File

DEPARTMENT OF STATISTICS

The University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT #16

June, 1963

A BAYESIAN APPROACH TO THE IMPORTANCE OF ASSUMPTIONS APPLIED TO THE COMPARISON OF VARIANCES

G. E. P. Box and George C. Tiao

This research was supported in part by the Wisconsin Alumni Research Foundation and by the United States Navy through the Office of Naval Research, under Contract Nonr-1202(17), Project NR 042 222.

Also issued as Systems Formulation and Methodology Workshop Paper No. 6306 by the Social Systems Research Institute, University of Wisconsin.

A Bayesian Approach To The Importance of Assumptions Applied To The Comparison of Variances

G.E.P. Box and George C. Tiao

I. Introduction.

Frequently the distribution of observations $\underline{y}$ depends not only upon a set of r parameters $\underline{\xi}_1 = (\xi_1, \ldots, \xi_r)$ of interest, but also on a set of, say, $t - r$ further nuisance parameters $\underline{\xi}_2 = (\xi_{r+1}, \ldots, \xi_t)$. Considerable difficulties may arise in dealing with nuisance parameters by non-Bayesian methods. When the use of Bayes' theorem is appropriate, however, the "overall" inference situation about $\underline{\xi}_1$ is completely exemplified by the posterior distribution of $\underline{\xi}_1$ obtained by simply "integrating out" the nuisance parameter $\underline{\xi}_2$.

Now we can write the joint posterior distribution of $(\underline{\xi}_1, \underline{\xi}_2)$ as the product,

$$p(\underline{\xi}_1, \underline{\xi}_2 | \underline{y}) = p(\underline{\xi}_1 | \underline{\xi}_2, \underline{y}) p(\underline{\xi}_2 | \underline{y}).$$

The posterior distribution of $\underline{\xi}_1$ can thus be written:

$$(1) \qquad p(\underline{\xi}_1 | \underline{y}) = \int_R p(\underline{\xi}_1 | \underline{\xi}_2, \underline{y}) p(\underline{\xi}_2 | \underline{y}) \, d\, \underline{\xi}_2$$

in which the marginal posterior distribution $p(\underline{\xi}_2 | \underline{y})$ of the nuisance parameters acts as a weight function multiplying the conditional distribution $p(\underline{\xi}_1 | \underline{\xi}_2, \underline{y})$ of the parameters of interest. It is frequently helpful in understanding the problem and the nature of the conclusions which can safely be drawn to consider not only $p(\underline{\xi}_1|\underline{y})$ but also the components of the integral on the right hand side of equation (1). One is thus led to consider the conditional distributions of $\underline{\xi}_1$ for particular values of the nuisance parameters $\underline{\xi}_2$ in relation to the probability of occurance of the postulated values of the nuisance parameters.

In particular, in judging the robustness of the inference relative to characteristics such as non-normality and lack of independence between errors the nuisance parameters $\underline{\xi}_2$ can be measures of departure from normality and independence. The distribution of the parameters of interest $\underline{\xi}_1$ conditional on some specific choice $\underline{\xi}_2 = \underline{\xi}_{20}$ will indicate the nature of the inference which we could draw if the corresponding set of assumptions (for example, the assumptions of normality with uncorrelated errors) are made, while the marginal posterior density $p(\underline{\xi}_2 = \underline{\xi}_{20}|\underline{y})$ reflects the plausibility of such assumptions being correct. The marginal distribution $p(\underline{\xi}_1|\underline{y})$ obtained by integrating out $\underline{\xi}_2$ indicates the overall inference which can be made when proper weight is given to the various possible assumptions in the light of the data and their initial plausibility.

The attractiveness of this approach is further increased by the fact that if we let $\pi(\xi_1)$ and $\pi(\underline{\xi}_2)$ be the prior distributions for $\underline{\xi}_1$ and $\underline{\xi}_2$ assumed independent, and let $\ell(\underline{\xi}_1,\underline{\xi}_2|\underline{y})$

represent the joint likelihood, we may then write equation (1) in the form,

$$p(\underline{\xi}_1|\underline{y}) = k_1 \int_R p(\underline{\xi}_1|\underline{\xi}_2,\underline{y})\ell(\underline{\xi}_2|\underline{y})\pi(\underline{\xi}_2)\, d\,\underline{\xi}_2$$

where

$\ell(\underline{\xi}_2|\underline{y}) = k_2 \int \ell(\underline{\xi}_1,\underline{\xi}_2|\underline{y})d\,\underline{\xi}_1$ is the "integrated" likelihood and k_1 and k_2 are normalizing constants. Thus the weight function $p(\underline{\xi}_2|\underline{y})$ which is proportional to the product $\ell(\underline{\xi}_2|\underline{y})\pi(\underline{\xi}_2)$, is separated into a part coming from the data itself (this would tell us what the data "has to say" about the assumptions) and a distinct part coming from the prior distribution. It is informative to consider the effect of varying $\pi(\underline{\xi}_2)$ to see how sensitive the final result is to changes in prior assumptions and also to study $\ell(\underline{\xi}_2|\underline{y})$ itself.

The above is, in fact, the approach which we adopted in our previous study of Darwin's data which concerned the value of a location parameter-- Box and Tiao (1962). In this paper, we further illustrate the attack by applying it to the problem of comparing two scale parameters. Standard tests to compare variances based on normal theory are known to be sensitive to non-normality. For example, Box and Anderson (1954) show that the same degree of kurtosis which changes the significance level of a test to compare twenty means from 5% to 4.9% changes the significance level of a corresponding test to compare twenty variances from 5% to 71.8%.

II. Scope of the Present Study.

We suppose that two samples are drawn from specific populations characterized by location parameters θ_1 and θ_2, scale parameters σ_1 and σ_2 and a common non-normality parameter β. We first assume that the location parameters θ_1 and θ_2 are known. Letting the ratio σ_2^2/σ_1^2 corresponds to the parameter $\underline{\xi}_1$ and β correspond to the nuisance parameter $\underline{\xi}_2$ in our general formulation we can then study $p(\sigma_2^2/\sigma_1^2|\beta,\underline{y})$, 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.

We then relax the assumption that the location parameters θ_1 and θ_2 are known. Although in principal this merely involves two further integrations, this proves to be laborious even on a fast electronic computer. We show that a close approximation to the integral can be obtained by replacing the unknown θ_1 and θ_2 by their maximum likelihood estimates in the integrand and changing the "degrees of freedom" by one unit. An example is given.

III. Choice of Parent and Prior Distributions.

A. Parent distribution

We would expect that inference about scale parameters would be particularly affected by kurtosis in the parent distributions, while the effect of skewness would be considerably less. We suppose, therefore, that the parent distributions in question are symmetric but not necessarily normal. A convenient class of such distributions are the power distributions used in our previous paper,

$$p(y|\theta,\sigma,\beta) = k \exp\left\{-\tfrac{1}{2}\left|\frac{y-\theta}{\sigma}\right|^{\frac{2}{1+\beta}}\right\} \qquad \begin{array}{l} -\infty < y < \infty \\ -\infty < \theta < \infty \\ 0 < \sigma < \infty \\ -1 < \beta < 1 \end{array} \tag{2}$$

with $$k = \left\{\Gamma\left(1+\frac{1+\beta}{2}\right) 2^{\left(1+\frac{1+\beta}{2}\right)}\sigma\right\}^{-1}.$$

When two samples are drawn from possibly different members of this class, the joint probability density will depend upon six unknown parameters, namely a set $(\beta_1,\theta_1,\sigma_1)$ associated with the first sample and a set $(\beta_2,\theta_2,\sigma_2)$ the other. We shall assume throughout this paper that the parents have the same parameter β. The ratio σ_2^2/σ_1^2 of the scale parameters is thus also the variance ratio.

B. Choice of prior distributions for θ_1, θ_2, and σ_1, σ_2.

We assume, as in our previous paper, that location parameters and the logarithms of the scale parameters are locally uniformly distributed a priori, namely

(3.1) $p(\theta_i) \propto c_1$

(3.2) $p(\log \sigma_i) \propto c_2$ or $p(\sigma_i) \propto \frac{1}{\sigma_i}$ $i = 1, 2.$

This assumption would be appropriate if we believed that any point in a region in which the likelihood for θ_1, θ_2, $\log \sigma_1$, and $\log \sigma_2$ was appreciable would have been as acceptable a priori as any other.

C. Prior distribution of β

In practice, the value of β would not be known. However, suppose the problem was the common one where errors could originate from a number of different sources, such as chemical sampling, chemical analysis and instrument deviations. Then the overall error would be function $y - \theta = f(\underline{\epsilon}) = f(\epsilon_1, \epsilon_2, \ldots.)$ of a number of independent component errors having variances $\sigma^2_{\epsilon_1}, \sigma^2_{\epsilon_2}, \ldots$. Often the percentage variation in each component would be small and $f(\underline{\epsilon})$ could be closely represented by the linear approximation $y-\theta = \alpha_o + \sum_i \alpha_i \epsilon_i$. If at least some of the products $\alpha_i \sigma_{\epsilon_i}$ were of the same order of magnitude as the largest such product so that no single source of error dominated the situation, then a central limit effect would control the error distribution. The normal distribution would then assume a central place in our thinking. We can represent this by regarding β now as a measure of non-normality and choosing a prior distribution for β with modal value at $\beta = 0$ and containing an adjustable parameter which controlled the degree of concentration about this mode. A convenient choice which was used in our previous paper is,

$$(4) \qquad p(\beta) = \frac{\Gamma(2a)}{\{\Gamma(a)\}^2 \, 2^{2a-1}} (1 - \beta^2)^{a-1} \qquad \begin{array}{r} -1 < \beta < 1 \\ a \geq 1 \end{array}$$

When "a" approaches infinity, $p(\beta)$ would approach a δ function at $\beta=0$, corresponding to an outright assumption that the parent is normal. When "a" has the value unity, $p(\beta)$ is rectangular, representing a situation where there is no more reason to expect a normal distribution than any other member of the family of distributions in equation (2). Any desired intermediate situation can be represented by an intermediate value of "a" and the sensitivity of the inference to the assumption of normality can be assessed by varying this concentration parameter.

IV. Derivation of the Posterior Distribution of the Variance Ratio σ_2^2/σ_1^2 for Fixed Values of θ_1, θ_2, and β.

From (2), 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:

$$(5) \quad \ell(\sigma_1,\sigma_2,\theta_1,\theta_2,\beta|\underline{y}_1,\underline{y}_2) = k\, \sigma_1^{-n_1}\, \sigma_2^{-n_2} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{2} n_i s_i(\beta,\theta) \Big/ \sigma_i^{\frac{2}{1+\beta}}\right\}$$

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

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

We first restrict ourselves to the situation where the values of the location parameters θ_1 and θ_2 are known. The joint posterior distribution of σ_1, σ_2 and β is then:

$$(6) \quad p(\sigma_1,\sigma_2, \beta|\theta_1,\theta_2, \underline{y}_1,\underline{y}_2)=k\ p(\sigma_1)p(\sigma_2)p(\beta)\ell(\sigma_1,\sigma_2,\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)$$

$$= p(\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)p(\sigma_1,\sigma_2|\beta,\theta_1,\theta_2, \underline{y}_1,\underline{y}_2).$$

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

$$(7) \quad p(\sigma_1,\sigma_2|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2) = \prod_{i=1}^{2} p(\sigma_i|\beta,\theta_i,\underline{y}_i)$$

where

$$p(\sigma_i|\beta,\theta_i,\underline{y}_i) = k_i\ \sigma_i^{-(n_i+1)} \exp\left\{-\tfrac{1}{2}\, n_i s_i(\beta,\theta)/\sigma_i^{\frac{2}{1+\beta}}\right\}$$

and

$$k_i = n_i \left\{\frac{n_i s_i(\beta,\theta)}{2}\right\}^{n_i(\frac{1+\beta}{2})} \Big/ \Gamma\left[1 + \frac{n_i(1+\beta)}{2}\right] .$$

It is seen to be the product of two independent inverted gamma distributions. The posterior distribution of σ_2^2/σ_1^2 is readily obtained by making the transformation $V = \sigma_2^2/\sigma_1^2$ and $W = \sigma_1$, and integrating our W. Thus,

$$(8) \quad p(V|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2) = k\ V^{\frac{n_1}{2} - 1}\left(1 + \frac{n_1 s_1(\beta,\theta)}{n_2 s_2(\beta,\theta)}\ V^{\frac{1}{1+\beta}}\right)^{-(n_1+n_2)(\frac{1+\beta}{2})}$$

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

It is convenient to consider the quantity $\frac{s_1(\beta,\theta)}{s_2(\beta,\theta)} V^{\frac{1}{1+\beta}}$, where it is to be remembered $V = \frac{\sigma_2^2}{\sigma_1^2}$ is a random variable and $\frac{s_1(\beta,\theta)}{s_2(\beta,\theta)}$ is a constant calculated from the observations. We have that,

$$(9) \quad p\left[\frac{s_1(\beta,\theta)}{s_2(\beta,\theta)} V^{\frac{1}{1+\beta}} \middle| \beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2\right] = p\left[F_{n_1(1+\beta),\ n_2(1+\beta)}\right],$$

an F distribution with $n_1(1+\beta)$ and $n_2(1+\beta)$ degrees of freedom.

In particular, when $\beta = 0$, the quantity $V \dfrac{\sum^{n_1} (y_{1i}-\theta_1)^2/n_1}{\sum^{n_2} (y_{2i}-\theta_2)^2/n_2}$ is distributed as F with n_1 and n_2 degrees of freedom. Further, when the value of β tends to -1 (the parent distributions tend to the rectangular form), the quantity $u = V\left\{\dfrac{\max|y_{1i}-\theta_1|}{\max|y_{2i}-\theta_2|}\right\}^2$ has the distribution,

$$(10) \quad \lim_{\beta\to -1} p(u|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2) = \frac{n_1 n_2}{2(n_1+n_2)} u^{\frac{n_1}{2}-1} \quad \text{for} \quad u \leq 1$$

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

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 simply

$$(11) \quad \Pr\left\{V > 1|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2\right\} = \Pr\left\{F_{n_1(1+\beta),\ n_2(1+\beta)} > \frac{s_1(\beta,\theta)}{s_2(\beta,\theta)}\right\}.$$

V. Relationship Between the Posterior Distribution $p(V|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)$ and Classical Procedures.

In the simplifying situation where θ_1, θ_2, and β are assumed known, a result parallel to that just derived can be obtained from the classical theory. It can easily be shown that in expression (5) the two power sums $n_1s_1(\beta,\theta)$ and $n_2s_2(\beta,\theta)$ when regarded as functions of the random variables $\underline{y}_1$ and $\underline{y}_2$ have as their joint moment generating functions,

$$(12) \qquad M_Y(t_1,t_2) = \prod_{i=1}^{2} \left\{1 - 2t_i\sigma_i^{\frac{2}{1+\beta}}\right\}^{-\frac{n_i(1+\beta)}{2}}$$

where $\quad Y = (n_1s_1(\beta,\theta),\ n_2s_2(\beta,\theta)\ )$.

Thus, letting $Y' = (n_1s_1(\beta,\theta),/\sigma_1^{\frac{2}{1+\beta}},\ n_2s_2(\beta,\theta)/\sigma_2^{\frac{2}{1+\beta}})$ we obtain

$$(13) \qquad M_{Y'}(t_1,t_2) = \prod_{i=1}^{2} (1 - 2t_i)^{-\frac{n_i(1+\beta)}{2}}$$

which is clearly the product of the moment generating functions of two independently distributed χ^2 distributions with $n_1(1+\beta)$ and $n_2(1+\beta)$ degrees of freedom respectively. Thus, the criterion $s_1(\beta,\theta)/s_2(\beta,\theta)$ 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 $s_1(\beta,\theta)/s_2(\beta,\theta)$ is

$$(14) \qquad \Pr\left\{F_{n_1(1+\beta),\ n_2(1+\beta)} > \frac{s_1(\beta,\theta)}{s_2(\beta,\theta)}\right\}$$

and is numerically equal to the probability for $V > 1$ given in equation (11). The level of significance associated with the observed ratio $s_1(\beta,\theta)/s_2(\beta,\theta)$ which can be derived from classical theory is thus precisely the probability a posteriori that the variance ratio σ_2^2/σ_1^2 exceeds unity when $\log \sigma_1$ and $\log \sigma_2$ are supposed locally uniform a priori.

A general test for comparing k variances for normal populations was derived by Neyman and Pearson (1931) using the likelihood ratio method. This test has been modified by Bartlett (1937) who showed that if we let

$$(15) \qquad \lambda(0) = \prod_{i=1}^{k}\left(\frac{N\, s_i(0,\theta)}{\sum_{i=1}^{k} n_i s_i(0,\theta)}\right)^{\frac{n_i}{2}}, \quad N = \sum_{i=1}^{k} n_i ,$$

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

$$(16) \qquad g(\beta) = 1 + [3k(1+\beta)]^{-1}\left\{\sum_{i=1}^{k} n_i^{-1} - N^{-1}\right\} .$$

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

$$(17) \qquad \lambda(\beta) = \prod_{i=1}^{k}\left(\frac{N\, s_i(\beta,\theta)}{\sum n_i s_i(\beta,\theta)}\right)^{\frac{n_i}{2}(1+\beta)} .$$

Now, provided β is not close to -1, the quantities $s_i(\beta,\theta)$ $(i = 1, 2, \ldots, k)$ follow scaled χ^2 distributions with appropriately adjusted degrees of freedom and the quantity $-2 \log \lambda(\beta)/g(\beta)$ is again approximately distributed as χ^2 with k degrees of freedom. It is hoped in later work to develop this generalization of Bartlett's test from the Bayesian point of view. For the present we concentrate on the problem for which $k = 2$ and do not consider this alternative form.

VI. Uncertainty Involved in Comparison of Variances when β is Not Known.

We continue to study the comparison of two variances and for the time being to assume that θ_1 and θ_2 are given. We have then demonstrated that "parallel" result for normal and non-normal parents belonging to the family of equation (2) can be obtained by Bayesian and by the Neyman-Pearson arguments. In most experimental situations we are unsure of the exact value of β and unfortunately, as we shall demonstrate, inferences about the variance ratio are sensitive to the choice of this value. This uncertainly can be easily taken into account using the Bayesian approach however.

Consider the data in Table I taken from Statistical Method in Research and Production, edited by O. L. Davies (1949).

This data were collected to compare the accuracy of an inexperienced analyst A_1 and an experienced analyst A_2 in their assay of carbon in a mixed powder. In the original analysis, the parent distributions were assumed normal; the null hypothesis was $\sigma_1^2/\sigma_2^2 = 1$ and this was tested against the alternative $\sigma_1^2/\sigma_2^2 > 1$. Here we shall assume instead that the parents are members of the class of distributions in (2) with identical β. The values of the location parameters θ_1 and θ_2 for this example are of course not known, but to simplify the analysis, we suppose at first that θ_1 and θ_2 are equal to the sample means of 6.55 and 5.77, respectively.

For each value of β, a significance level obtained by using the appropriate uniformly most powerful criterion can be calculated. This is numerically equal to the probability that the variance ratio V exceed unity given β, and is plotted against β in Figure (1a). It is 2.08% if the parent distribution is rectangular ($\beta \to -1$), 7.59% if the parent is normal ($\beta = 0$) and 9.90% if the parent is double exponential ($\beta = 1$). It is well known that the normal theory F criterion lacks robustness to non-normality, particularly to kurtosis but the type of sensitivity now being considered is, of course, of a different character. The criterion itself is being changed appropriately as the parent distribution is changed.

VII. The Posterior Distribution of V when β is Regarded as a Variable Parameter.

The dilemma arising from uncertainty about the value of β is easily remedied when a Bayesian standpoint is adopted and β is included in the formulation as a variable parameter. The joint posterior distribution of V and β can be written,

$$(18) \quad p(V,\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2) = p(\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)p(V|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)$$

where $p(V|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)$ is given by equation (6). The marginal distribution of β can be written as the product,

$$(19) \quad p(\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2) = p(\beta)\, \ell(\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)$$

where $p(\beta)$ is given by equation (4) and

$$\ell(\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2) = k\left\{\Gamma\left(1+\frac{1+\beta}{2}\right)\right\}^{-(n_1+n_2)} \prod_{i=1}^{2} \Gamma\left[1+n_i\left(\frac{1+\beta}{2}\right)\right]\left[n_i\, s_i(\beta,\theta)\right]^{-n_i\left(\frac{1+\beta}{2}\right)}$$

which is the integrated likelihood for β. We can thus regard $p(\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)$ as containing information of two kinds: The knowlege a priori about β is characterized by $p(\beta)$ and the information coming from the sample concerning β is represented by $\ell(\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)$. The marginal posterior distribution of β is shown in Figure (1b) with "a" taken to be unity. This marginal distribution thus essentially represents the information on β coming from the sample alone.

The difficulty in comparing variances rests on the "confounding" which occurs between the effect of inequality of variances and the effect of kurtosis. In terms of the Neyman-Pearson theory, the critical regions for tests on variances and tests of kurtosis overlap substantially -- see Box (1953a,b) -- so that it is usually difficult to know whether an observed discrepancy results from one or the other or a mixture of both.

Figures (1a) and (1b) together illustrate how, in line with our general discussion in section I, a satisfactory resolution is provided by Bayesian procedures. Figure (1b) shows what we are entitled to believe about the extent of the kurtosis of the parent distributions, while the former shows what we should believe about the equality of the variances given any particular degree of kurtosis.

In general, the posterior distribution of V is obtained by integrating out β from equation (18) giving

$$(20)\quad p(V|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2) = \int_{-1}^{+1} p(\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)p(V|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)\, d\beta$$

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

$$(21)\quad \Pr\left\{V > 1|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2\right\} = \int_{-1}^{+1} \Pr\left\{V > 1|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2\right\}p(\beta|\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)\, d\beta$$

where the first factor in the integrand is given in (11). In obtaining the above integral we are, in fact, averaging the various probabilities shown in Figure (1a) by the weight function given in Figure (1b). By so doing uncertainty in our inference about V induced by our uncertainty about β is removed. In particular, for the analyst example, $\Pr\{V > 1 | \theta_1, \theta_2, \underline{y}_1, \underline{y}_2\}$ is found to be 7.91%. For this particular set of data, this final value agrees very closely with the value 7.59% obtained using normal theory. The reason for this is seen from the figures. Figure (1b) shows the posterior distribution of β is very nearly centered about the value 0. Further, a transformation on β that would tend to make the posterior distribution in Figure (1b) approximately symmetric would also tend to transform the curve in Figure (1a) into a straight line. The averaging of equation (21) can therefore be expected for this particular sample to yield result close to the normal theory value.

In order to facilitate comparison with the result obtained by the more familiar Neyman-Pearson formulation, we have in the above calculated only posterior probability of V being greater than unity, since this parallel directly the significance level. We should perhaps emphasize that in most problems, we should be interested not in a single probability but in the whole posterior density function. An example illustrating this fuller analysis is used to illustrate the situation which we now discuss where the unrealistic assumption that θ_1 and θ_2 are known a priori is relaxed.

VIII. Posterior Distribution of V when θ_1 and θ_2 are regarded as Variable parameters.

To concentrate attention on important issues, we supposed initially that the location parameters θ_1 and θ_2 were known. No difficulties of principle associated with the elimination of nuisance parameters such as θ_1 and θ_2 arise of course within the Bayesian formulation, and this supposition is now relaxed.

We follow our discussion in section III and assume that θ_1 and θ_2 are locally uniformly distributed a priori as in (3.1). Upon "integrating out" these two parameters from the joint posterior distribution of the set $(\theta_1,\theta_2,V,\beta)$, we can then write the posterior distribution of β and V as:

$$(22) \quad p(V,\beta|\underline{y}_1,\underline{y}_2) = p(V|\beta,\underline{y}_1,\underline{y}_2)p(\beta|\underline{y}_1,\underline{y}_2)$$

The conditional posterior distribution of V for fixed value of β is given by:

$$(23) \quad p(V|\beta,\underline{y}_1,\underline{y}_2) = k\, V^{\frac{n_1}{2}-1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [n_2 s_2(\beta,\theta)+V^{\frac{1}{1+\beta}} n_1 s_1(\beta\theta)]^{-\frac{n_1+n_2}{2}(1+\beta)} d\theta_1 d\theta_2$$

where

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

and $s_i(\beta,\theta)$, i=1, 2 are given in (5).

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

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

has an F distribution with (n_1-1) and (n_2-1) degrees of freedom. Also, when the parents tends to the rectangular form $(\beta \to -1)$, the quantity $\omega = V(\frac{h_1}{h_2})^2$, where h_1 and h_2 are respectively the ranges of the first and the second sample, has the following limiting distribution,

$$(24) \quad \lim_{\beta \to -1} p(\omega|\beta,\underline{y}_1,\underline{y}_2) = k\,\omega^{\frac{n_1-1}{2}-1} [(n_1+n_2) - (n_1+n_2-2)\omega^{\frac{1}{2}}] \text{ for } \omega \le 1$$

$$= k\,\omega^{-\frac{n_2-1}{2}-1} [(n_1+n_2) - (n_1+n_2-2)\omega^{-\frac{1}{2}}] \text{ for } \omega > 1$$

with

$$k = \frac{n_1 n_2}{2(n_1+n_2)} \frac{(n_1-1)\ (n_2-1)}{(n_1+n_2-1)(n_1+n_2-2)} .$$

For other choice of parents, it does not appear possible to express the posterior distribution of V in (23) in terms of simple functions. Methods which can be employed to yield a close approximation to this distribution are now derived.

IX. Computational Procedures for the Posterior Distribution $p(V|\beta,\underline{y}_1,\underline{y}_2)$.

The numerical evaluation of equation (23) involves, among other things, computing a double integral for each value of V. This is laborous even on a fast electronic computer. However, a general expression for the moments of V is readily obtained in the form:

(25) $$E(V^r|\beta,\underline{y}_1,\underline{y}_2) = k \times$$

$$\frac{\int_{-\infty}^{\infty} [n_1 s_1(\beta,\theta)]^{-\frac{1}{2}(n_1+2r)(1+\beta)} d\theta_1 \int_{-\infty}^{\infty} [n_2 s_2(\beta,\theta)]^{-\frac{1}{2}(n_2-2r)(1+\beta)}}{\int_{-\infty}^{\infty} [n_1 s_1(\beta,\theta)]^{-\frac{1}{2}n_1(1+\beta)} d\theta_1 \int_{-\infty}^{\infty} [n_2 s_2(\beta,\theta)]^{-\frac{1}{2}n_2(1+\beta)} d\theta_2}$$

with

$$k = \Gamma[\tfrac{1}{2}(n_1+2r)(1+\beta)]\ \Gamma\ [\tfrac{1}{2}(n_2-2r)(1+\beta)] / \prod_{i=1}^{2} \Gamma[\tfrac{1}{2}n_i(1+\beta)]$$

Computation for the moments in the above expression only involves the comparatively simple evaluation of one dimensional integrals. It would now be possible to proceed by evaluating these moments and fitting appropriate forms of distributions suggested by the exact results obtainable when $\beta=0$ and $\beta \to -1$. The simpler and more intuitively satisfying procedure we actually employed is as follows: We can write equation (25) as:

$$E(V^r|\beta,\underline{y}_1,\underline{y}_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(V^r|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2)p(\theta_1,\theta_2|\beta,\underline{y}_1,\underline{y}_2)\ d\theta_1 d\theta_2$$

In the integrand, the moments of the conditional posterior distribution of V for fixed choices of θ_1 and θ_2 are given by

(26) $E(V^r|\beta,\theta_1,\theta_2,\underline{y}_1,\underline{y}_2) =$

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

and the joint posterior distribution of θ_1 and θ_2 is

(27) $$p(\theta_1,\theta_2|\beta,\underline{y}_1,\underline{y}_2) = \prod_{i=1}^{2}[n_i s_i(\beta,\theta)]^{-\frac{1}{2}n_i(1+\beta)} \Big/ \int_{-\infty}^{\infty}[n_i s_i(\beta,\theta)]^{-\frac{1}{2}n_i(1+\beta)}\,d\theta_i$$

It was shown in our previous paper that, for fixed β, the function

(28) $$f(\theta) = n\,s(\beta,\theta) = \sum_{i=1}^{n}|y_i-\theta|^{\frac{2}{1+\beta}}$$

has continuous first derivative and an unique minimum which is attained at some point in the interval $[y_S, y_L]$. When $\beta=0$ or $\beta < -\frac{1}{3}$, it is easy to see that $f^3(\theta)$ exists and is continuous. Thus, for these values of β, we can employ Taylor's theorem to expand $f(\theta)$ into:

(29) $$f(\theta) \backsim f(\hat{\theta}) + \frac{1}{2}f''(\hat{\theta})(\theta-\hat{\theta})^2 ,$$

where $\hat{\theta}$ is the point at which $f(\theta)$ attains minimum. This approximation will be satisfactory when β is not close to -1. Using this result, we find that the moments of V in equation (25) is approximately:

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

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

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

$$(31) \qquad c(V) = V^{\frac{1}{1+\beta}}\,\frac{n_1 s_1(\beta,\hat{\theta})/[n_1(1+\beta)-1]}{n_2 s_2(\beta,\hat{\theta})/[n_2(1+\beta)-1]}$$

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 posterior distribution of $c(V)$ can be closely approximated using ordinary F tables. In this approximation, the nuisance parameters θ_1 and θ_2 in the posterior 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. Although calculating these maximum likelihood values still require the use of numerical methods, this is a procedure of great simplicity compared with the exact evaluation of multiple integrals for each V.

The justification supplied above for this simple approximation is, unfortunately, only valid when $\beta=0$ and when β is less than $\frac{1}{3}$ but not close to -1. However, it seems that in practice the approximation has a much wider usefulness. Using the previous analytical data without now the assumption that the location parameters θ_1 and θ_2 are known, we find empirically that (30) and (31) give a very close approximation over the range $-0.6 < \beta < 1$ and a tolerable approximation even when $\beta = -0.8$, although this representation rapidly becomes inadequate as β approaches -1. The excellent approximation over the range $-0.8 < \beta < 0.8$ given by equation (30) for the first four moments of $V^{-1} = \sigma_1^2/\sigma_2^2$ using the analytical example is shown in Table II. The exact and approximate posterior distribution of V^{-1} has been obtained for various values of β and specimen values of the probability densities are given in Table III. Some discrepancies occur when $\beta = -.8$, but for the other distributions the agreement is very close.

The authors are somewhat embarrassed by the fact that the approximation seems most accurate where the present mathematical justification is weakest, namely for $-\frac{1}{3} < \beta < 0$ and $0 < \beta < 1$. A possible reason that such good agreement is found in these cases is that only values close to $\hat{\theta}$ have sufficient weight to matter very much.

X. Posterior Distribution of the Variance Ratio with all Nuisance Parameters Eliminated.

We now reconsider the analytical data of Table I. The diagrams in Figure (2) show the computed distributions of

$$V^{-1} = \frac{\sigma_1^2}{\sigma_2^2} = \frac{\text{Variance of Analyst } A_1}{\text{Variance of Analyst } A_2}$$

for various choices of β where this time the simplifying assumption that the location parameters θ_1 and θ_2 are known and equal to the sample means is not made. In computing the posterior distributions of V^{-1} for $\beta = -0.6$ and $\beta = 0.99$, θ_1 and θ_2 are eliminated by actual integration and also by the approximation procedure mentioned above. In each of these two cases, the curves obtained by the two methods are practically indistinguishable. For the cases $\beta=0$ and $\beta = -1$, the exact distributions are easily calculated.

We see from these figures how the posterior distribution of the variance ratio changes as the value of β is changed. When the parents are rectangular ($\beta \to -1$), the distribution is sharply concentrated around its model value. The degree of concentration is considerably less pronounced when the parents are normal ($\beta=0$) and the model value is smaller. Finally, when the parents are nearly double exponential ($\beta=.99$), the distribution has an even smaller modal value and becomes rather flat with a long tail toward the right. It is evident that inferences concerning the variance ratio depend heavily upon the form of the parent distributions.

As in the case θ_1 and θ_2 assumed known, the uncertainty in our inferences about V can be removed when information concerning the parameter β is taken into account. In the joint posterior distribution of V and β in (22), the marginal posterior distribution of β is given by

(32) $$p(\beta|\underline{y}_1,\underline{y}_2) = k\, \ell(\beta|\underline{y}_1,\underline{y}_2)\, p(\beta)$$

where

$$\ell(\beta|\underline{y}_1,\underline{y}_2) = \frac{\prod_{i=1}^{2} \Gamma[1+\frac{n_i(1+\beta)}{2}]}{\left\{\Gamma(1+\frac{1+\beta}{2})\right\}^{n_1+n_2}} \prod_{i=1}^{2} \int_{-\infty}^{\infty} [n_i s_i(\beta,\theta)]^{-\frac{n_i(1+\beta)}{2}} d\,\theta_i$$

which is the integrated likelihood of β, $p(\beta)$ is the prior distribution of β given in (4) and k is the normalizing constant. Upon integrating out β from (22) the final posterior distribution of the variance ratio V is then

(33) $$p(V|\underline{y}_1,\underline{y}_2) = \int_{-1}^{+1} p(V|\beta,\underline{y}_1,\underline{y}_2,)p(\beta|\underline{y}_1,\underline{y}_2)\, d\beta$$

For the analytical data, the posterior distribution of β in (32) and the final posterior distribution of the variance ratio $V^{-1} = \sigma_1^2/\sigma_2^2$ are shown, respectively, in Figure (3a) and (3b)*.

---

* In computing (32), we again take the value of the parameter "a" in $p(\beta)$ to be unity.

Table I

Results From Analysis of Identical Sample

$x = (\% \text{ of carbon} - 4.50) \times 100$

Analyst A_1 Sample No.	x_1	Analyst A_2 Sample No.	x_2
1	-10	1	- 8
2	16	2	- 3
3	- 8	3	20
4	9	4	22
5	5	5	3
6	- 5	6	5
7	5	7	10
8	-11	8	14
9	25	9	-21
10	22	10	2
11	16	11	7
12	3	12	8
13	40	13	16
14	0		
15	- 5		
16	16		
17	30		
18	-14		
19	25		
20	-28		
Mean	6.55		5.77

Table II.

Moments of the Variance Ratio $V^{-1} = \sigma_1^2/\sigma_2^2$ for the Analytical Data for Various choices of β^+.

β	-.8	-.6	-.4	-.2	.2	.4	.6	.8
μ_1'	2.16	2.03	2.07	2.20	2.55	2.72	2.89	3.05
μ_1^*	2.11	2.02	2.06	2.20	2.55	2.72	2.89	3.04
μ_2'	5.12	4.71	5.13	6.10	9.00	10.79	12.77	14.97
μ_2^*	4.99	4.69	5.11	6.08	9.01	10.79	12.74	14.84
μ_3'	13.31	12.50	15.23	21.24	44.02	62.14	86.19	117.63
μ_3^*	13.14	12.49	15.18	21.17	44.07	62.14	85.69	115.94
μ_4'	38.04	38.16	54.53	93.54	300.60	525.11	896.08	1497.80
μ_4^*	39.00	38.44	54.51	93.36	301.04	524.67	888.07	1465.83

+ In this table, μ_r' gives moment obtained by direct evaluation of (25) using Simpson's Rule, while μ_r^* shows moment approximated by the method given in (30).

Table III

Specimen of Probability Densities of the Variance Ratio $V^{-1}=\sigma_1^2/\sigma_2^2$ for the Analytical Data for Various values of β^+

σ_1^2/σ_2^2	$\beta = -.8$		$\beta = -.6$		$\beta = -.2$	
0.5	.005	.020	.017	.021	.048	.049
1.0	.099	.159	.229	.245	.299	.301
1.2	.209	.264	.378	.388	.389	.390
1.4	.353	.388	.507	.510	.444	.444
1.6	.503	.510	.583	.580	.461	.461
1.8	.619	.597	.594	.588	.449	.448
2.2	.628	.579	.474	.468	.372	.371
2.6	.412	.380	.301	.297	.276	.275
3.0	.206	.194	.168	.167	.192	.192
4.0	.026	.028	.032	.032	.070	.069
	$\beta = .2$		$\beta = .6$		$\beta = .8$	
0.5	.070	.069	.090	.089	.100	.100
1.0	.274	.274	.258	.257	.253	.253
1.2	.329	.329	.294	.293	.282	.283
1.4	.359	.359	.311	.311	.294	.295
1.6	.367	.368	.313	.313	.295	.296
1.8	.359	.360	.305	.306	.287	.287
2.2	.314	.315	.272	.273	.256	.257
2.6	.255	.256	.231	.231	.219	.219
3.0	.200	.200	.190	.190	.183	.183
4.0	.100	.100	.110	.110	.111	.111

+ For each value of β, the densities in the first column are obtained by direct evaluation of equation (23) using Simpson's Rule, and those in the second column are calculated using (31).

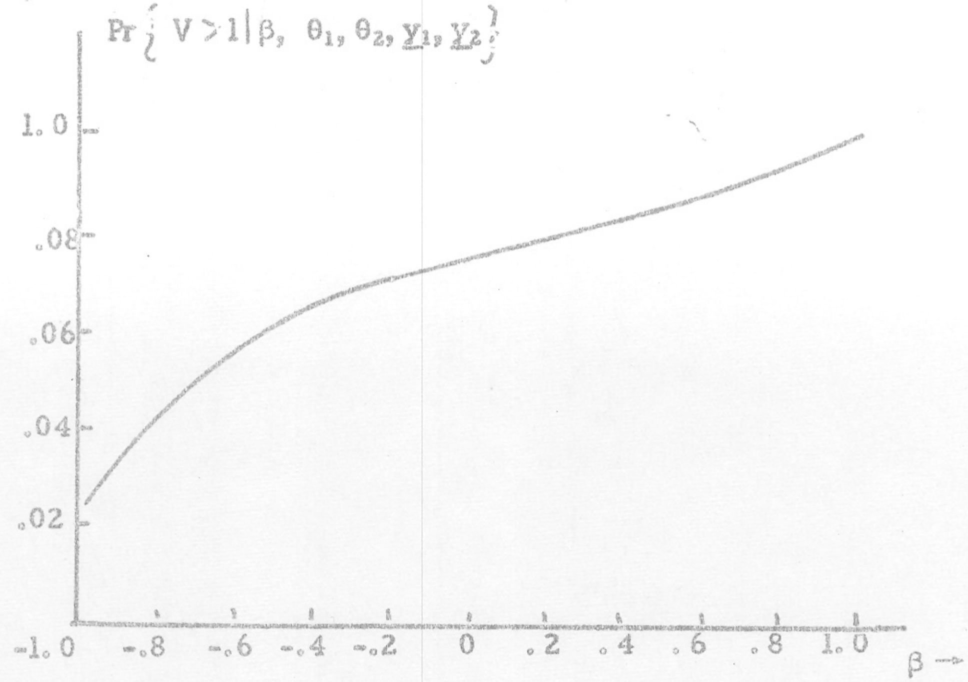

Fig. 1a. Posterior Probability that V Exceeds Unity for Various Choices of β.

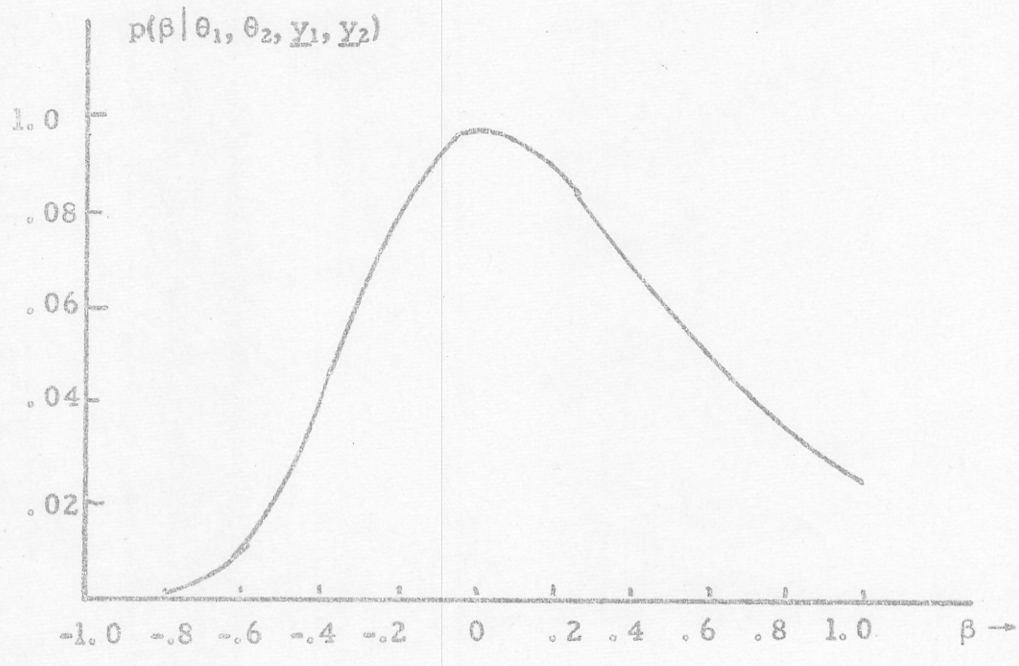

Fig. 1b. Posterior Distribution of β for Given θ_1 and θ_2.

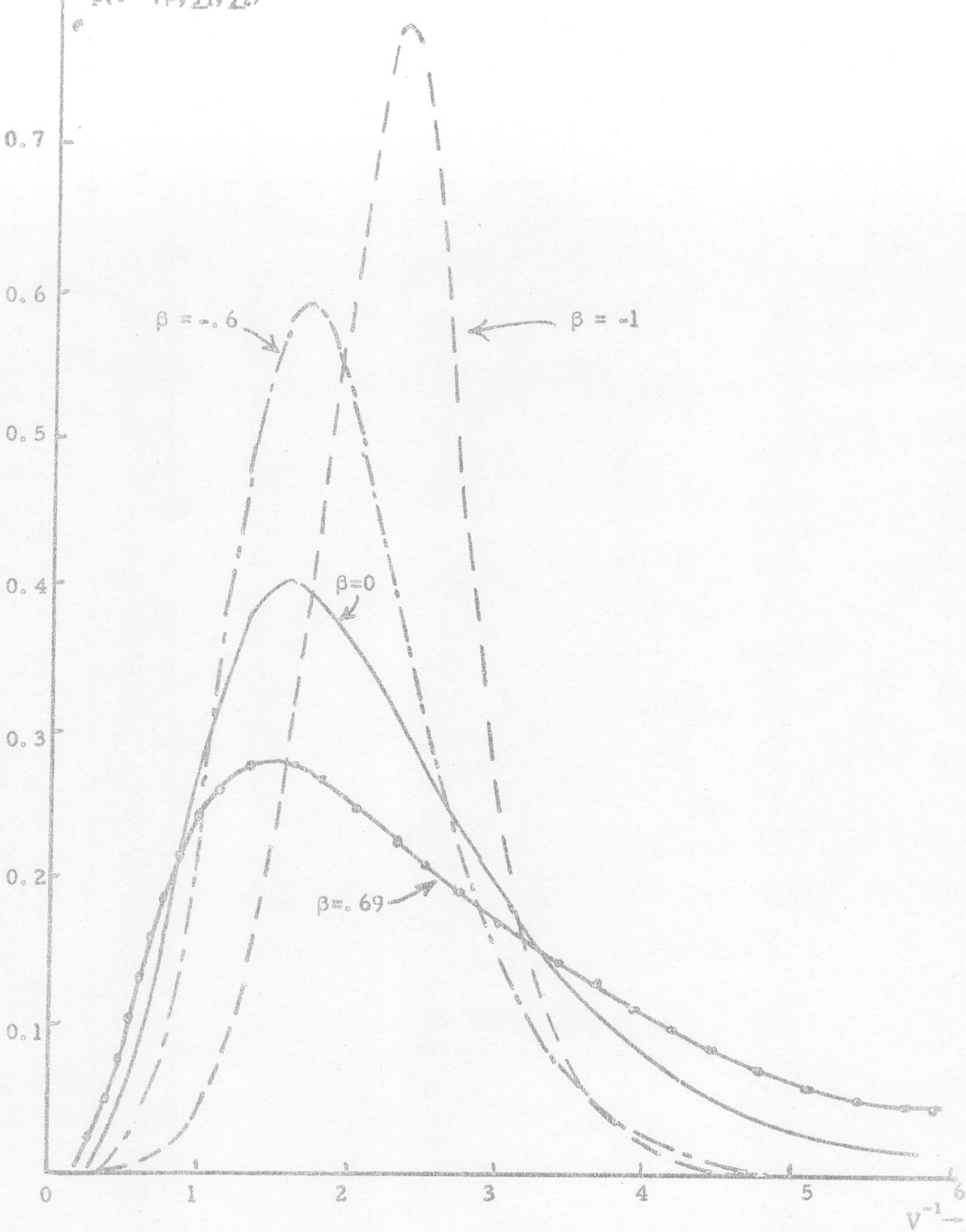

Fig. 2. Posterior Distribution of V^{-1} for Various Choices of β.

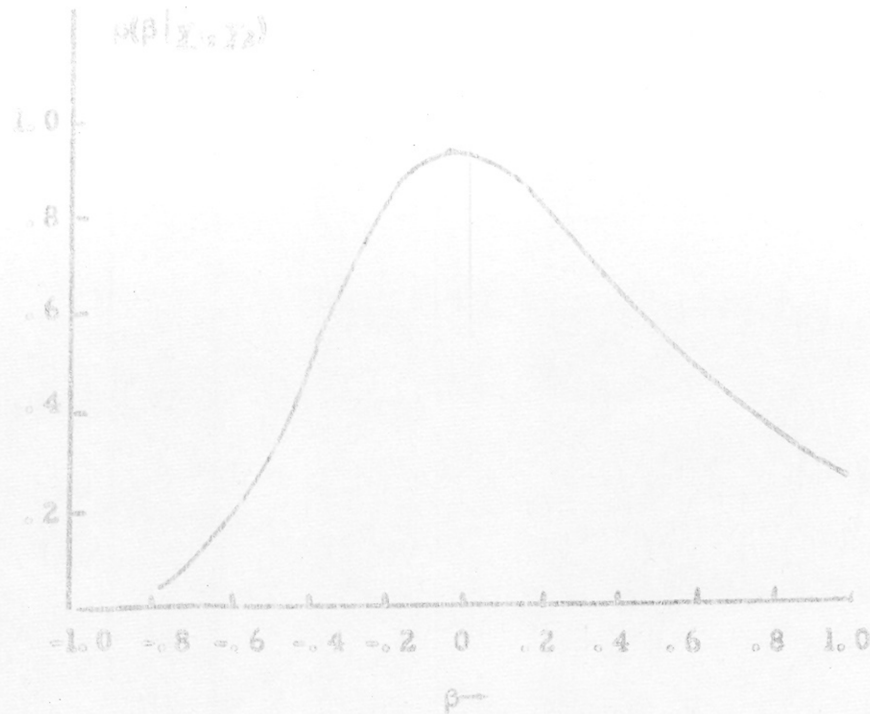

Fig. 3a. Posterior Distribution of β

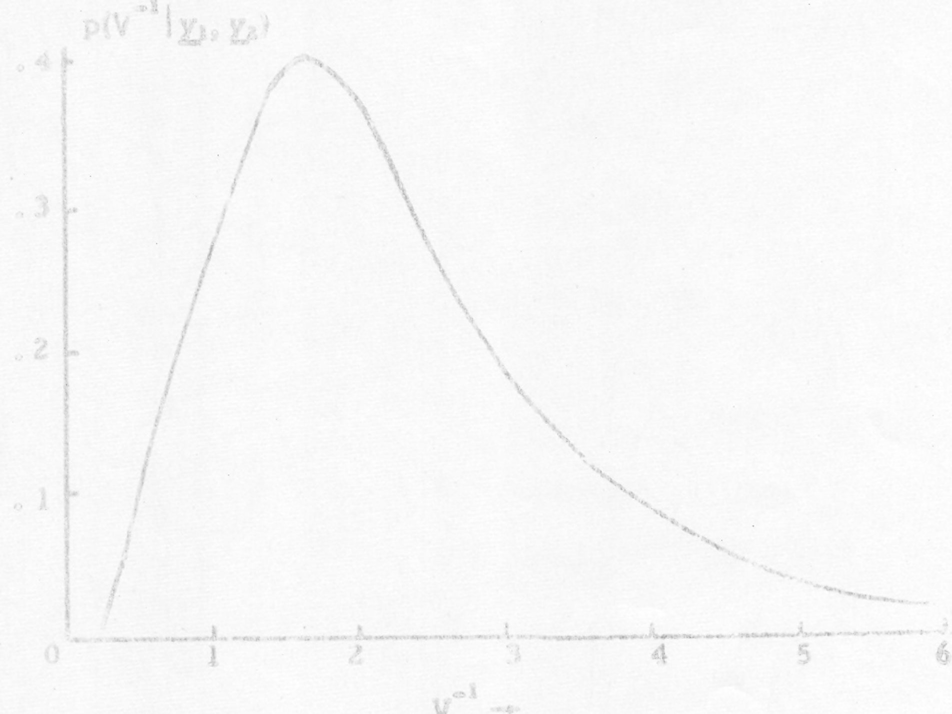

Fig. 3b. Posterior Distribution of V^{-1}

References

Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. Proc. Roy. Soc., A, 160, 268.

Box, G.E.P. (1953a). A note on regions for tests of Kurtosis. Biometrika, 40, 465.

Box, G.E.P. (1953b). Non-normality and tests on variances. Biometrika, 40, 318.

Box, G.E.P. and Anderson, S.L. (1955). Permutation theory in the derivation of robust creteria and the study of departures from assumptions. J.R.S.S., B, 17, 1.

Box, G.E.P. and Tiao, G.C. (1962). A further look at robustness via Bayes' theorem. Biometrika, 49, 419.

Davies, O.L. editor (1949). Statistical Methods in Research and Production-- With Special Reference to the Chemical Industry. Oliver and Boyd: London.

Neyman, J. and Pearson, E.S. (1931). On the problem of k samples. Bull. Acad. Polonaise Sci. Lett., A, 460.