DEPARTMENT OF STATISTICS
University of Wisconsin
Madison, Wisconsin

Technical Report No. 56
October, 1965.

BAYESIAN APPROACHES TO SOME BOTHERSOME PROBLEMS
IN DATA ANALYSIS

by

George E. P. Box


Partially supported by Office of Naval Research under Contract ONR 1202(17), Project No. NR 042 222.
BAYESIAN APPROACHES TO SOME BOTHERSOME PROBLEMS IN DATA ANALYSIS

George E. P. Box
Department of Statistics
University of Wisconsin

1. INTRODUCTION

The data analyst often has statistical problems which do not coincide with those described in standard texts. For example, rather than assume normality he may wish to consider a wider class of error distributions. He may wish to compare means in data which arise in the form of time series and he would like to use a model where the assumption of independence between successive observations (or perhaps even the assumption of stationarity) is not made. The fact that satisfactory solutions to such problems have not been available is largely the result of the limitations of classical Neyman-Pearson and Fisherian theories. These theories work well only in cases where sufficient statistics exist and restrict us to making references about unknown parameters in terms of the sampling distributions of statistics. By this route we can claim to know what the data has to tell us about the parameters only if the problem happens to be one in which all aspects of the data are irrelevant except those features of it which are condensed in the statistics. Even when these conditions are met the probabilities we calculate are essentially those with which different sets of data (other than that which has actually happened) could occur for some fixed values of the parameters.

Many people feel that it is much more natural to base inferences on probabilities associated with different values of the parameters leading to a fixed set of data. After all we do have only one set of data and there are a variety of possible values for the parameters. Once we become involved in
discussing the probabilities of sets of data which have not actually occurred we have to decide which "reference set" of groups of data which have not actually occurred we are going to contemplate and this can lead to serious difficulties (see, for example, Barnard [1]).

The association of probabilities with different values of the parameters can be done by the use of Bayes' Theorem. This says that the probability density distribution for parameters \( \theta \) after the data \( \mathbf{y} \) has been collected (i.e., the posterior density \( p(\theta | \mathbf{y}) \)) is proportional to the product of the probability density for \( \theta \) before the data was collected (i.e., the prior density \( p_0(\theta) \)) and the probability density \( p(\mathbf{y} | \theta) \) for the sample (numerically equal to the likelihood \( likelihood \ 2(\theta | \mathbf{y}) \)). Thus

\[
p(\theta | \mathbf{y}) = k p_0(\theta) \times p(\mathbf{y} | \theta)
\]

where \( k \) is to be chosen so that the integral of the probability density with respect to \( \theta \) is unity.

Advocates of the theorem have sometimes urged its use "because it allows evidence not coming from the data to be incorporated." On the other hand when one tries to incorporate such evidence one meets head on the age old problem that "the appropriate prior distribution may not be precisely known" and the criticism that inferences from Bayes' Theorem are arbitrary to the extent that the prior distribution is arbitrary.

Three situations can be distinguished

(a) Prior probabilities important and known exactly. Examples of this kind are rare but do occur. One such arising from a genetic problem is discussed in some detail by R.A.Fisher (2).
The two extreme situations are illustrated above. In one case, very little information is supplied by the data through the likelihood and this scarcely modifies at all the sharply peaked prior distribution which exists. In the other situation the prior information is extremely vague and is completely dominated by the likelihood. Following Savage (3) and (4), I shall assume that Bayes' Theorem can be applied to subjective probabilities and shall argue that in the majority of experimental situations we are in the circumstances of the second diagram where the likelihood dominates. Savage refers to this second situation as that in which the principle of "stable estimation" or "precise measurement" applies. The principle says, in effect, that we do not need to know exactly what the prior distribution of \( \theta \) 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 like that illustrated in Figure (b) in which the likelihood dominates and is inapplicable in the situation illustrated in Figure (a) in which the prior probability density dominates. What makes this principle of particular importance is that most actual experimental situations are represented by Figure (b) rather than by Figure (a). The reason for this is that if the situation is really like that in Figure (a), then there is little point in doing the experiment. For instance, suppose that the value of the gravitational constant in suitable units had been estimated as 32.2 \pm .1, then there would be little justification for making further measurements with a method whose accuracy was, say, \pm 2, but considerable justification for conducting further experiments using a method whose accuracy was \pm 0.2.
On classical theory, once having assumed the form of the parent distribution, we can derive a criterion which is appropriate on this assumption. For example, on the assumption of normality, for the comparison of two means we would derive the t statistic. It is then customary to justify the use of such a normal theory criterion in the practical circumstance in which normality cannot be guaranteed by arguing that the distribution of the criterion is but little affected by non-normality of the parent distribution—that is, it 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 normal-theory statistic. It is easy to produce examples in which the distribution of the normal theory criterion is little affected if the parent is assumed to be some distribution other than the normal; and yet, the inference to be drawn when a criterion appropriate to this other distribution is employed is markedly different.

In a recent paper [6] the analysis of Darwin's paired data on the heights of self and cross-fertilized plants quoted by Fisher in "The Design of Experiments" [7] is reconsidered. In this development the parent
distribution is not assumed to be normal, but only a member of a class of symmetric distributions

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

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

\[ \infty < y < \infty \quad 0 < \sigma < \infty \]

\[ \infty < \theta < \infty \quad -1 < \beta < 1 \]

which include the normal, and whose kurtosis is measured by a parameter \( \beta \).

In particular, we see that when \( \beta = 0 \), we have the normal distribution, when \( \beta \) is 1, we have the double exponential; and when \( \beta \) tends to -1, our distribution tends to the uniform distribution.

We are concerned to make inferences about the true difference in height \( \theta \) between cross fertilized and self fertilized plants. Rather remarkably we find that, on the assumption (made throughout this work), that \( \theta \) and \( \log \sigma \) are locally uniform, the posterior distribution of \( \theta \) for any \( \beta_o \) has the remarkably simple form

\[ p(\theta | y, \beta_o) = k[M(\theta)]^{-\frac{n(\beta_o+1)}{2}} \]

where

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

yielding the familiar t distribution when \( \beta_o = 0 \).

The family of distributions for various values of \( \beta_o \) together with the original data is shown in Figure 2. We see that very different inferences about \( \theta \), will be drawn depending upon which value of \( \beta_o \) is assumed. The
chief reason for this wide discrepancy is the fact that in Darwin's data, the
center of the posterior distribution changes markedly as $\beta$ is changed. In
particular, for this sample, the median, mean, and the mid-point are
respectively 24.0, 20.9, 4.0, and these are the modes of the posterior
distributions for the double exponential, normal, and uniform parent respectivel
Because of the wide differences which occur in the posterior distribution
of $\theta$ depending on which parent distribution (that is, which value of $\beta_o$)
we employ, it might be thought there would be considerable uncertainty as to
the nature of the valid inference that could be drawn from this data. When
we use appropriate evidence concerning the value of $\beta$ this turns out not to
be the case. We have two sources of information about the value of $\beta$, one
from the data itself and the other from our knowledge a priori that a central
limit effect would operate in the circumstances of the experiment. Both types
of evidence can be injected into our analysis by allowing $\beta$ to be a variable
parameter associated with a prior distribution.

We can represent the central limit tendency of the errors by choosing
a prior distribution for $\beta$ which has a maximum value at $\beta = 0$, and which
extends from -1 to +1. A convenient distribution for this purpose is the
beta distribution having mean zero and extending from -1 to +1 and, consequently,
possessing only one adjustable parameter which we call $a$. We assume then:

$$P(\beta) = \text{constant} \times (1 - \beta^2)^{a-1} \quad -1 < \beta < 1$$
$$a \geq 1$$

(3)

When $a = 1$, this distribution is uniform. With $a > 1$, it is a symmetric
distribution having its mode at the normal theory value $\beta = 0$. 
From the joint posterior distribution of $\theta$ and $\beta$

\[ p(\theta, \beta / y) = p(\theta / \beta, y) \cdot \phi(\beta) \cdot p(\beta) \]

the posterior distribution of $\theta$ is obtained by integrating out $\beta$ yielding

\[ p(\theta / y) = \int_{-1}^{1} p(\theta, \beta / y) \, d\beta \]
\[ = \int_{-1}^{1} p(\theta / \beta, y) \cdot \phi(\beta) \cdot p(\beta) \, d\beta \]

In obtaining this integral, we are averaging the t-like distributions $p(\theta / \beta, y)$ with a weight function $\phi(\beta) \cdot p(\beta)$ which is, in fact, $p(\beta / y)$, the posterior distribution of $\beta$. The value of this weight function is seen to depend partly upon information from the sample through $\phi(\beta)$ and partly from prior information characterized by $p(\beta)$. The way in which this weight function $\phi(\beta) \cdot p(\beta)$ changes as the assumed central limit effect is increased is shown in Figure 3. In these diagrams, the dotted curve is, in each case, the prior distribution $p(\beta)$. When $a = 1$ (Figure 3(i)) $p(\beta)$ is uniform and $p(\beta / y)$ equals $\phi(\beta)$. This represents the situation where the information concerning $\beta$ is essentially coming from the sample itself. The value of the parameter $a$ is 3 in Figure 3(ii), 6 in Figure 3(iii), and 10 in Figure 3(iv). These three diagrams show how increasing certainty of a central limit effect tends to override the information from the sample. Finally, when $a$ tends to infinity, both $p(\beta)$ and $p(\beta / y)$ would approach a delta function at $\beta = 0$.

The integration

\[ p(\theta / y) = \int_{-1}^{1} (\theta / \beta, y) \phi(\beta) \cdot p(\beta) \, d\beta \]
has been actually carried out for each of these weight functions and the results are shown in Figure 4 together with the $t$ distribution which would be appropriate for the case $a \to \infty$ corresponding to an assumption of exact normality.

Figure 4 then represents the final inference we could draw for $\theta$ depending on how strong a central limit effect would be appropriate in the physical situation. In view of the very large differences exhibited by the $t$-like distributions in Figure 2, it seems remarkable how alike these distributions are. In particular, it will be seen that the tail areas which have been traditionally regarded as the most important part of the distribution are very little affected even with no "central limit effect." The main reason for this is that those widely discrepant $t$-like distributions generated by parents which approach the uniform are almost ruled out by information coming from the sample itself. (See Figure 3(i).)

It would also seem that the calculation of $\phi(\delta)$, i.e., $p(\delta|y)$ for uniform $p(\delta)$ as shown in Figure 3(i) would provide a very satisfactory way of summarizing what the data has to tell us concerning the nature of the parent distribution from which the sample is drawn. It will be noted that in our approach, we have done more than merely "test" the assumption of normality and then, in the absence of "a significant" result, assume it. The information concerning $\delta$ coming from the sample is included in the formulation itself and we have seen in the case of Darwin's data it plays an important role in virtually eliminating the influence of unlikely parent distributions.

A similar approach can be taken [8] to problem of the comparison of variances where the normal assumption is particularly restrictive.

3. A BAYESIAN APPROACH TO TRANSFORMATIONS

In recent work [9] the Bayesian approach has been applied to the problem of choosing a suitable data transformation. Consider the data of Table 1.
Table 1. Survival times (unit, 10 hr.) of animals in a $3 \times 4$ factorial experiment

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>.31</td>
<td>.82</td>
<td>.43</td>
<td>.45</td>
</tr>
<tr>
<td></td>
<td>.45</td>
<td>1.10</td>
<td>.45</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>.46</td>
<td>.88</td>
<td>.63</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td>.43</td>
<td>.72</td>
<td>.76</td>
<td>.62</td>
</tr>
<tr>
<td>Poison</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>.36</td>
<td>.92</td>
<td>.44</td>
<td>.56</td>
</tr>
<tr>
<td></td>
<td>.29</td>
<td>.61</td>
<td>.35</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>.40</td>
<td>.49</td>
<td>.31</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>.23</td>
<td>1.24</td>
<td>.40</td>
<td>.38</td>
</tr>
<tr>
<td>III</td>
<td>.22</td>
<td>.30</td>
<td>.23</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>.21</td>
<td>.37</td>
<td>.25</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>.18</td>
<td>.38</td>
<td>.24</td>
<td>.31</td>
</tr>
<tr>
<td></td>
<td>.23</td>
<td>.29</td>
<td>.22</td>
<td>.33</td>
</tr>
</tbody>
</table>

The usual normal theory analysis will be valid, efficient, and easily comprehensible if in addition to independence of errors, we can assume

(a) the expected value of the variate in any cell can be represented by additive row and column constants, i.e., that no interaction terms are needed,

(b) the error variance is constant,

(c) the observations are normally distributed.

Even a cursory examination of these data suggest that these assumptions will not be true. We suppose instead (less restrictively) that the assumptions may be true after some non-linear data transformation has been applied. Although the method can be applied for any type of non-linear transformation, we shall illustrate it only in the case of the power transformation. We assume that for some $\lambda$ the transformed observation $y^\lambda$ satisfies the above assumptions. If
If we then conduct our analysis in terms of the transformed observations
\[ x = \frac{1}{\lambda} \cdot (y/\bar{y})^\lambda \]
where \( \bar{y} \) is the geometric mean of all the observations then on sensible assumptions we find that the posterior distribution of \( \lambda \) is simply

\[ p(\lambda) = \text{constant} \times p_0(\lambda) \times S(x)^{-\frac{1}{2}} \]

where \( S(x) \) is the residual sum of squares.

Figure 5 shows the posterior distribution of \( \lambda \) for the data of Table 1. We see that no transformation (\( \lambda = 1 \)) and the log transformation (\( \lambda = 0 \)) are unacceptable and a transformation close to the reciprocal is indicated as best.

The remarkable improvement which occurs when this transformation is actually employed is shown in Figure 5. The variance ratio \( F(\lambda) \) for interaction against error and Bartlett’s criterion for equality of cell variances are there plotted as functions of \( \lambda \) with significance levels indicated. The Analysis of Variance in Table 2 shows that in the transformed variate not only do we eliminate all suspicion of an interaction but also we increase the sensitivity of the analysis about threefold. One transformed observation is about as good as three untransformed ones.

<table>
<thead>
<tr>
<th>Table 2.</th>
<th>Biological data. Analyses of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean squares × 1000</td>
<td></td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>Untransformed</td>
</tr>
<tr>
<td>Poisons</td>
<td>2</td>
</tr>
<tr>
<td>Treatments</td>
<td>3</td>
</tr>
<tr>
<td>P × T</td>
<td>6</td>
</tr>
<tr>
<td>Within groups</td>
<td>36</td>
</tr>
</tbody>
</table>
A further interesting feature of the Bayesian analysis is the possibility which it offers of analyzing separately the contributions of Normality (N), Homogeneity of variance (H), and Additivity (A) to the choice of transformation.

The analysis for the biological data is shown in Figure 7. We see that the information about $\lambda$ coming from within group normality is very slight, values of $\lambda$ as far apart as -1 and 2 being acceptable on this basis. The requirement of constant variance, however, has a major effect on the choice of $\lambda$; further, some information is contributed by the requirement of additivity.

Many other examples might be quoted to illustrate the practical value of the Bayesian approach. In particular, at Wisconsin we have found that the approach has illuminated such diverse topics as variance component analysis [10], regression analysis [11], regression with auto-correlated errors [12], estimation of common parameters from several responses [13], sequential design of non-linear experiments [14], sequential discrimination among models [15]. To illustrate the broad scope of this approach I quote a final example [16].

4. A CHANGE IN LEVEL OF A NON-STATIONARY TIME SERIES

Suppose that observations $z_t$ of a time series are available at equally spaced time intervals. We consider the problem of making inferences about a possible shift in level of the series associated with the occurrence of an event $E$ at some particular time. For example, the observations might be of some economic indicator and we might suspect a change in level to occur in a particular interval because of a change in fiscal policy. Alternatively, $z_t$
could be the daily output of a chemical process and the event E might be a change in the supplier of raw material.

Suppose we have available \( n_1 \) observations before the event E and further \( n_2 \) observations afterwards. It is well known that if the observations \( z_1, \ldots, z_{n_1} \) are a random sample from the normal distribution \( N(\mu_1, \sigma^2) \) and \( z_{n_1+1}, \ldots, z_{n_1+n_2} \) a random sample from \( N(\mu_2, \sigma^2) \), then inferences concerning the change in level \( \delta = \mu_2 - \mu_1 \) can be made using the Student-t distribution.

The criterion

\[
\frac{(y_2 - y_1) - \delta}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}
\]

would follow a t-distribution and could be used for testing hypotheses and for obtaining confidence intervals for \( \delta \). The fiducial distribution of \( \delta \) would be a scaled t-distribution. With appropriate prior assumptions this would also be the posterior distribution of \( \delta \). If the observations were not independent but the physical situations "event E" and "no event E" could be applied in random order then the t-distribution would again be appropriate as a close approximation to the null randomization distribution and so could supply a valid basis for inference.

In examples like those quoted above where the observations are almost certainly dependent and no possibility for randomization exists, the usual procedure based on the t-distribution would of course be invalid and could be extremely misleading. Even moderate dependence between observations within a sample can seriously invalidate such procedures.
In practice not only would successive observations usually be dependent, but frequently the time series would be non-stationary. A simple stochastic model which seems to provide adequate representation of a surprisingly large number of such time series arising in economic and industrial applications (Brown 1959, Holt et al. 1960, Box and Jenkins 1962, 1963) is:

\[
Z_p = M + \gamma_o \sum_{j=1}^{\infty} \alpha_{p-j} + \alpha_p \quad 0 \leq \gamma_o < 2
\]

(4)

\(M\) reflects the initial location of the series at the infinite past and the \(\alpha\)'s are independent random normal deviates having variance \(\sigma^2\).

Values of the constant \(\gamma_o\) between zero and one are most frequently found in practice. When necessary the series can be suitably generalized to allow for more complicated situations.

Suppose the parameter \(\delta\) measures the shift in level of the series associated with the event \(E\). Then, for the \(n_1\) available observations before \(E\),

\[
Z_1 = L + \alpha_1
\]

\[
Z_p = L + \gamma_o \sum_{j=1}^{L} \alpha_{p-j} + \alpha_p
\]

(5)

\[= \gamma_o \sum_{j=0}^{p-2} (1 - \gamma_o)^j Z_{p-1-j} + (1 - \gamma_o)^{p-1} L + \alpha_p\]

\[p = 2, \ldots, n_1\]
For the observations after $E$,

$$
  z_p = L + \delta + \gamma_o \sum_{j=1}^{p-1} \alpha_{p-j} + \alpha_p
$$

$$
  = \gamma_o \sum_{j=1}^{p-2} (1-\gamma_o)^j z_{p-1-j} + (1-\gamma_o)^{p-1} L + (1-\gamma_o)^{p-(n_1+1)} \delta + \alpha_p
$$

$$
  p = n_1 + 1, \ldots, n_1 + n_2.
$$

Assuming at first that $\gamma$ is known and that $L$, $\delta$, and $\log \sigma$ have locally independent uniform distributions, $\delta$ is distributed a posteriori so that

$$
  (\delta - \hat{\delta}) \left\{ \frac{[1 - (1-\gamma_o)^{2n_1}][1 - (1-\gamma_o)^{2n_2}]}{[1 - (1-\gamma_o)^{2N}] \gamma_o(2-\gamma_o) s^2} \right\}^{1/2}
$$

exactly follows a Student-$t$ distribution with $(n_1 + n_2 - 2)$ degrees of freedom with

$$
  s^2 = \frac{1}{n_1 + n_2 - 2} \left\{ \sum_{t=1}^{n_1} \left[ y_t - \hat{L}(1-\gamma_o)^{t-1} \right]^2 + \sum_{t=n_1+1}^{N} \left[ y_t - \hat{L}(1-\gamma_o)^{t-1} - \hat{\delta}(1-\gamma_o)^{t-n_1-1} \right]^2 \right\}.
$$

$$
  \hat{\delta} = \frac{\gamma_o}{1-(1-\gamma_o)^{2n_2}} \left[ \sum_{s=1}^{n_2} (1-\gamma_o)^{s-1} z_{n_1+s} + (1-\gamma_o)^{n_2} \sum_{s=1}^{n_2} (1-\gamma_o)^{n_2-s} z_{n_1+s} \right]
$$

$$
  - \frac{\gamma_o}{1-(1-\gamma_o)^{2n_1}} \left[ \sum_{s=1}^{n_1} (1-\gamma_o)^{n_1-s} z_s + (1-\gamma_o)^{n_1} \sum_{s=1}^{n_1} (1-\gamma_o)^{s-1} z_s \right].
$$
and \[ \hat{L} = \frac{\gamma_o}{1 - (1 - \gamma_o)^2} \left[ \sum_{s=1}^{n_1} (1 - \gamma_o)^{s-1} z_s + (1 - \gamma_o) \sum_{s=1}^{n_1} (1 - \gamma_o)^{n_1-s} z_s \right] \] (10)

At first sight, the above expression appears somewhat complicated but, in fact, it has a remarkably simple interpretation as is seen by considering an example. The 50 observations \( z_t \) plotted in Figure 8 were generated from a table of standard random normal deviates using equations (5) and (6) and setting \( L = 5 \), \( \gamma_o = 0.3 \). A change in level \( \delta = 1.5 \) was introduced after the 25th observation.

Suppose at first we know that \( \gamma_o = 0.3 \) and we desire to make inferences about a possible change in level associated with the event \( E \) which occurs immediately after the 25th observation. The weight function applied to the observations \( z_t \) in calculating \( \hat{\delta} \) in which is our estimate of the change in level \( \delta \), is shown in Figure 8 immediately above the series. It is seen that \( \hat{\delta} \) is simply the difference between two exponentially weighted averages, one having maximum weight immediately prior to the event \( E \) and the other having maximum weight immediately after.

Most of the complications which occur in the formulae are to take care of the possibility that the weight functions may not have "died out" before the series begins and ends. The effect of truncation occurring in the weight function is illustrated by showing what happens in the example when only \( n_1 = 7 \) observations are available immediately prior to the event \( E \) and only \( n_2 = 5 \) after it. The weight function appropriate to this situation is shown in Figure 8 immediately below the series. To allow for truncation, the weight function is seen to "double back" on itself. It is the weight added
in this doubling back which is taken care of by the second terms in the square brackets of (9). When the truncation is negligible, these terms can be ignored and we then have simple exponentially weighted averages. In this example a 95% interval for \( \delta \) is (0.236, 2.8574).

To illustrate how our inferences about \( \delta \) may be affected by changes in the value of \( \gamma_o \),

\[
t(\gamma_o) = \sqrt{\frac{[1-(1-\gamma_o)^{2n_1}] [1-(1-\gamma_o)^{2n_2}]}{[1-(1-\gamma_o)^{2n}] \gamma_o (2-\gamma_o) s^2}}
\]

is plotted together with the function \( h(\gamma_o \mid z) \) in Figure 9. This would be the posterior distribution of \( \gamma_o \) if the prior distribution \( p_o(\gamma_o) \) were taken locally uniform. It can thus be used to calculate the posterior distribution of \( \gamma_o \) by combining it with any desired \( p_o(\gamma_o) \). If as in many situations "no information" about \( \gamma_o \) were available a priori, then it could be regarded as indicating what we know about \( \gamma_o \) from the sample itself. We see that over the range in which \( h(\gamma_o \mid z) \) is appreciable the value of \( t(\gamma_o) \) is close to 2 and consequently for plausible values of \( \gamma_o \), 95% of the conditional posterior probability mass of \( \delta \) is over the positive range.

CONCLUSION

I have tried to show with the above examples how Bayesian analysis sets us free from the limitations of the assumptions necessary to make sampling theory work. What he is entitled to infer from an experiment is, in my opinion, best conveyed to the experimenter by the appropriate inspection of the
posterior distribution. With the ready accessibility of electronic computers to make laborious calculations, and the availability of Bayes' Theorem to make general inferences it is now open to the experimenter to generate and analyze data unimpeded by unnecessary restrictions.
REFERENCES


FIGURES 3 : (i) - 3 : (iv)
PRIOR AND POSTERIOR DISTRIBUTIONS OF \( \beta \) FOR VARIOUS CHOICES OF \( \alpha \).

3 : (i)

\( \alpha = 1 \)

\( \phi(\beta) = P(\beta) \)

3 : (ii)

\( \alpha = 3 \)

3 : (iii)

\( \alpha = 6 \)

3 : (iv)

\( \alpha = 10 \)

\( P(\beta/Y) \)

\( P(\beta) \)
Fig. 5