected utility. Take the case of two-action problems, for example, deciding whether a new medical treatment is better or worse than a standard treatment. This procedure is in no conflict with the traditional approach of selecting designs by comparing operating characteristics, although it formalizes certain things - prior probabilities and utilities - that often are treated intuitively in the traditional approach.

DERIVATION OF THE t-TEST VIA BAYES' THEOREM

As has been noted, according to the Bayesian argument there exists a priori distributions for the mean μ and variance σ^2. Assume that the local a priori distribution of the parameter μ and σ^2 are independent. Also assume that the a priori distribution of μ is locally uniform. Now the Savage (1960) principle of precise measurement says

> . . . that we do not need to know exactly what the a priori distribution of μ is if we can say only that in the region in which the likelihood is appreciable it does not change very much, and at no other point is it of sufficiently great magnitude as to become appreciable when multiplied by the likelihood. This principle would be applicable in situations where the likelihood dominates but is not applicable in situations where the a priori probability density dominates.

The importance of this principle lies in the fact that in actual practice most of the experiments are conducted only when it is expected that the likelihood will exert a much stronger influence in the final result than the a priori distribution. Otherwise, there is little point in doing the experiment. For example, suppose that the value of the gravitational constant in suitable units had been estimated as 32.2 ± 0.1 then there would be little justification for making further measurements with a method whose accuracy was, say, ± 0.2, but considerable justification for conducting further experiments using a method whose accuracy was ± 0.02.

The argument that if μ is taken as locally uniform, then log μ, $\frac{1}{\mu}$ etc; will not be, loses its force if it is remembered that unless the range of values of μ over which the likelihood is appreciable is large compared with the average magnitude of μ over the same range, then such transformations will make little practical difference in the range considered. In the example considered above, for instance, if the a priori distribution of μ were assumed uniform from, say, $\mu = 32.0$ to $\mu = 32.4$, then to a close approximation, the a priori distribution of, for example, log μ and $\frac{1}{\mu}$ would be approximately uniform over corresponding ranges.

Assume also that either σ or its logarithm or some power of σ has a distribution which is locally uniform. Then

(4) $$p_1(\mu) \propto k, \quad p_2(\sigma) \propto \begin{cases} \sigma^{q-1} & \text{if distribution of } \sigma^q \text{ assumed uniform} \\ \sigma^{-1} & \text{if distribution of } \log\sigma \text{ assumed uniform} \end{cases}$$

where k is a constant and "$\propto$" means "proportional to".

Let $\ell(\mu,\sigma|Y)$ denote the likelihood function given the sample Y, then the a posteriori distribution for μ and σ would be

(5) $$p(\mu,\sigma|Y) = k\,\ell(\mu,\sigma|Y)p_1(\mu)\; p_2(\sigma)$$

where $$k^{-1} = \iint_R \ell(\mu,\sigma|Y).\; p_1(\mu)\; .\; p_2(\sigma)\; d\mu d\sigma$$

(6) Now $$p(\mu,\sigma|Y) = p(\mu|\sigma,\bar{y}).\; p(\sigma|s)$$

where $$p(\mu|\sigma,\bar{y}) = \left\{n/(2\pi\sigma^2)\right\}^{\frac{1}{2}} \exp\left\{-(\tfrac{1}{2}\, n/\sigma^2)(\bar{y}-\mu)^2\right\}$$

and $$p(\sigma|s) = 2\left\{\Gamma\left[\tfrac{1}{2}(\nu-q)\right]\right\}^{-1} (\tfrac{1}{2}\,\nu s^2)^{\frac{1}{2}(\nu-q)}\, \sigma^{q-(\nu+1)} \exp\left\{-\tfrac{1}{2}\,\nu s^2/\sigma^2\right\}$$

$(\nu = n-1, \text{ and } q<\nu)$

On integrating out σ one obtains

$$p\left(\frac{\mu-\bar{y}}{s/\sqrt{n}}\,|Y\right) = p\left[t_{\nu-q}\right] \qquad \text{(Box \& Tiao (1962))}$$

where $p\left[t_{\nu-q}\right]$ is the t-distribution with $\nu-q$ degrees of freedom.

In particular, if log σ is assumed to be locally uniform, then the a posteriori distribution of μ is a t-distribution with $\nu = n-1$ d.f. If σ is assumed locally uniformly distributed, then the a posteriori distribution will be a t-distribution with (n-2) d.f. and if σ^2 is locally uniform then one obtains the t-distribution with (n-3) degrees of freedom.

SELECTION OF THE PARENT DISTRIBUTION

Assume that the parent distribution is a member of a class of symmetric distributions which includes, in particular, the normal, together with other distributions on the one hand more leptokurtic, and on the other hand more platykurtic than the normal. A convenient choice is the class of power distributions employed by Diananda (1949), Box (1953), and Turner (1960), where

$$(7) \qquad p(y|\mu,\sigma,\beta) = W \exp\left[-\frac{1}{2}\left|\frac{y-\mu}{\sigma}\right|^{2/(1+\beta)}\right]$$

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

$$(-\infty < y < \infty,\ 0 < \sigma < \infty,\ -\infty < \mu < \infty,\ -1 < \beta < 1)$$

where β denotes a non-normality parameter. In particular, when $\beta=0$, one has the normal distribution; when β is 1, it turns out to be the double exponential; and when $\beta \longrightarrow -1$, the distribution tends to the uniform distribution.

When two samples are drawn from possibly different members of this class, the joint probability density will depend upon six unknown parameters i.e; a set (β_1,μ_1,σ_1) associated with the first sample and a set (β_2,μ_2,σ_2) with the others. It will be assumed throughout the remaining discussion that the parents have the same parameter β, and the ratio $\frac{\sigma_2^2}{\sigma_1^2}$ of the scale parameters is the variance ratio.

DERIVATION OF THE POSTERIOR DISTRIBUTION OF μ FOR A SPECIFIC SYMMETRIC PARENT

Suppose one selects a parent distribution as given above with β assumed to have a fixed value β_o. By doing so, he will adopt the same assumptions *a priori* as are necessary to derive the t-distribution when β is assumed to be zero. One has,

$$\text{(8)} \qquad \ell(\mu,\sigma|Y,\beta_o) = \left[\Gamma\left\{1+\tfrac{1}{2}(1+\beta_o)\right\} 2^{\left\{1+\frac{1}{2}(1+\beta_o)\right\}} \sigma\right]^{-n} \cdot$$

$$\exp\left[-\frac{1}{2}\sum_i \left|\frac{y_i-\mu}{\sigma}\right|^{2/(1+\beta_o)}\right], \; p_1(\mu) \propto k', \; p_2(\sigma) \propto \sigma^{-1}$$

So that

$$\text{(9)} \qquad p(\mu,\sigma|Y,\beta_o) = k\sigma^{-(n+1)} \exp\left\{-\frac{1}{2}\sum_i \left|\frac{y_i-\mu}{\sigma}\right|^{2/(1+\beta_o)}\right\}$$

assuming at least two of the observations are not equal, where

$$k^{-1} = \iint_R \sigma^{-(n+1)} \exp\left\{-\frac{1}{2}\sum_i \left|\frac{y_i-\mu}{\sigma}\right|^{2/(1+\beta_o)}\right\} d\mu d\sigma$$

By integrating out σ, one obtains for the *a posteriori* distribution of μ for any fixed $\beta = \beta_o$ in the permissible range the simple expression

$$\text{(10)} \qquad p(\mu|Y,\beta_o) = k\left[M(\mu)\right]^{-\frac{1}{2}\left[n(\beta_o+1)\right]}$$

where

$$M(\mu) = \left[\sum_i |y_i-\mu|^{2/(1+\beta_o)}\right]$$

and $M(\mu)/n$ is the absolute moment of order $2/(1+\beta_o)$ of the observations about μ . The integral

$$k^{-1} = \int_{-\infty}^{\infty} \left[M(\mu)\right]^{-\frac{1}{2}n(1+\beta_o)} d\mu$$ is merely a normalizing factor which ensures that the total area under the distribution is unity. Usually it is difficult to express it as a simple function but it can be computed easily by use of computers.

Since $p(\mu|Y,\beta_o)$ is a monotonic function of $M(\mu)$, then

(i) $p(\mu|Y,\beta_o)$ is continuous, differentiable and unimodal, although not necessarily symmetric, the mode being attained in the interval $[y_S, y_L]$ where y_L and y_S are respectively the largest and the smallest of the observations.

(ii) When $\beta_o = 0$, $M(\mu) = \sum(y_i-\mu)^2 = (n-1)s^2 + n(\bar{y}-\mu)^2$ and making the necessary substitution in (10), one obtains for the a posteriori distribution of μ

$$p\left(\frac{\mu-\bar{y}}{s/\sqrt{n}}\,\Big|\,Y,\beta_o\right) = p(t_{n-1}) \text{ as obtained earlier.}$$

(iii) When $\beta \longrightarrow -1$, $\lim\limits_{\beta_o \longrightarrow -1} [M(\mu)]^{\frac{1}{2}(\beta_o+1)} = (h + |m - \mu|)$ and making the necessary substitutions

$$(11) \qquad \lim_{\beta_o \longrightarrow -1} p(\mu|Y,\beta_o) = k[h + |m - \mu|]^{-n}$$

where $h = \frac{1}{2}[y_L - y_S]$ and $m = \frac{1}{2}[y_L + y_S]$

$$k^{-1} = \int_{-\infty}^{\infty} (h + |m-\mu|)^{-n} d\mu$$

$$\text{so that} \lim_{\beta_o \to -1} p\left(\frac{|\mu - m|}{h/(n-1)} \,|Y, \beta_o\right) = p(F2,\ 2(n-1))$$

Thus notice that when the parent is normal ($\beta_o = 0$) the expression (10) yields the t-distribution, and when the parent distribution approaches the uniform ($\beta_o \to -1$), the expression (10) gives the double F-distribution with 2 and 2(n-1) d.f. In each of these cases, the a posteriori distribution can be expressed in terms of simple functions of the observations which provide the minimal sufficient statistics for μ and σ. (Box and Tiao (1962))

(iv) In certain other cases, it is possible to express the a posteriori distribution of μ in terms of a fixed number of functions of the observations. For instance, when

$\beta = (1-q)/q$ $(q = 1, 2, 3, \ldots)$, one has

$$(12) \qquad p(\mu,\sigma|Y,\beta_o) \propto \sigma^{-(n+1)} \quad \exp\left\{-\frac{1}{2}\ \sigma^{-2q} \sum_{r=0}^{2q} (-1)^r \binom{2q}{r} \mu^r S_{2q-r}\right\}$$

and

$$(13) \qquad p(\mu|Y,\beta_o) \propto \left[\sum_{r=0}^{2q} (-1)^r \binom{2q}{r} \mu^r S_{2q-r}\right]^{-n/2q}$$

where $S_r = \sum_i y_i^r$ (Box and Tiao (1962))

and it is seen that the set of 2q functions, $S_1, S_2, \ldots S_{2q}$ of the observations are jointly sufficient for μ and σ.

In general, however, the a posteriori distribution cannot be expressed in terms of a few functions of the observations and the minimal sufficient statistics are the observations themselves.

CHOICE OF PRIOR DISTRIBUTIONS FOR μ_1, μ_2, σ_1, σ_2 AND β

As mentioned earlier in (4), assume that the location parameters and the logarithms of the scale parameters are locally uniformly distributed a priori i.e;

(14) $$p_1(\mu_i) \propto k_1$$

(15) $$p_2(\log \sigma_i) \propto k_2 \text{ or } p_2(\sigma_i) \propto \frac{1}{\sigma_i}, \quad i = 1, 2$$

This assumption is appropriate, so long as it is assumed that any point in a region in which the likelihood for μ_1, μ_2, log σ_1 and log σ_2 was appreciable would have been as acceptable a priori as any other. (Assumption used in Savage Principle of Precise Measurement)

Suppose β is a measure of non normality. Choose a priori distribution for β with modal value at $\beta = 0$ and containing an adjustable parameter which controls the degree of concentration about this mode. A convenient choice (Box and Tiao (1962)) is

(16) $$p(\beta) = \frac{\Gamma 2a}{[\Gamma a]^2\, 2^{2a-1}} (1 - \beta^2)^{a-1} \qquad -1 < \beta < 1, \quad a \geq 1$$

When a = 1, this distribution is uniform. This parameter "a" can be adjusted to allow for any desired strength of central limit effect. The case a = 1 giving a uniform distribution for $p(\beta)$ corresponds to no central limit effect. When 'a' tends to infinity, $p(\beta)$ becomes a delta function and represents an overwhelming strong central effect. This corresponds to the assumption of exact normality for the parent distribution.

DERIVATION OF THE POSTERIOR DISTRIBUTION OF THE VARIANCE RATIO $\frac{\sigma_2^2}{\sigma_1^2}$ FOR FIXED VALUES OF μ_1, μ_2 AND β.

From (7), the joint likelihood function of the two samples

$\underline{y}_1 = (y_{11}, y_{12}, \ldots, y_{1n_1})$ and $\underline{y}_2 = (y_{21}, y_{22}, \ldots y_{2n_2})$ is

(17) $$\ell(\sigma_1,\sigma_2,\mu_1,\mu_2,\beta|\underline{y}_1,\underline{y}_2) = k\,\sigma_1^{-n_1}\,\sigma_2^{-n_2}\,.$$

$$\exp\left\{-\frac{1}{2}\sum_{i=1}^{2} n_i s_i(\beta,\mu)\,/\sigma_i^{2/1+\beta}\right\}$$

where $$s_i(\beta,\mu) = \frac{1}{n_i}\sum_{j=1}^{n_i} |y_{ij} - \mu_i|^{2/1+\beta} \qquad i = 1, 2$$

and $$k = \left\{\Gamma\left(1 + \frac{1+\beta}{2}\right) 2^{\left(1 + \frac{1+\beta}{2}\right)}\right\}^{-(n_1+n_2)}$$

Here μ_1, μ_2 are assumed to be known.

The joint posterior distribution of σ_1, σ_2 and β is then

(18) $$p(\sigma_1,\sigma_2,\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)$$

$$= p(\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)\; p(\sigma_1,\sigma_2|\beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)$$

$$= kp(\sigma_1)p(\sigma_2)p(\beta)\;\ell(\sigma_1,\sigma_2,\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)$$

The conditional posterior distribution of σ_1 and σ_2 for given value of β is

(19) $$p(\sigma_1,\sigma_2|\beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2) = \prod_{i=1}^{2} p(\sigma_i|\beta,\mu_i,\underline{y}_i)$$

where $p(\sigma_i|\beta,\mu_i,\underline{y}_i) = k_i\sigma_i^{-(n+1)} \exp\left\{\left[-\frac{1}{2}\ n_i s_i(\beta,\mu)\ /\ \sigma_i\right]^{2/1+\beta)}\right\}$

and $k_i = n_i\left[\dfrac{n_i s_i(\beta,\mu)}{2}\right]^{n_i(1+\beta)/2} \Big/ \ \Gamma\left[1+\dfrac{n_i(1+\beta)}{2}\right]$

which seems to be the product of two inverted gamma distributions. The

a posteriori distribution of $\dfrac{\sigma_2^2}{\sigma_1^2}$ is obtained by making the transformation

$V = \dfrac{\sigma_2^2}{\sigma_1^2}$, $W = \sigma_1$, and integrating out W.

Thus,

(20) $$p(V|\beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2) = kV^{\frac{n_1}{2}-1}\left[1+\frac{n_1 s_1(\beta,\mu)}{n_2 s_2(\beta,\mu)}\ V^{1/1+\beta}\right]^{-(n_1+n_2)\frac{1+\beta}{2}}$$

where $$k = \left(\frac{1}{1+\beta}\right)\ \frac{\Gamma\left[(n_1+n_2)\frac{1+\beta}{2}\right]}{\prod_{i=1}^{2}\Gamma\left[n_i\frac{(1+\beta)}{2}\right]}\ \left[\frac{n_1 s_1(\beta,\mu)}{n_2 s_2(\beta,\mu)}\right]^{\frac{n_1}{2}(1+\beta)}$$

(Box and Tiao (1963))

Now consider the quantity $\dfrac{s_1(\beta,\mu)}{s_2(\beta,\mu)}\ V^{1/1+\beta}$ where $V = \dfrac{\sigma_2^2}{\sigma_1^2}$ is a random variable and $s_1(\beta,\mu)\ /\ s_2(\beta,\mu)$ is a constant calculated from the observations.

After an appropriate transformation

$$(21) \qquad p\left[\frac{s_1(\beta,\mu)}{s_2(\beta,\mu)}\, V^{1/1+\beta} \,\middle|\, \beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2\right] = p\left\{F\left[n_1(1+\beta),\ n_2(1+\beta)\right]\right\}$$

an F-distribution with $n_1(1+\beta)$ and $n_2(1+\beta)$ d.f.

In particular, when $\beta = 0$, the quantity $V \dfrac{\sum^{n_1}(y_{1i}-\mu_1)^2/n_1}{\sum^{n_2}(y_{2i}-\mu_2)^2/n_2}$

is distributed as F with n_1 and n_2 d.f.

Further, when the value of $\beta \longrightarrow -1$ (the parent distributions tend to the rectangular form), the quantity

$$u = V\left\{\frac{\max|y_{1i}-\mu_1|}{\max|y_{2i}-\mu_2|}\right\}^2$$ has the distribution,

$$(22) \qquad \lim_{\beta \longrightarrow -1} p(u|\beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2) = \frac{n_1 n_2}{2(n_1+n_2)}\, u^{\frac{n_1}{2}-1} \qquad \text{for } u \leq 1$$

$$= \frac{n_1 n_2}{2(n_1+n_2)}\, u^{\frac{-n_2}{2}-1} \qquad \text{for } u > 1$$

(Box and Tiao (1963))

Thus, for given β not close to -1, probability levels of V can be obtained from the F-table. In particular, the probability a posteriori that the variance ratio V exceeds unity is

$$\Pr\left\{V > 1 \mid \beta, \mu_1, \mu_2, \underset{\sim}{y}_1, \underset{\sim}{y}_2\right\} = \Pr\left\{F_{n_1(1+\beta),\ n_2(1+\beta)} > \frac{s_1(\beta,\mu)}{s_2(\beta,\mu)}\right\} \tag{23}$$

RELATIONSHIP BETWEEN THE POSTERIOR DISTRIBUTION $p(V \mid \beta, \mu_1, \mu_2, \underset{\sim}{y}_1, \underset{\sim}{y}_2)$ AND CLASSICAL PROCEDURES

From (17), it can be shown that the two power sums $n_1 s_1(\beta,\mu)$ and $n_2 s_2(\beta,\mu)$ when regarded as functions of the random variables $\underset{\sim}{y}_1$ and $\underset{\sim}{y}_2$ have their joint moment generating functions,

$$M_Y(t_1,t_2) = \prod_{i=1}^{2} \left\{1 - 2t_i\sigma_i^{2/1+\beta}\right\}^{\frac{-n_i(1+\beta)}{2}} \tag{24}$$

where $Y = (n_1 s_1(\beta,\mu),\ n_2 s_2(\beta,\mu))$

Thus, letting $Y' = \left(n_1 s_1(\beta,\mu)\Big/\sigma_1^{2/1+\beta},\ n_2 s_2(\beta,\mu)\Big/\sigma_2^{2/1+\beta}\right)$

one obtains

$$M_{Y'}(t_1,t_2) = \prod_{i=1}^{2} (1-2t_i)^{-n_i(1+\beta)/2} \tag{25}$$

(Box and Tiao (1963))

This is a product of the moment generating functions of the independently distributed χ^2 distribution with $n_1(1+\beta)$ and $n_2(1+\beta)$ degrees of freedom respectively. Therefore, $s_1(\beta,\mu)/s_2(\beta,\mu)$ on the hypothesis that $\sigma_1^2/\sigma_2^2 = 1$ is distributed as F with $n_1(1+\beta)$ and $n_2(1+\beta)$ degrees of freedom and in fact provides a uniformly most powerful similar test for this hypothesis against the alternative that $\sigma_1^2/\sigma_2^2 > 1$. The significance level associated with the observed $s_1(\beta,\mu) \,/\, s_2(\beta,\mu)$ is

$$(26) \qquad \Pr\left\{\left[F n_1(1+\beta),\ n_2(1+\beta)\right] > \frac{s_1(\beta,\mu)}{s_2(\beta,\mu)}\right\}$$

and is numerically equal to the probability for $V > 1$ given in equation (23).

A general test derived by Neyman and Pearson and later modified by Bartlett (1937) for comparing k variances for normal populations using the likelihood ratio method is given as follows (This result is due to Bartlett 1937) Let

$$(27) \qquad \lambda(0) = \prod_{i=1}^{k} \left[\frac{N\, s_i(0,\mu)}{\sum_{i=1}^{k} n_i s_i(0,\mu)}\right]^{\frac{n_i}{2}}, \quad N = \sum_{i=1}^{k} n_i ,$$

the quantity $-2 \log \lambda(0) \,/\, g(0)$ is distributed approximately as χ^2 with k degrees of freedom where

$$(28) \qquad g(\beta) = 1 + \left[3k(1+\beta)\right]^{-1} \left[\sum_{i=1}^{k} n_i^{-1} - N^{-1}\right]$$

In general, the likelihood ratio $\lambda(\beta)$ is given by

$$(29) \qquad \lambda(\beta) = \prod_{i=1}^{k} \left[\frac{N\, s_i(\beta,\mu)}{\sum n_i s_i(\beta,\mu)}\right]^{\frac{n_i}{2}(1+\beta)}$$

The quantity $-2 \log \lambda(\beta) / g(\beta)$ is approximately distributed as χ^2 with k degrees of freedom.

THE POSTERIOR DISTRIBUTION OF V WHEN β IS REGARDED AS A VARIABLE PARAMETER

The joint posterior distribution of V and β can be written

$$(30) \qquad p(V,\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2) = p(\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)\; p(V|\beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)$$

where $p(V|\beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)$ is given by equation (20).

The marginal distribution of β can be written as the product

$$(31) \qquad p(\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2) = p(\beta)\,\ell(\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)$$

where $p(\beta)$ is given by equation (16) and

$$\ell(\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2) = k\left[\Gamma\left(1+\frac{1+\beta}{2}\right)\right]^{-(n_1+n_2)} \prod_{i=1}^{2} \Gamma\left[1+n_i\left(\frac{1+\beta}{2}\right)\right]\cdot \left[n_i s_i(\beta,\mu)\right]^{-n_i\frac{1+\beta}{2}}$$

which is the integrated likelihood for β. Thus $p(\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)$ contains information of two kinds i.e; the knowledge a priori about β is characterized by $p(\beta)$ and the information coming from the sample concerning β is represented by $\ell(\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)$

The posterior distribution of V is obtained by integrating out β from equation (30) giving

$$(32) \qquad p(V|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2) = \int_{-1}^{+1} p(\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)\; p(V|\beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)\, d\beta$$

In particular, the probability a posteriori that the variance ratio V exceeds unity is

$$(33) \quad \Pr\left[V>1\,|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2\right] = \int_{-1}^{+1} \Pr\left[V>1\,|\beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2\right]\cdot p(\beta|\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)\, d\beta$$

where the first factor in the integrand is given in (23).

POSTERIOR DISTRIBUTION OF V WHEN μ_1 AND μ_2 ARE REGARDED AS VARIABLE PARAMETERS'

As usual suppose μ_1 and μ_2 are locally uniformly distributed *a priori* as in (14). Upon integrating out these two parameters from the joint posterior distribution of the set (μ_1, μ_2, V, β), one can write the *a posteriori* distribution of β and V as

$$(34) \qquad p(V,\beta|\underline{y}_1,\underline{y}_2) = p(V|\beta,\underline{y}_1,\underline{y}_2)\; p(\beta|\underline{y}_1,\underline{y}_2)$$

The conditional *a posteriori* distribution of V for fixed value of β is given by

$$(35) \quad p(V|\beta,\underline{y}_1,\underline{y}_2) = kV^{\frac{n_1}{2}-1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[n_2 s_2(\beta,\mu) + V^{\frac{1}{1+\beta}} n_1 s_1(\beta,\mu)\right]^{-\frac{n_1+n_2}{2}(1+\beta)} d\mu_1 d\mu_2$$

where

$$k^{-1} = \frac{(1+\beta)}{\Gamma\left[\frac{(n_1+n_2)(1+\beta)}{2}\right]} \prod_{i=1}^{2} \Gamma\left[\frac{n_i}{2}(1+\beta)\right] \int_{-\infty}^{\infty}\left[n_i s_i(\beta,\mu)\right]^{-\frac{n_i}{2}(1+\beta)} d\mu_i$$

and $s_i(\beta,\mu)$, i = 1, 2 are given in (17).

When the parents are normal ($\beta = 0$), the quantity

$$F = V\;\frac{\sum(y_{1i}-\bar{y}_1)^2/(n_1-1)}{\sum(y_{2i}-\bar{y}_2)^2/(n_2-1)}$$

has an F-distribution with (n_1-1) and (n_2-1) degrees of freedom. When the parents tend to the rectangular form ($\beta \longrightarrow -1$) the quantity

$w = V\left(\frac{h_1}{h_2}\right)^2$, where h_1 and h_2 are respectively the ranges of the first and the second sample, has the following limiting distribution

$$(36) \quad \lim_{\beta \to 1} p(w|\beta,\underline{y}_1,\underline{y}_2) = kw^{\frac{n_1-1}{2}-1}\left[(n_1+n_2) - (n_1+n_2-2)\, w^{\frac{1}{2}}\right] \text{ for } w \leq 1$$

$$= kw^{-\frac{n_2-1}{2}}\left[(n_1+n_2) - (n_1+n_2-2)\, w^{-\frac{1}{2}}\right] \text{for } w > 1$$

$$\text{with } k = \frac{n_1 n_2}{2(n_1+n_2)} \quad \frac{(n_1-1)(n_2-1)}{(n_1+n_2-1)(n_1+n_2-2)}$$

COMPUTATIONAL PROCEDURES FOR THE POSTERIOR DISTRIBUTION $p(V|\beta,\underline{y}_1,\underline{y}_2)$

To avoid complexities in evaluating the double integral in (35), the following procedure is adopted. The general expression for the moments of V is obtained in the form

$$(37) \quad E(V^r|\beta,\underline{y}_1,\underline{y}_2) = k \,.$$

$$\frac{\int_{-\infty}^{\infty}\left[n_1 s_1(\beta,\mu)\right]^{-\frac{1}{2}(n_1+2r)(1+\beta)} d\mu_1}{\int_{-\infty}^{\infty}\left[n_1 s_1(\beta,\mu)\right]^{\frac{-n_1}{2}(1+\beta)} d\mu_1} \qquad \frac{\int_{-\infty}^{\infty}\left[n_2 s_2(\beta,\mu)\right]^{-\frac{1}{2}(n_2-2r)(1+\beta)} d\mu_2}{\int_{-\infty}^{\infty}\left[n_2 s_2(\beta,\mu)\right]^{\frac{-n_2}{2}(1+\beta)} d\mu_2}$$

$$\text{with } k = \Gamma\left[\frac{1}{2}(n_1+2r)(1+\beta)\right] \cdot \Gamma\left[\frac{1}{2}(n_2-2r)(1+\beta)\right] \Big/ \prod_{i=1}^{2} \Gamma\left[\frac{n_i(1+\beta)}{2}\right]$$

This expression involves only a simple evaluation of one dimensional integrals.

The more simpler method can be employed as

$$E(V^r|\beta,\underline{y}_1,\underline{y}_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(V^r|\beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2)\ p(\mu_1,\mu_2|\beta,\underline{y}_1,\underline{y}_2)\ d\mu_1 d\mu_2$$

In the integrand, the moments of the conditional a posteriori distribution of V for fixed choices of μ_1 and μ_2 are given by

(38) $$E(V^r|\beta,\mu_1,\mu_2,\underline{y}_1,\underline{y}_2) =$$

$$\frac{\Gamma\left[\frac{1}{2}(n_1+2r)(1+\beta)\right] . \Gamma\left[\frac{1}{2}(n_2-2r)(1+\beta)\right]}{\prod_{i=1}^{2} \Gamma\left[\frac{n_i}{2}(1+\beta)\right]} \left[\frac{n_2 s_2(\beta,\mu)}{n_1 s_1(\beta,\mu)}\right]^{r(1+\beta)}$$

and the joint a posteriori distribution of μ_1 and μ_2 is

(39) $$p(\mu_1,\mu_2|\beta,\underline{y}_1,\underline{y}_2) = \prod_{i=1}^{2} \left[n_i s_i(\beta,\mu)\right]^{-\frac{1}{2} n_i(1+\beta)} \Big/ \int_{-\infty}^{\infty}\left[n_i s_i(\beta,\mu)\right]^{-\frac{1}{2} n_i(1+\beta)} d\mu$$

As shown in the paper by Box and Tiao (1962) that for fixed β, the function

(40) $$f(\mu) = ns(\beta,\mu) = \sum_{i=1}^{n} \left|y_i - \mu\right|^{2/1+\beta}$$

has continuous first derivative and a unique minimum at some point in the interval $[y_S, y_L]$. When $\beta = 0$ or $\beta < -1/3$, it can be shown that $f^3(\mu)$ exists and is continuous, Thus, for these values of β, one can employ Taylor's theorem to expand $f(\mu)$ into $f(\mu) \sim f(\hat{\mu}) + 1/2\ f''(\hat{\mu})\ (\mu-\hat{\mu})^2$ where $\hat{\mu}$ is the point at which $f(\mu)$ attains a minimum. This approximation will be satisfactory when β is not close to -1. From this result, one finds that the moments of V in equation (37) is approximately

(41) $E(V^r|\beta, \underline{y}_1, \underline{y}_2) \sim$

$$\frac{\Gamma\left[\frac{(n_1+2r)(1+\beta)-1}{2}\right]\cdot\Gamma\left[\frac{(n_2-2r)(1+\beta)-1}{2}\right]}{\prod_{i=1}^{2}\Gamma\left[\frac{n_i(1+\beta)-1}{2}\right]}\left\{\frac{n_2 s_2(\beta,\hat{\mu})}{n_1 s_1(\beta,\hat{\mu})}\right\}^{r(1+\beta)}.$$

This implies that, to this degree of approximation, the moments of

$$(42) \qquad C(V) = V^{\frac{1}{1+\beta}} \quad \frac{n_1 s_1(\beta,\hat{\mu}) \Big/ \left[n_1(1+\beta)-1\right]}{n_2 s_2(\beta,\hat{\mu}) \Big/ \left[n_2(1+\beta)-1\right]}$$

are the same as those of an F variable with $n_1(1+\beta)-1$ and $n_2(1+\beta)-1$ degrees of freedom, and hence that the _a posteriori_ distribution of C(V) can be closely approximated using ordinary F-tables. In this approximation, the nuisance parameters μ_1 and μ_2 in the _a posteriori_ distribution of V are eliminated by the very simple process of replacing them by their maximum likelihood values and reducing the degrees of freedom by one unit.

The justification supplied above for this simple approximation is, unfortunately, only valid when $\beta = 0$ and $\beta < 1/3$ but not close to -1. But in actual practice, the approximation has a much wider usefulness.

BAYESIAN ANALYSIS OF THE REGRESSION MODEL

(i) The regression model with the coefficient vector $\beta = (\beta^1, \beta^2, \ldots, \beta^p)$ can be written as

(43) $y = X\beta + \epsilon$

where y is a Tx1 vector of observations, X is a Txp matrix of fixed elements with rank p, and ε is a Tx1 vecotr of random disturbances. Assume that the ε's are NID$(0,\sigma^2)$. Under these assumptions the likelihood function is

$$(44) \qquad \ell(\beta,\sigma|y) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^T \exp\left[\frac{-1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right]$$

Denote the quadratic form in variables β centered at η and with matrix A by $Q(\beta,\eta,A) \equiv (\beta-\eta)' A(\beta-\eta)$

Then the likelihood function can be re-written as

$$(45) \qquad \ell(\beta,\sigma|y) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^T \exp\left[\frac{-1}{2\sigma^2}\left\{\nu s^2 + Q(\beta,\hat{\beta},Z)\right\}\right]$$

where $Z = X'X$, $\hat{\beta} = Z^{-1}X'Y$, $\nu = T-p$ and $s^2 = \frac{1}{\nu}(y-X\hat{\beta})'(y-X\hat{\beta})$

Using Bayes' theorem, the joint posterior distribution can be written as

$$(46) \qquad p(\beta,\sigma|y) = kp(\beta,\sigma)\,\ell(\beta,\sigma|y)$$

where $k^{-1} = \int_R p(\beta,\sigma)\,\ell(\beta,\sigma|y)\,d\beta d\sigma$

and $p(\beta,\sigma)$ is the prior distribution of the parameters β and σ.

When there is nothing known about β and σ, then the <u>a priori</u> distribution of β and $\log\sigma$ could be taken as locally uniform and independent, i.e.

$$(47) \qquad p(\beta) \propto k_1;\ p(\log\sigma) \propto k_2 \quad \text{or } p(\sigma) \propto \frac{1}{\sigma}$$

Combining (45) and (47) in (46), the joint posterior distribution of β and σ is

(48) $$p(\beta,\sigma|y) = \text{const}\ \ \sigma^{-(T+1)}\ \ \exp\left\{\frac{-1}{2\sigma^2}\left[\nu s^2 + Q(\beta,\hat{\beta},Z)\right]\right\}$$

The marginal posterior distribution of β is obtained by integrating the joint posterior density function over σ which gives

(49) $$p(\beta|y) = \text{const}\ \left[1 + \frac{Q(\beta,\hat{\beta},Z)}{\nu s^2}\right]^{-\frac{\nu+p}{2}}$$ (Savage (1962))

This distribution is in the form of a multivariate t-distribution.

In particular, the marginal distribution of a single element β^i can be expressed in terms of a univariate t-distribution with T-p degrees of freedom.

(ii) According to Raiffa and Schlaifer, consider $\sigma_1/\sigma_2 = k$ where k is known. Suppose k = 1, so that $\sigma_1 = \sigma_2 = \sigma$. For this case, they show that

(50) $$p(\beta,\sigma|y_1,y_2) = \text{const}\ \sigma^{-(T_1+T_2+1)}\ \ \exp\left\{-\frac{1}{2\sigma^2}\left[\nu_1 s_1^2 + \nu_2 s_2^2 + Q(\beta,\tilde{\beta},Z)\right]\right\}$$

where $Z = Z_1 + Z_2$, $\tilde{\beta} = Z^{-1}(Z_1\hat{\beta}_1 + Z_2\hat{\beta}_2)$ and the quantities $(\nu_i, s_i^2, \hat{\beta}_i, Z_i)$ i = 1, 2 are defined in connection with (45).

On integrating out σ, the posterior distribution of β is

(51) $$p(\beta|y_1,y_2) = \text{const}\ \left[1 + \frac{Q(\beta,\tilde{\beta},Z)}{\nu s^2}\right]^{-\frac{1}{2}(\nu+p)}$$

with $\nu = T_1+T_2-p$ and $s^2 = \frac{1}{\nu}(\nu_1 s_1^2 + \nu_2 s_2^2)$ which is in the same form as (49)

(iii) Theil (1963) considered the case when σ_1 and σ_2 are different and σ_1 is known. He proceeded within the sampling theory framework to construct the following estimator for β which incorporates information from both samples,

$$(52)\quad \tilde{\tilde{\beta}} = \left(\frac{1}{\sigma_1^2} Z_1 + \frac{1}{s_2^2} Z_2\right)^{-1} \left(\frac{1}{\sigma_1^2} X_1' y_1 + \frac{1}{s_2^2} X_2' y_2\right)$$

The a posteriori distribution of β is given by

$$(53)\quad p(\beta|y_1, y_2) = \text{const}\, \exp\left\{\frac{-1}{2\sigma_1^2} Q(\beta,\hat{\beta}_1,Z_1)\right\}\left\{1 + \frac{Q(\beta,\hat{\beta}_2,Z_2)}{\nu_2 s_2^2}\right\}^{-\frac{\nu_2+p}{2}}$$

where σ_1 is known and β and $\log \sigma_2$ are locally uniform a priori. The expression (53) is the product of two factors, the first is a multivariate normal form and the other a multivariate t-form.

(iv) Suppose σ_1 and σ_2 are independent and unknown. In such cases, with locally uniform a priori distribution for β, $\log \sigma_1$ and $\log \sigma_2$, one finds that the a posteriori distribution of β based upon two samples is given by

$$(54)\quad p(\beta|y_1,y_2) = k^{-1}\left[1+\frac{Q(\beta,\hat{\beta}_1,Z_1)}{\nu_1 s_1^2}\right]^{-\frac{\nu_1+p}{2}}\left[1+\frac{Q(\beta,\hat{\beta}_2,Z_2)}{\nu_2 s_2^2}\right]^{-\frac{\nu_2+p}{2}}$$

$$\text{with } k = \int_R \left[1+\frac{Q(\beta,\hat{\beta}_1,Z_1)}{\nu_1 s_1^2}\right]^{-\frac{\nu_1+p}{2}}\left[1+\frac{Q(\beta,\hat{\beta}_2,Z_2)}{\nu_2 s_2^2}\right]^{-\frac{\nu_2+p}{2}} d\beta$$

This distribution is the product of two multivariate 't' distributions which is known as multivariate "double-t" distribution. The normalizing factor k, here is a p dimensional integral.

The result obtained in (54) is applicable to the problem of making inferences about a population mean when samples are drawn from two normal populations with common mean and unequal variances.

In this case expression (54) reduces to

$$(55)\quad p(\beta|y_1,y_2) = k^{-1}\left[1+\frac{(\nu_1+1)(\beta-\bar{y}_1)^2}{\nu_1 s_1^2}\right]^{-\frac{\nu_1+1}{2}}\left[1+\frac{(\nu_2+1)(\beta-\bar{y}_2)^2}{\nu_2 s_2^2}\right]^{-\frac{\nu_2+1}{2}}$$

$$\text{where } k = \int_{-\infty}^{\infty}\left[1+\frac{(\nu_1+1)(\beta-\bar{y}_1)^2}{\nu_1 s_1^2}\right]^{-\frac{\nu_1+1}{2}}\left[1+\frac{(\nu_2+1)(\beta-\bar{y}_2)^2}{\nu_2 s_2^2}\right]^{-\frac{\nu_2+1}{2}} d\beta$$

and the quantities $\bar{y}_1, \bar{y}_2, s_1^2$ and s_2^2 are respectively, the sample means and the sample variances for the two sets of experiments.

Generalizing the above discussion, suppose that the likelihood function for the ith sample is in the form of (44) with parameters (β,σ_i) and data (y_i, X_i, T_i) . $i = 1, 2, . . . , k$. Then by taking the σ_i's as independent scale parameters, one obtains the following posterior distribution of

$$(56) \qquad p(\beta|y) = w \prod_{i=1}^{k} \left[1 + \frac{Q(\beta,\hat{\beta}_i,Z_i)}{\nu_i s_i^2}\right]^{-\frac{\nu_i+p}{2}}$$

where $w^{-1} = \int_R \prod_{i=1}^{k} \left[1 + \frac{Q(\beta,\hat{\beta}_i,Z_i)}{\nu_i s_i^2}\right]^{-\frac{\nu_i+p}{2}} d\beta$ (Tiao and Zellner (1963))

This distribution is the product of k factors each of which can be expressed as a multivariate 't' distribution.

ASYMPTOTIC EXPRESSION FOR THE MULTIVARIATE "DOUBLE-t" POSTERIOR DISTRIBUTION.

In the previous section, the problem of p dimensional integration is rather laborious and difficult. This can be simplified, by expanding the a posteriori distribution into an asymptotic series in powers of ν_1^{-1} and ν_2^{-1}, and one can reduce the problem of integration to a problem of evaluating the mixed moments of two quadratic forms.

Let $M_1 = \frac{1}{s_1^2} Z_1$, $M_2 = \frac{1}{s_2^2} Z_2$, $Q_1 = Q(\beta,\hat{\beta}_1,M_1)$ and $Q_2 = Q(\beta,\hat{\beta}_2,M_2)$

The expression (54) then becomes

(57) $$p(\beta|y_1,y_2) = k^{-1}\left[1+\frac{Q_1}{\nu_1}\right]^{-\frac{\nu_1+p}{2}}\left[1+\frac{Q_2}{\nu_2}\right]^{-\frac{\nu_2+p}{2}}$$

$$\text{with } k = \int_R \left[1+\frac{Q_1}{\nu_1}\right]^{-\frac{\nu_1+p}{2}}\left[1+\frac{Q_2}{\nu_2}\right]^{-\frac{\nu_2+p}{2}} d\beta$$

The expression $\left[1+\frac{Q_1}{\nu_1}\right]^{-\frac{\nu_1+p}{2}}$ can be written as

$$\left[1+\frac{Q_1}{\nu_1}\right]^{-\frac{\nu_1+p}{2}} = \exp\left[-\frac{1}{2}\ Q_1\right]\ \exp\left[\frac{1}{2}\ Q_1 - \frac{\nu_1+p}{2}\cdot \log\left(1+\frac{Q_1}{\nu_1}\right)\right] .$$

Expand the second factor on the right in powers of ν_1^{-1}

(58) $$\left[1+\frac{Q_1}{\nu_1}\right]^{-\frac{\nu_1+p}{2}} = \exp\left[-\frac{1}{2}\ Q_1\right]\sum_{i=0}^{\infty} p_i\nu_1^{-i}$$

where $p_o = 1$

$$p_1 = \frac{1}{4}\left[Q_1^{\ 2} - 2p\ Q_1\right]$$

$$p_2 = \frac{1}{96}\left[3Q_1^{\ 4} - 4\ (3p+4)\ Q_1^{\ 3} + 12p(p+2)\ Q_1^{\ 2}\right]$$

. . .

Similarly,

$$(59)\quad \left[1+\frac{Q_2}{\nu_2}\right]^{-\frac{\nu_2+p}{2}} = \exp\left[-\frac{1}{2}Q_2\right]\sum_{i=0}^{\infty} q_i\ \nu_2^{-i}$$

where
$$q_0 = 1$$
$$q_1 = \frac{1}{4}\left[Q_2^2 - 2p\ Q_2\right]$$
$$q_2 = \frac{1}{96}\left[3Q_2{}^4 - 4(3p+4)\ Q_2{}^3 + 12p(p+2)\ Q_2{}^2\right]$$
$$\cdots$$

Substitute (58) and (59) into (57) and after simplifying obtain

$$(60)\quad p(\beta|y_1,y_2) = \frac{1}{w}\ \frac{|D|^{1/2}}{(2\pi)^{p/2}}\ \exp\left[-\frac{1}{2}\ Q(\beta,\bar{\beta},\ D)\right]\sum_{i=0}^{\infty}\ \sum_{j=0}^{\infty} p_i q_j \nu_1{}^{-i}\nu_2{}^{-j}$$

where $D = M_1 + M_2$, $\bar{\beta} = D^{-1}(M_1\hat{\beta}_1+M_2\hat{\beta}_2)$ and

$$(61)\quad w = \int_R \frac{|D|^{1/2}}{(2\pi)^{p/2}}\quad \exp\left[-\frac{1}{2}Q(\beta,\ \bar{\beta},D)\right]\sum_{i=0}^{\infty}\ \sum_{j=0}^{\infty} p_i q_j \nu_1{}^{-i}\nu_2{}^{-j}\quad d\beta$$

(Tiao & Zellner (1963))

This integral can be integrated term by term. It appears from the expression (58) and (59), that each term is, in fact, a bivariate polynomial in the mixed moments of the quadratic forms $Q_1 = Q(\beta,\hat{\beta}_1,M_1)$ and $Q_2 = Q(\beta,\hat{\beta}_2,M_2)$ where the variables β have a multivariate normal distribution with mean $\bar{\beta}$ and covariance matrix D^{-1} .

Another simpler method for obtaining mixed moments is first to find the mixed cumulants.

The cumulating generating function of Q_1 and Q_2 is

$$(62) \quad k(t_1,t_2) = \log \int_R \frac{|D|^{1/2}}{(2\pi)^{p/2}} \exp\left[t_1Q_1+t_2Q_2-\frac{1}{2}Q(\beta,\bar{\beta},D)\right] d\beta$$

$$= -\frac{1}{2}\log\left|I-2D^{-1}(t_1M_1+t_2M_2)\right| + t_1\eta_1{}'M_1\eta_1+t_2\eta_2{}'M_2\eta_2$$

$$+ 2(t_1M_1\eta_1+t_2M_2\eta_2)'\ (D-2t_1M_1-2t_2M_2)^{-1}(t_1M_1\eta_1+t_2M_2\eta_2)$$

where $\eta_1 = \bar{\beta} - \hat{\beta}_1$ and $\eta_2 = \bar{\beta} - \hat{\beta}_2$

Upon differentiating (62), one can find the various cumulants. The general form of which is given below

$$(63) \quad k_{rs} = 2^{r+s-1}(r+s-2)!\left[(r+s-1)\ \mathrm{tr.}\ D^{-1}\ G^{rs} + (r\eta_1+s\eta_2)'\cdot\right.$$

$$\left. G^{rs}(r\eta_1+s\eta_2) - r\eta_1{}'\ G^{rs}\ \eta_1 - s\eta_2{}'\ G^{rs}\ \eta_2\right] \qquad r+s \geq 2$$

where $G^{rs} = D(D^{-1}M_1)^r\ (D^{-1}M_2)^s$ (Tiao and Zellner (1963))

Employing the bivariate moment-cumulant formula as given by Cook (1951), the integral in (61) can be written as

$$(64)\quad W = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} b_{ij}\, \nu_1^{-i}\, \nu_2^{-j}$$

where $b_{00} = 1$

$$b_{10} = \frac{1}{4}\left[k_{20} + k_{10}^{\;2} - 2pk_{10}\right]$$

$$b_{01} = \frac{1}{4}\left[k_{02} + k_{01}^{\;2} - 2pk_{01}\right]$$

. . .

Substituting the results in (64) into (60), one obtains

$$(65)\quad p(\beta|y_1,y_2) = \frac{|D|^{1/2}}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}\, Q(\beta,\bar{\beta},D)\right] \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} d_{ij}\, \nu_1^{-i}\, \nu_2^{-j}$$

where

$$d_{00} = 1 \qquad d_{11} = (p_1 - b_{10})(q_1 - b_{01}) + b_{10}\, b_{01} - b_{11}$$

$$d_{10} = p_1 - b_{10} \qquad d_{20} = p_2 - b_{20} - p_1\, b_{10} + b_{10}^{\;2}$$

$$d_{01} = q_1 - b_{01} \qquad d_{02} = q_2 - b_{02} - q_1\, b_{01} + b_{01}^{\;2}$$

The posterior distribution is expressed in the form of a multivariate normal distribution multiplied by a power series in ν_1^{-1} and ν_2^{-1} .

When ν_1 and $_2 \longrightarrow \infty$, all terms of the power series except the leading one vanish so that, in the limit, the posterior distribution is multivariate normal with mean $\bar{\beta}$ and covariance matrix D^{-1} . For finite values of ν_1 and ν_2, the terms in the power series can be regarded as "corrections" in a normal approximation to the multivariate "double-t" distribution.

THE MARGINAL POSTERIOR DISTRIBUTION

When interest centers on a subset of the elements of β, say $\beta_{(\ell)} = (\beta^1, \ldots, \beta^\ell)$, an asymptotic expression for the corresponding marginal posterior distribution can be obtained by integrating out the remaining elements,

$\beta_{(m)} = (\beta^{\ell+1}, \ldots, \beta^p)$ from the joint distribution in (65).

(66) One obtains $p(\beta_{(\ell)}|y_1,y_2) = \dfrac{|D|^{1/2}}{(2\pi)^{p/2}}$.

$$\int_{R'} \exp\left[-\frac{1}{2}\ Q\ (\beta,\ \bar{\beta}\ ,\ D)\right] \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} d_{ij}\ \nu_1^{-1}\nu_2^{-j}\ d\beta_{(m)}$$

Denoting $\bar{\beta} = (\bar{\beta}_{(\ell)} \vdots \bar{\beta}_{(m)})$ and partitioning the matrices D and D^{-1} into

$$D = \begin{array}{c} \begin{array}{cc} \ell & p-\ell \end{array} \\ \left[\begin{array}{c|c} D_{\ell\ell} & D_{\ell m} \\ \hline D_{m\ell} & D_{mm} \end{array}\right] \end{array}, \quad D^{-1} = \begin{array}{c} \begin{array}{cc} \ell & p-\ell \end{array} \\ \left[\begin{array}{c|c} V_{\ell\ell} & V_{\ell m} \\ \hline V_{m\ell} & V_{mm} \end{array}\right] \end{array}$$

One can write the marginal posterior distribution as

(67) $$P(\beta_{(\ell)}|y_1,y_2) = \frac{\left|V_{\ell\ell}^{\ -1}\right|^{1/2}}{(2\pi)^{\ell/2}} \exp\left[-\frac{1}{2}\ Q(\beta_{(\ell)},\bar{\beta}_{(\ell)},V_{\ell\ell}^{\ -1})\right] f(\beta_{(\ell)}|y_1,y_2)$$

where $f(\beta_{(\ell)}|y_1,y_2) = \dfrac{|D_{mm}|^{1/2}}{(2\pi)^{(p-\ell)/2}} \displaystyle\int_{R'}$

$$\exp\left[-\frac{1}{2}\,Q\ (\beta_{(m)},\theta,D_{mm})\right] \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} d_{ij}\nu_1^{-i}\nu_2^{-j}\ d\beta_{(m)}$$

with $\theta = \bar{\beta}_{(m)} - D_{mm}^{\ -1}\ D_{m\ell}\ (\beta_{(\ell)} - \bar{\beta}_{(\ell)})$

From the expression for d_{ij} given in (65), each term in the integral $f(\beta_{(\ell)}|y_1,y_2)$ is a bivariate polynomial in the quadratic forms $Q(\beta,\hat{\beta}_1,M_1)$ and $Q(\beta,\hat{\beta}_2,M_2)$ where $\beta_{(\ell)}$ is considered fixed and $\beta_{(m)}$ has a multivariate normal distribution with mean θ and covariance matrix D_{mm}^{-1}. Adopting the same procedure as done in the previous section

$$\hat{\beta}_1 = (\hat{\beta}_{1(\ell)} \vdots \hat{\beta}_{1(m)}),\quad \hat{\beta}_2 = (\hat{\beta}_{2(\ell)} \vdots \hat{\beta}_{2(m)})$$

$$M_1 = \begin{array}{c} \begin{array}{cc} \ell & p-\ell \end{array} \\ \left[\begin{array}{c|c} B_{\ell\ell} & B_{\ell m} \\ \hline B_{m\ell} & B_{mm} \end{array}\right] \end{array}, \qquad M_1^{-1} = \begin{array}{c} \begin{array}{cc} \ell & p-\ell \end{array} \\ \left[\begin{array}{c|c} E_{\ell\ell} & E_{\ell m} \\ \hline E_{m\ell} & E_{mm} \end{array}\right] \end{array}$$

$$M_2 = \begin{array}{c} \begin{array}{cc} \ell & p-\ell \end{array} \\ \left[\begin{array}{c|c} C_{\ell\ell} & C_{\ell m} \\ \hline C_{m\ell} & C_{mm} \end{array}\right] \end{array}, \qquad M_2^{-1} = \begin{array}{c} \begin{array}{cc} \ell & p-\ell \end{array} \\ \left[\begin{array}{c|c} F_{\ell\ell} & F_{\ell m} \\ \hline F_{m\ell} & F_{mm} \end{array}\right] \end{array}$$

The general form for the mixed cumulants of $Q(\beta,\hat{\beta}_1,M_1)$ and $Q(\beta,\hat{\beta}_2,M_2)$ is given below

$$(68)\quad W_{rs} = 2^{r+s-1}(r+s-2)!\left[(r+s-1)\,\mathrm{tr}\; D_{mm}^{-1} H^{rs} + (r\gamma_1+s\gamma_2)' H^{rs} (r\gamma_1+s\gamma_2) - r\gamma_1' H^{rs}\gamma_1 - s\gamma_2' H^{rs}\gamma_2\right] \quad r+s \geq 2$$

where $H^{rs} = D_{mm}\left[D_{mm}^{-1} B_{mm}\right]^r \left[D_{mm}^{-1} C_{mm}\right]^s$

$$\gamma_1 = \theta - \hat{\beta}_{1(m)} + B_{mm}^{-1} B_{m\ell}\left[\beta_{(\ell)} - \hat{\beta}_{1(\ell)}\right]$$

$$\gamma_2 = \theta - \hat{\beta}_{2(m)} + C_{mm}^{-1} C_{m\ell}\left[\beta_{(\ell)} - \hat{\beta}_{2(\ell)}\right]$$

Using the result in (68), the marginal posterior distribution of $\beta(\ell)$ can be expressed as

$$(69)\quad p(\beta_{(\ell)}|y_1,y_2) = \frac{|V_{\ell\ell}|^{-1}}{(2\pi)^{\ell/2}} \exp\left[-\frac{1}{2} Q(\beta_{(\ell)},\bar{\beta}_{(\ell)},V_{\ell\ell}^{-1})\right] \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \delta_{ij} \nu_1^{-i} \nu_2^{-j}$$

where $\delta_{00} = 1$ $\qquad \delta_{10} = g_{10} - b_{10}$

$\delta_{01} = g_{01} - b_{01}$, $\qquad \delta_{11} = g_{11} - b_{11} - g_{10}b_{01} - g_{01}b_{10} + 2b_{01}b_{10}$

$\delta_{20} = g_{20} - b_{20} - g_{10}b_{10} + b_{10}^2$ $\qquad \cdots$

$\delta_{02} = g_{02} - b_{02} - g_{01}b_{01} + b_{01}^2$

The quantities g_{ij} are functions of the mixed cumulants W_{ij} with functional relationships exactly the same as those between b_{ij} and k_{ij} shown in (63). It should be noted that when $\beta_{(\ell)}$ consist of one variable, the quantity δ_{ij} in (67) are simply polynomials in that variable. Employing the well known expression for the moments of a normal variable one can easily derive an asymptotic expression for moments.

This finishes our discussion on application of Bayes' Theorem in Regression Analysis. For an illustrative example, the reader may refer to the paper on 'Bayes' Theorem and the use of Prior Knowledge in Regression Analysis' by George C. Tiao and Arnold Zellner, University of Wisconsin.

SUMMARY

The use of Bayes' Theorem in Statistical inferences has recently been reconsidered in the works of Jeffreys, Savage, Box and Tiao and others. One advantage of a Bayesian approach is that prior knowledge about parameters of interest can be combined in a well-defined mathematical way with information obtained from experiments. Such prior knowledge may come either from some general theoretical considerations or from the results of previous experience.

Through the use of Bayes' theorem one can obtain the posterior distribution of a certain parameter on the basis of a likelihood function and the prior distribution of that parameter. It has been shown that the best Bayesian point estimate is the mean of the posterior distribution. The Bayesian solution of a design problem requires that one looks beyond the prior distribution to the ultimate decisions that will be made in the light of this distribution.

After assuming the form of our parent distribution, which is not necessarily considered to be normal, but only a member of a class of symmetric distributions which includes normal, one can derive a criterion which is appropriate on this assumption. For example, on the assumption of normality, for the comparison of two means one would derive the t-statistic. It seems natural to justify the use of such a normal theory criterion in the practical circumstances in which normality cannot be guaranteed. These situations lead one to adopt Bayes' method for solution of such problems where normality cannot be assumed.

ACKNOWLEDGMENTS

The author wishes to take this opportunity to express his sincere thanks to his major professor, Dr. Arlin M. Feyerherm, for his interest and skillful guidance not only throughout the writing of this report but also throughout this portion of the writer's graduate study. Appreciation is also expressed to Dr. Leslie F. Marcus for his valuable criticisms in the preparation of this report.

The writer would also like to express his appreciation to Marie Jirak, the Secretary of the Department of Statistics for the many hours she spent in typing the difficult material of this report.

REFERENCES

Bartlett, M. S.
"Properties of sufficiency and statistical tests." Proc. Roy. Soc., 1937.

Box, G. E. P.
"A note on regions for test of kurtosis." Biometrika, 1953, vol. 40.

Box, G. E. P. and Tiao, G. C.
"A further look at robustness via Bayes' theorem." Biometrika, 1962, vol. 49.

Box, G. E. P. and Tiao, G. C.
"A Bayesian approach to the importance of assumptions applied to the comparison of variances." Technical report no. 16, June, 1963.

Cook, M. B.
"Bivariate k-statistics and cumulants of their joint sampling distribution." Biometrika, 1951, vol. 38.

Diananda, P. H.
"Note on some properties of maximum likelihood estimates." Proc. Camb. Phil. Soc., 1949, vol. 45.

Neyman, J. and Pearson, E. S.
"On the problem of k-samples." Bull. Acad. Polonaise Sci. Lett., 1931.

Raiffa, H. and Schlaifer, R.
Applied Statistical Decision Theory. Harvard University, 1961.

Savage, L. J.
"Subjective probability and statistical practice." Technical note 59-1161, Air Force office of Scientific Research, 1959.

Savage, L. J.
Bayesian Statistics in Decision and Information Process, Macmillan, New York, 1962.

Theil, H.
"On the use of incomplete prior information in regression analysis." Jour. of the Am. Stat. Asso., 1963, vol. 58.

Tiao, G. C. and Zellner, Arnold
"Bayes' theorem and the use of prior knowledge in regression analysis." University of Wisconsin, Madison, Wisconsin, 1963. Unpublished Paper.

Turner, M. C.
"On heuristic estimation methods." Biometrics, 1960, vol. 16.

SOME RECENT DEVELOPMENTS IN
BAYESIAN INFERENCE

by

DEVKI NANDAN KHANNA

B. A., Panjab University, Chandigarh (INDIA), 1957
M. A., University of Delhi, Delhi (INDIA), 1959

AN ABSTRACT OF
A MASTER'S REPORT

Submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1964

The Bayesian approach emphasizes the fact that given the likelihood function and the prior distribution of a certain parameter one can find the posterior distribution of that parameter by using the Bayes' theorem. One of the advantages is that the condition of normality need not be assumed while deriving certain distributions. By use of Bayes' theorem, it has been shown how prior knowledge can be utilized in conjunction with sample information in making inferences about the parameters of the regression model.

The 't' distribution was derived given the likelihood function and prior distributions for parameters μ and σ . It was also shown that in sampling from a parent distribution which is a member of a class of symmetric distributions, one can find the posterior distribution of the mean μ by integrating out σ from $p(\mu,\sigma|y,\beta)$, for any fixed β .

A two-sample problem was considered where the two samples are drawn from specified populations with location parameters μ_1 and μ_2 and scale parameters σ_1 and σ_2 and a common non-normality parameter β. Assume that μ_1 and μ_2 are known. Let the ratio $V = \frac{\sigma_2^2}{\sigma_1^2}$ and β correspond to the nuisance parameter in our general formulation. One can then study $p(\frac{\sigma_2^2}{\sigma_1^2} | \beta_1\underline{y})$, the conditional posterior distribution of the squared scale parameter ratio, for any chosen degree of non-normality together with the associated $p(\beta|\underline{y})$ which indicates the plausibility of that value. The posterior density $p(\beta|\underline{y})$ can be written as the product $\ell(\beta|\underline{y})\ p(\beta)$ whose elements are associated with (i) the information concerning non-normality coming from the data and (ii) that injected _a priori_.

Further, if one removes the assumption that μ_1 and μ_2 are known, then the problem involves two laborious integrations. But a close approximation to the integrand can be obtained by replacing the unknown μ_1 and μ_2 by their maximum likelihood estimators in the integrand and changing the degrees of freedom by one unit.

In the case of a regression model, attention has been directed at developing procedures for using information from one sample as prior knowledge in the analysis of a subsequent sample. It is assumed that the two samples drawn from the population have unequal variances. Given a regression model with specified coefficient vector β, one can write the likelihood function which can again be utilized for use of Bayes' theorem in the development of _a posteriori_ distribution of β. Suppose that the likelihood function for the ith sample is in the form

$$\ell(\beta,\sigma|y) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^T \exp\left[\frac{-1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right]$$

with parameters (β,σ_i) and data (y_i,X_i,T_i), $i = 1, \ldots, k$. By taking σ_i's as a independent scale parameters, one can find the posterior distribution of β as a product of k factors which can be expressed as a multivariate 't' distribution.

In the above case, the problem of p dimensional integration is laborious and difficult. This can be remedied by expanding the posterior distribution into an asymptotic series, and thus reducing the problem of integration into a problem of evaluating the mixed moments of two quadratic forms.