MULTIPARAMETER PROBLEMS FROM A BAYESIAN POINT OF VIEW[1]

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

University of Wisconsin

1. Introduction. It has long been known that when attention is focused on a single parameter or comparison of parameters such as the difference in means or the ratio of variances from the normal samples, then using "non-informative" prior distributions Bayesian results exactly paralleling classical procedures and involving the standard t and F distributions can be obtained. When there are many parameters certain "portmanteau" multi-comparison procedures can be derived on the sampling theory approach and they are frequently of great value when used to supplement individual comparisons. These include the χ^2-goodness of fit test, the analysis of variance test and Bartlett's test to compare variances (when it is suitably modified so as to be robust to non-normality). This paper shows that these portmanteau procedures do have a simple and natural Bayesian interpretation. Furthermore, using Bayesian methods, it is possible to cover important cases not amenable to treatment by classical techniques.

Suppose in a particular investigation we have computed an appropriate posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$ where $\boldsymbol{\theta}$ is a k-dimensional vector of parameters of interest and $\mathbf{y}$ is an n-dimensional vector of observations. Then from the Bayesian point of view, all inferential problems concerning $\boldsymbol{\theta}$ may be answered in terms of $p(\boldsymbol{\theta} \mid \mathbf{y})$. In practice, inference involves a communication with the mind, and usually it is difficult to comprehend a function in k-dimensions. Fortunately, there are often specific individual features of $p(\boldsymbol{\theta} \mid \mathbf{y})$ of interest which can be appreciated by one, two or three dimensional thought. For example, marginal distributions may be of interest. Or we may inspect conditional distributions of a small subset of the parameters for specific values of the other parameters. With high speed electronic computers available, print-outs of two dimensional sections of such distributions can be readily obtained. The value of such appraisals of the estimation situation is very great, as has been repeatedly pointed out by Barnard in connection with the likelihood principle.

Another feature of $p(\boldsymbol{\theta} \mid \mathbf{y})$ which is of value is a posterior probability region for the parameters. Often the region over which the posterior density is non-zero extends over infinite ranges in the parameter space. Nevertheless over a substantial part of the parameter space the density may be small or negligible. It is therefore possible to delineate a comparatively small region which contains most (say 95%) of the probability mass. Obviously there is an infinite number of ways such a region can be chosen. We must therefore decide what properties we would like the region to have. Either of the following two principles seems to be

Received 8 February 1965; revised 12 April 1965.

[1] This research was supported by the Office of Naval Research under Contract Nonr-1202(17), Project Nro 42-222.

intuitively sensible:

(i) A "best" region should be such that the probability density of every point inside it is at least as large as any point outside it.

(ii) A "best" region should be such that for a given probability content, it occupies the smallest possible volume in the parameter space.

Either of the above principles could be adopted as axiomatic and it is easy to show that the other follows as a natural consequence. It is desirable to give such a region a name and we will call it a region of highest posterior density or a H.P.D. region for short.

We will adopt the first principle in order to give a formal definition.

DEFINITION. Let $p(\boldsymbol{\theta} \mid \mathbf{y})$ be a posterior density function. A region R in the parameter space of $\boldsymbol{\theta}$ is called a H.P.D. region of content $(1 - \alpha)$ if

(1.1) (i) $\Pr\{\boldsymbol{\theta} \varepsilon R \mid \mathbf{y}\} = 1 - \alpha$, and

(ii) for $\boldsymbol{\theta}_1 \varepsilon R$ and $\boldsymbol{\theta}_2 \notin R$, $p(\boldsymbol{\theta}_1 \mid y) \geqq p(\boldsymbol{\theta}_2 \mid \mathbf{y})$.

Some properties of the H.P.D. region.

(i) It follows immediately from the above definition that for a given probability content $(1 - \alpha)$, the H.P.D. region has the smallest possible volume in the parameter space of $\boldsymbol{\theta}$.

(ii) If we make the assumption that $p(\boldsymbol{\theta} \mid \mathbf{y})$ is non-uniform over any region in the space of $\boldsymbol{\theta}$, then the H.P.D. region of content $(1 - \alpha)$ is unique. Further, if $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are two points such that $p(\boldsymbol{\theta}_1 \mid \mathbf{y}) = p(\boldsymbol{\theta}_2 \mid \mathbf{y})$, then these two points are simultaneously included in or excluded by a $(1 - \alpha)$ H.P.D. region. The converse is also true. That is, if $p(\boldsymbol{\theta}_1 \mid \mathbf{y}) \neq p(\boldsymbol{\theta}_2 \mid \mathbf{y})$, then there exists a $(1 - \alpha)$ H.P.D. region which includes one point but not the other.

(iii) *Effect of transformation.* Let $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\theta})$ be, say, a one to one transformation of the parameters $\boldsymbol{\theta}$ to $\boldsymbol{\phi}$. It is obvious that any region of content $(1 - \alpha)$ in the space of $\boldsymbol{\theta}$ transforms into a region of the same content in the space of $\boldsymbol{\phi}$. But it is clear from their definition that H.P.D. regions in $\boldsymbol{\theta}$ will not in general transform into H.P.D. regions in $\boldsymbol{\phi}$. Such regions are, however, invariant under linear transformation.

This exactly parallels the situation for smallest confidence regions which are similarly not invariant under general transformation. On the implied assumption that such lack of invariance is bad, it has been suggested for example that the region should be based on the likelihood itself. That is, the boundary of the region should follow a likelihood contour. In particular, Hildreth (1963) has proposed that a $100(1 - \alpha)\%$ region R be based on

$$\int_R l(\boldsymbol{\theta} \mid \mathbf{y})\, d\boldsymbol{\theta} / \int_\Omega l(\boldsymbol{\theta} \mid \mathbf{y})\, d\boldsymbol{\theta} = (1 - \alpha) \tag{1.2a}$$

with the property that for $\boldsymbol{\theta}_1 \varepsilon R$ and $\boldsymbol{\theta}_2 \notin R$,

$$l(\boldsymbol{\theta}_1 \mid \mathbf{y}) > l(\boldsymbol{\theta}_2 \mid \mathbf{y}). \tag{1.2b}$$

It will be observed that although the Inequality (1.2b) is preserved under general transformation, the Equality (1.2a) will not be. A posterior region based

upon the likelihood which is in a sense invariant under general transformation can be obtained as follows. For a fixed prior distribution $p_0(\boldsymbol{\theta})$, choose a region R such that

$$\int_R l(\boldsymbol{\theta} \mid \mathbf{y}) p_0(\boldsymbol{\theta})\, d\boldsymbol{\theta} / \int_\Omega l(\boldsymbol{\theta} \mid \mathbf{y}) p_0(\boldsymbol{\theta})\, d\boldsymbol{\theta} = (1 - \alpha) \tag{1.3a}$$

$$l(\boldsymbol{\theta}_1 \mid \mathbf{y}) > l(\boldsymbol{\theta}_2 \mid \mathbf{y}), \quad \text{for} \quad \boldsymbol{\theta}_1 \varepsilon R \quad \text{and} \quad \boldsymbol{\theta}_2 \varepsilon R. \tag{1.3b}$$

Both (1.3a) and (1.3b) are invariant under general transformations. This region which tries to make the best of both worlds is, however, a somewhat artificial construction. If we believe in the appropriateness of the prior distribution $p_0(\boldsymbol{\theta})$, then we should surely not adopt a region for which the posterior density for points outside can be greater than that for points inside.

It seems that we cannot hope for invariance for a genuine measure of credibility. It needs to be remembered that invariance under transformation and virtues are not synonymous. For problems which should not be invariant under transformation, a search for invariance serves only to guarantee inappropriate solutions.

Graphical representation. Clearly when there are only two parameters, a diagram showing the point of maximum posterior density and, say, a 95 % H.P.D. region would advise the investigator of most of what the data had to tell him. A more informative plot would be one showing simultaneously the boundaries of, say, the 50 %, 25 %, and 10 % H.P.D. regions. In such a case we would be back to the plot of posterior density contours labelled according to their interior content. This graphical approach could be extended to three or four parameters by exhibiting a "grid" of two dimensional θ_1, θ_2 plots for various combinations of θ_3 and θ_4. An instrument such as the "Calcomp" plotter can produce such plots automatically from digital computer output and such plotting should be part of the normal stock in trade of the modern practicing statistician. This technique is valuable for appreciating peculiarities in an estimation situation, and is therefore particularly important in exploring new problems.

The use of the distribution of the function $p(\boldsymbol{\theta} \mid \mathbf{y})$. There are certain problems which are "standard" and for which properties of the H.P.D. regions are readily comprehended. The inferential problems concerning k means and k variances discussed below are, for example, of this type. When there are a large number of parameters of interest and no special peculiarities of the estimation situation, it is useful to have a way of knowing whether or not a parameter point $\boldsymbol{\theta}_0$ lies inside or outside a H.P.D. region of content $(1 - \alpha)$. From the definition and properties of H.P.D. regions, we see that if R_α is a H.P.D. region of content $(1 - \alpha)$, then the event $\boldsymbol{\theta} \varepsilon R_\alpha$ is equivalent to the event that $p(\boldsymbol{\theta} \mid \mathbf{y}) > c$, where c is a suitably chosen positive constant. It follows that the parameter point $\boldsymbol{\theta}_0$ is covered by the H.P.D. region of content $(1 - \alpha)$ if and only if

$$\Pr\{p(\boldsymbol{\theta} \mid \mathbf{y}) > p(\boldsymbol{\theta}_0 \mid \mathbf{y}) \mid \mathbf{y}\} \leqq 1 - \alpha. \tag{1.4}$$

Thus, once the posterior distribution of the quantity $p(\boldsymbol{\theta} \mid \mathbf{y})$ or some function of it can be determined, this question can be answered. The main purpose of this

paper is to consider the specific nature of the region $p(\boldsymbol{\theta} \mid \mathbf{y}) > c$ for a number of examples of interest. While preparing this paper, we have become aware of the recent work of Lindley (1964) which contains much of the same ideas discussed in Sections 1–4 of this paper.

2. The linear model. As a first example, we consider the familiar linear model

$$\mathbf{y} = X\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{2.1}$$

where $\mathbf{y}$ is a $n \times 1$ vector of observations, X a $n \times k$ matrix of constants, $\boldsymbol{\theta}$ a $k \times 1$ vector of unknown coefficients and $\boldsymbol{\varepsilon}$ a $n \times 1$ vector of disturbances. It is assumed that $\boldsymbol{\varepsilon}$ is normally distributed with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = I\sigma^2$ where I is a $n \times n$ identity matrix. Then, on the usual assumption that locally the prior distribution $p(\boldsymbol{\theta}, \log \sigma)$ is approximately constant, the posterior distribution of $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto \{1 + [(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})/\nu s^2]\}^{-\frac{1}{2}(\nu+k)} \tag{2.2}$$

with $\hat{\boldsymbol{\theta}} = (X'X)^{-1}X'\mathbf{y}$, $\nu = n - k$ and $s^2 = \nu^{-1}(\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}})$. This is of course the multivariate-t distribution as discovered by Cornish (1954) and Dunnett and Sobel (1954). Further, the quantity $(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})/ks^2$ is distributed as F with (k, ν) degrees of freedom. Suppose we are now interested in the question: Is the parameter point $\boldsymbol{\theta}_0 = (\theta_{10}, \cdots, \theta_{k0})$ included in the H.P.D. region of content $(1 - \alpha)$? According to the above argument, we then need to calculate the probability of the event $p(\boldsymbol{\theta} \mid \mathbf{y}) > p(\boldsymbol{\theta}_0 \mid \mathbf{y})$.

Now $p(\boldsymbol{\theta} \mid \mathbf{y})$ is a monotonic decreasing function of the quantity $(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'X'X \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})/ks^2$ which is distributed *a posteriori* as $F_{(k,\nu)}$. The particular point $\boldsymbol{\theta}_0$ is then included in the H.P.D. region of content $(1 - \alpha)$ if and only if

$$(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) < ks^2F_\alpha(k, \nu) \tag{2.3}$$

where $F_\alpha(k, \nu)$ is the upper $100\alpha\%$ point of an F distribution with (k, ν) degrees of freedom. In this particular example, the H.P.D. region is identical with the confidence region and (2.3) is appropriate to decide if a given point $\boldsymbol{\theta}_0$ lies inside or outside the corresponding confidence region. Equivalently, the quantity

$$\Pr\{F_{(k,\nu)} < (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})/ks^2\} \tag{2.4}$$

gives the content of the H.P.D. region which just covers the point $\boldsymbol{\theta}_0$. In the Neyman-Pearson framework, the complement of (2.4) gives the significance level associated with the null hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Generalization to the corresponding linear multivariate model can be readily obtained, but we shall not pursue this further here.

3. The goodness of fit problem. As a second example, we consider some aspects of the classical goodness of fit problem. Suppose we have observed k frequencies $\mathbf{f} = (f_1, \cdots, f_k)$. These frequencies are *independently* distributed as Poisson variables with means $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_k)$ respectively. The likelihood function is

$$L(\boldsymbol{\mu} \mid \mathbf{f}) \propto \prod_{s=1}^{k} \mu_s^{f_s} e^{-\mu_s}. \tag{3.1}$$

Let us assume that the prior distribution of $\boldsymbol{\mu}$ is of the form

$$p(\boldsymbol{\mu}) \propto \prod_{s=1}^{k} \mu_s^{\gamma} \tag{3.2}$$

where γ is some constant. The posterior distribution of $\boldsymbol{\mu}$ is then,

$$p(\boldsymbol{\mu} \mid \mathbf{f}) = \prod_{s=1}^{k} [1/\Gamma(m_s + 1)] \mu_s^{m_s} e^{-\mu_s} \tag{3.3}$$

with $m_s = f_s + \gamma$, which is seen to be the product of k independent gamma distributions. Again, in answering the question whether a specific parameter point $\boldsymbol{\mu}_0 = (\mu_{10}, \cdots, \mu_{k0})$ lies inside a H.P.D. region of content $(1 - \alpha)$, we need to calculate the posterior probability of the event

$$p(\boldsymbol{\mu} \mid \mathbf{f}) > p(\boldsymbol{\mu}_0 \mid \mathbf{f}). \tag{3.4}$$

Following the approach adopted for example by Box (1949), we consider the quantity

$$M = -2 \log W \tag{3.5}$$

where

$$W = \prod_{s=1}^{k} (\mu_s/m_s)^{m_s} \exp -(\mu_s - m_s). \tag{3.6}$$

It is clear that $p(\boldsymbol{\mu} \mid \mathbf{f})$ is a monotonic decreasing function of M, so that the statement in (3.4) is equivalent to the statement

$$M < M_0, \tag{3.7}$$

where

$$M_0 = -2 \log \prod_{s=1}^{k} (\mu_{s0}/m_s)^{m_s} \exp -(\mu_{s0} - m_s). \tag{3.8}$$

We now obtain the cumulant generating function of M. It is straightforward to verify that the characteristic function of M is

$$\begin{aligned} E(e^{itM}) &= E(W^{-2it}) \\ &= \prod_{s=1}^{k} (e/m_s)^{-2itm_s} \{\Gamma[m_s(1 - 2it)]/\Gamma(m_s)\}(1 - 2it)^{-m_s(1-2it)}. \end{aligned} \tag{3.9}$$

Taking logarithms and employing Stirling's series, we obtain the cumulant generating function,

$$\kappa_M(it) = a - \tfrac{1}{2}k \log (1 - 2it) + \sum_{r=1}^{\infty} \alpha_r (1 - 2it)^{-(2r-1)} \tag{3.10}$$

where "a" is some constant independent of t,

$$\alpha_r = [B_{2r}/2r(2r - 1)] \sum_{s=1}^{k} m_s^{-(2r-1)} \tag{3.11}$$

and the B_{2r} are Bernoulli numbers. Thus, the density function of M can be expressed as a weighted series of χ^2 densities, the leading term having k degrees of freedom.

Approximation to the distribution of M. Once an asymptotic series of this type is established, it is of course possible to obtain approximations in various ways,

e.g. Bartlett (1937), Hartley (1940) and Box (1949). From (3.10) the rth cumulant of M is, to order m_s^{-1},

$$\kappa_r(M) = 2^{r-1}(r-1)!k\{1 + Ar\} \tag{3.12}$$

where $A = (1/6k)\sum m_s^{-1}$. The Bartlett type of approximation is to take

$$M \sim (1 + A)\chi_k^2. \tag{3.13}$$

The rth cumulant of this approximate form is

$$\begin{aligned} \kappa_1(M) &= k(1 + A) \\ \kappa_r(M) &= 2^{r-1}(r-1)!\, k\{1 + rA + \tbinom{r}{2}A^2 + \cdots\}, \quad r \geqq 2. \end{aligned} \tag{3.14}$$

Thus, the first cumulants in (3.12) and (3.14) are exactly the same and, to order m_s^{-1}, the higher order cumulants in the two expressions are also identical.

Alternatively, an approximation to the distribution of M can be obtained by equating κ_1 and κ_2 in (3.12) to that of a scaled χ^2 variable, say $a\chi_b^2$. It is readily seen that

$$a = [(1 + 2A)/(1 + A)] \quad \text{and} \quad b = k\cdot(1 + A)^2/(1 + 2A). \tag{3.15}$$

The rth cumulant of this approximation is

$$\kappa_r(M) = 2^{r-1}(r-1)!\, k\cdot(1 + 2A)^{r-1}/(1 + A)^{r-2} \tag{3.16}$$

which can be written as

$$\begin{aligned} \kappa_1(M) &= k(1 + A) \\ \kappa_2(M) &= 2k(1 + 2A) \\ \kappa_r(M) &= 2^{r-1}(r-1)!\, k\{1 + rA + \tbinom{r-1}{2}A^2 + \cdots\}, \quad r > 2. \end{aligned} \tag{3.17}$$

We see that, to order m_s^{-1}, the cumulants in (3.12) and (3.17) are again identical. Further, the error committed in (3.17) is less than that in (3.14).

Relationship with K. Pearson's "Chi-square". When $\mu_{10}, \cdots, \mu_{k0}$ are *large* the quantity M_0 in (3.8) can be written

$$\begin{aligned} M_0 &= 2\textstyle\sum_{s=1}^{k}\{(\mu_{s0} - m_s) + m_s \log[1 + (m_s - \mu_{s0})/\mu_{s0}]\} \\ &= 2\textstyle\sum_{s=1}^{k}\{(\mu_{s0} - m_s) + [1 + (m_s - \mu_{s0})/\mu_{s0}] \\ &\quad \cdot\mu_{s0}\log[1 + (m_s - \mu_{s0})/\mu_{s0}]\} \\ &= \textstyle\sum_{s=1}^{k}\{[(m_s - \mu_{s0})^2/\mu_{s0}] + O(\mu_{s0}^{-\frac{1}{2}})\}. \end{aligned} \tag{3.16}$$

Suppose we take the prior distribution of μ in (3.2) to be uniform, i.e., $\gamma = 0$, so that $m_s = f_s$. Then M_0 becomes approximately

$$M_0 \cong \textstyle\sum_{s=1}^{k}[(f_s - \mu_{s0})^2/\mu_{s0}] \tag{3.17}$$

which, we recognize, is the traditional Chi-square statistic in the problem of comparing a set of observed frequencies $(f_1, \cdots, f_k)$ with the corresponding set of

theoretical frequencies $(\mu_{10}, \cdots, \mu_{k0})$ when there is no constraint involved. It follows that if we take M as approximately distributed as χ_k^2, then

$$\Pr\{M < M_0 \mid \mathbf{f}\} \cong \Pr\{\chi_k^2 < \sum_{s=1}^{k} [(f_s - \mu_{s0})^2/\mu_{s0}]\}. \tag{3.18}$$

The complement of (3.18) is thus numerically equivalent to the significance level associated with the classical goodness of fit test. We are grateful to a referee for pointing out that the approximation (3.17) is in fact valid for *any* γ (and indeed for any prior) if we assume $f_s = \mu_{s0} + O(\mu_{s0}^{\frac{1}{2}})$ which is reasonable.

Finally, we remark here that if one adopts the "invariance" argument given by Jeffreys (1961) to take the prior distribution of μ_s as proportional to $\mu_s^{-\frac{1}{2}}$, i.e., $\gamma = -\frac{1}{2}$ in (3.2), the effect would be to change $m_s = f_s$ to $m_s = (f_s - \frac{1}{2})$. For a more detailed discussion of the prior distribution of Poisson parameters, see Jeffreys (1961).

4. The goodness of fit problem when the frequencies are subjected to the linear constraint $\sum f_s = n$. The above argument can be readily extended to the more common situation in which the variables $(f_1, \cdots, f_k)$ are subjected to the linear constraint $\sum f_s = n$. In this case, the likelihood function is

$$L(\boldsymbol{\mu} \mid \mathbf{f}, \textstyle\sum f_s = n) = L(\boldsymbol{\theta} \mid \mathbf{f}, \textstyle\sum f_s = n) \propto \prod_{s=1}^{k} \theta_s^{f_s} \tag{4.1}$$

where $\theta_s = \mu_s/\sum \mu_s$ and $\sum \theta_s = 1$, a result first given by Fisher. From (3.2), the prior distribution of $\boldsymbol{\theta}$ is then of the form

$$p(\boldsymbol{\theta}) \propto \prod_{s=1}^{k} \theta_s^{\gamma} \tag{4.2}$$

so that the posterior distribution of $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta} \mid \mathbf{f}, \textstyle\sum f_s = n) = \{\Gamma[\sum (m_s + 1)]/\prod_{s=1}^{k} \Gamma(m_s + 1)\} \prod_{s=1}^{k} \theta_s^{m_s} \tag{4.3}$$

which is a $(k - 1)$ dimensional multivariate beta distribution.

Adopting an argument similar to that given in the preceding section, we consider the quantity

$$M^* = -2 \log W^* \tag{4.4}$$

where

$$W^* = (\textstyle\sum m_s)^{\sum m_s} \prod_{s=1}^{k} (\theta_s/m_s)^{m_s}. \tag{4.5}$$

The cumulant generating function of M^* is readily found to be

$$\kappa_{M^*}(it) = a - \tfrac{1}{2}(k - 1) \log (1 - 2it) + \sum_{r=0}^{\infty} \alpha_r (1 - 2it)^{-r} \tag{4.6}$$

where

$$\alpha_r = [(-1)^r/r(r + 1)][(B_{r+1}(k)/(\textstyle\sum m_s)^r) - \sum_{s=1}^{k} (B_{r+1}(1)/m_s^r)]$$

and $B_r(x)$ is the Bernoulli polynomial of degree r. This is the same type of asymptotic series as that given in (3.10). As might be expected, the leading term corresponds to the cumulant generating function of a χ^2 variable with $(k - 1)$ degrees of freedom. Better approximations to the distribution of M^* can now be obtained by methods discussed before.

An important application of this result is of course the classical goodness of fit problem in which a sample of n observations are drawn from some population, the range of which is divided into k non-overlapping intervals. f_s is the number of observations in the sth interval and θ_s the probability that a single observation will fall in that interval. Thus we may be interested in the possibility that the observations are coming from a particular population H_0. This population H_0 gives rise to a specific set of values $\boldsymbol{\theta}_0 = (\theta_{10}, \cdots, \theta_{k0})$. We can then decide whether $\boldsymbol{\theta}_0$ lies inside or outside the $(1 - \alpha)$ H.P.D. region by calculating the probability

$$\Pr\{M^* < -2 \log W_0^*\} \tag{4.7}$$

with

$$W_0^* = (\sum m_s)^{\sum m_s} \prod_{s=1}^{k} (\theta_{s0}/m_s)^{m_s}. \tag{4.8}$$

We note that the quantity W_0^* is in exactly the same form as the likelihood ratio statistic in testing the hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. They are in fact identical if the prior distribution of $\boldsymbol{\theta}$ is uniform so that $m_s = f_s$. It follows that to order $(\sum m_s)^{-1}$, $-2 \log W_0^*$ can also be written

$$-2 \log W_0^* = \sum_{s=1}^{k} \frac{(m_s - \theta_{s0} \sum m_u)^2}{\theta_{s0} \sum m_u}. \tag{4.9}$$

In particular, when $m_s = f_s$ so that $\sum m_u = n$, the right hand side of (4.9) is recognized as K. Pearson's "Chi-square" statistic. As in the preceding section, we are again able to see the connection between the traditional sampling theory result and the Bayesian result.

5. Comparison of parameters. In the previous sections we have discussed problems of deciding whether a particular parameter point $\boldsymbol{\theta}_0$ is or is not included in the $(1 - \alpha)$ H.P.D. region. In much practical statistical work, we are often concerned with the comparative values of parameters rather than with the absolute values. Suppose in general we have k parameters $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$. We shall define $(k - 1)$ non-redundant comparisons as $(k - 1)$ independent functions

$$\phi_i = f_i(\boldsymbol{\theta}), \qquad i = 1, \cdots, (k - 1), \tag{5.1}$$

which are all equal to zero if and only if $\theta_1 = \cdots = \theta_k$. There is clearly a very wide range of choices of functions of this kind. Since the H.P.D. regions, like the confidence regions, are not invariant under non-linear transformation, some thought must be given as to how we parameterize such comparisons. Two of the most important problems in statistics are concerned with (a) comparison of location of distributions and (b) comparison of spread of distributions.

Comparison of location of k distributions. If we wish to compare k distributions which are identical except for location, we can do this in terms of any location parameter. A location parameter such as mean, median, quantile, etc., has the property that addition of a fixed constant to the observation produces a cor-

responding change in the parameter. Further, the location parameters of a distribution differ from each other by fixed constants. In comparing location of k distributions, it seems unreasonable not to require that the comparison functions should be independent of the choice of the location parameter. It follows that we must take the comparisons f_i as functions of linear contrasts of the k specific location parameters chosen. In the usual sense, a linear contrast in a set of parameters $\boldsymbol{\theta}$ is the function

$$l = \sum a_j\theta_j \qquad \text{where} \quad \sum a_j = 0. \tag{5.2}$$

The simplest such functions of linear contrasts are the contrasts themselves and the problem of comparing location will be expressed in these terms.

Comparison of spread of k distributions. In a similar way the spread of a distribution can be measured in a variety of ways in terms of standard deviation, variance, mean deviation, precision constant, etc. A parameter θ qualifies as a scale parameter if it has the property that a linear transformation of the observation from y to $(a + by)$ changes the parameter θ to $|b|^q\theta$. Further, if γ and θ are two scale parameters of a distribution then they are related by

$$\gamma = c\theta^\alpha. \tag{5.3}$$

In comparing spread of k distributions, it would be natural to require that the comparison functions are independent of the choice of the scale parameter. It follows that we must take $f_1, \cdots, f_{k-1}$ as functions of linear contrasts of the logarithms of the k scale parameters chosen. As before, the simplest functions to consider are the linear contrasts themselves so that our comparisons will be expressed in these terms.

It is, of course, accepted that other choices might be of interest and H.P.D. regions corresponding to these choices could be obtained by the method we give.

6. Comparison of location of k normal populations. We now return to the linear model discussed in Section 2. Consider the special case that the observations $\mathbf{y}$ are independent samples of size $n_1, \cdots, n_k (\sum n_s = n)$ from k normal populations with means $(\theta_1, \cdots, \theta_k)$ respectively and common variance σ^2. The posterior distribution of $\boldsymbol{\theta}$ in (2.2) reduces to

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto \{1 + [\sum n_i(\theta_i - \bar{y}_i)^2/\nu s^2]\}^{-\frac{1}{2}(\nu+k)} \tag{6.1}$$

with $\bar{y}_i = n_i^{-1} \sum_j y_{ij}$, $\nu = n - k$ and $s^2 = (n - k)^{-1} \sum \sum (y_{ij} - \bar{y}_i)^2$.

This distribution would then allow us to decide whether a particular set of values of the means $\boldsymbol{\theta}_0 = (\theta_{10}, \cdots, \theta_{k0})$ is or is not included in the $(1 - \alpha)$ H.P.D. region. In practice, however, we are frequently concerned with the problem of comparing the location of the k normal populations. Following the argument of the previous section, we are led to consider a set of $(k - 1)$ linearly independent contrasts in the k means $\theta_1, \cdots, \theta_k$,

$$\phi_i = \textstyle\sum_j a_{ij}\theta_j \qquad \text{where} \quad \sum_j a_{ij} = 0,\ i = 1, \cdots, k - 1. \tag{6.2}$$

Since the H.P.D. region is invariant under linear transformation, it is convenient

to consider the particular set

$$\phi_i = \theta_i - \bar{\theta} \qquad \text{where} \quad \bar{\theta} = (1/n) \sum n_i\theta_i\,, \; i = 1, \cdots, k-1. \tag{6.3}$$

It follows from the properties of the multivariate-t distribution that the posterior distribution of $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_{k-1})$ is

$$p(\boldsymbol{\phi} \mid \mathbf{y}) \propto \{1 + [\textstyle\sum^k n_i[\phi_i - (\bar{y}_i - \bar{y})]^2/\nu s^2]\}^{-\frac{1}{2}[\nu+(k-1)]} \tag{6.4}$$

where $\bar{y} = n^{-1} \sum n_i\bar{y}_i$, $\phi_k = \theta_k - \bar{\theta}$ and $\sum^k n_i\phi_i = 0$.

To decide whether a particular point $\boldsymbol{\phi}_0$ is or is not included in the $(1 - \alpha)$ H.P.D. region, we refer the quantity

$$f(\boldsymbol{\phi}_0) = \textstyle\sum^k n_i[\phi_{i0} - (\bar{y}_i - \bar{y})]^2/(k-1)s^2 \tag{6.5}$$

to $F_\alpha(k - 1, \nu)$, which is the upper $100\alpha\%$ point of an F distribution with $(k - 1, \nu)$ degrees of freedom. In particular, we may be interested to discover if $\boldsymbol{\phi}_0 = \mathbf{0}$ is so included. The point $\boldsymbol{\phi}_0 = \mathbf{0}$ corresponds to the situation where $\theta_1 = \cdots = \theta_k$ and is often of special concern in comparing location of distributions. In this case,

$$f(\mathbf{0}) = \textstyle\sum n_i(\bar{y}_i - \bar{y})^2/(k-1)s^2 \tag{6.6}$$

which is recognized as the usual F-statistic in the analysis of variance. It is hoped in later work to generalize this result to the situation in which the variances of the k normal populations are not necessarily equal.

7. Comparison of k scale parameters in a class of power distributions. We now consider the problem of comparing the spread of k populations. Because of the known lack of robustness to non-normality of tests to compare scale parameters, e.g. variances, due to confounding of variance inequality and kurtosis, it seems appropriate to consider a general class of parent populations with variable kurtosis. A convenient choice is the following three parameter family of power distributions

$$p(y \mid \theta, \beta, \sigma) = c(\beta, \sigma) \exp\{-\tfrac{1}{2}\,|(y - \theta)/\sigma|^{2/(1+\beta)}\} \qquad -\infty < y < \infty \tag{7.1}$$

with $c(\beta, \sigma) = \{\Gamma[1 + \frac{1}{2}(1 + \beta)]2^{1+\frac{1}{2}(1+\beta)}\sigma\}^{-1}$ and $-\infty < \theta < \infty$, $0 < \sigma < \infty$, $-1 < \beta < 1$.

In (7.1), θ is a location parameter and σ a scale parameter. When $\beta = 0$, the distribution is normal. Thus β can be regarded as a parameter measuring the departure from normality. This class of distributions was employed by the authors in studying the effect of non-normality on the inference about a location parameter (1962) and later on the inference about the equality of two scale parameters (1964). Our present result is a generalization of the later paper.

We shall assume that the k populations have the same parameter β, but possibly different values of θ and σ. At first, we suppose that β and the location parameter θ in each population are known. When independent samples of size $n_1, \cdots, n_k$ are drawn from the k populations, the likelihood function is

$$(7.2)\qquad l(\boldsymbol{\sigma} \mid \boldsymbol{\theta}, \beta, Y) = \prod_{i=1}^{k} \{c(\beta, \sigma_i)\}^{n_i} \exp\{-\tfrac{1}{2} n_i s_i(\beta, \theta_i)/\sigma^{2/(1+\beta)}\}$$

where

$$s_i(\beta, \theta_i) = n_i^{-1} \sum_{j=1}^{n_i} |y_{ij} - \theta_i|^{2/(1+\beta)}$$

$\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$, $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_k)$, $Y = (\mathbf{y}_1, \cdots, \mathbf{y}_k)$, and $\mathbf{y}_i = (y_{i1}, \cdots, y_{in_i})$, $i = 1, \cdots k$.

On the usual assumption that

$$(7.3)\qquad p(\log \sigma_i) \propto 1 \quad \text{or} \quad p(\sigma_i) \propto \sigma_i^{-1}, \qquad i = 1, \cdots, k,$$

the posterior distribution of $\boldsymbol{\sigma}$ is readily found to be

$$(7.4)\qquad p(\boldsymbol{\sigma} \mid \boldsymbol{\theta}, \beta, \mathbf{y}) = \prod_{i=1}^{k} p(\sigma_i \mid \theta_i, \beta, \mathbf{y}_i)$$

where

$$p(\sigma_i \mid \theta_i, \beta, \mathbf{y}_i) = d_i(\beta, \theta_i)\sigma_i^{-(n_i+1)} \exp\{-\tfrac{1}{2} n_i s_i(\beta, \theta_i)/\sigma_i^{2/(1+\beta)}\}$$

and

$$d_i(\beta, \theta_i) = n_i\{\tfrac{1}{2} n_i s_i(\beta, \theta_i)\}^{\frac{1}{2}n_i(1+\beta)}/\Gamma\{1 + \tfrac{1}{2} n_i(1 + \beta)\}.$$

This distribution is seen to be in the form of the product of k independent inverted gamma distributions. We are interested in comparing the equality of the spread of the k distributions. Following the argument in Section 5 and noting that the H.P.D. region is invariant under linear transformation, we consider the $(k - 1)$ linear contrasts in $\log \sigma_i$,

$$(7.5)\qquad \phi_i = [2/(1 + \beta)](\log \sigma - \log \sigma_i), \qquad i = 1, \cdots, k - 1.$$

It is straightforward to verify that the posterior distribution of $\boldsymbol{\phi}$ is

$$(7.6)\qquad p(\boldsymbol{\phi} \mid \boldsymbol{\theta}, \beta, Y) = \{\Gamma[\tfrac{1}{2}N(1 + \beta)]/\prod_{i=1}^{k} \Gamma[\tfrac{1}{2} n_i(1 + \beta)]\} T_1^{\frac{1}{2}n_1(1+\beta)} \cdots T_{k-1}^{\frac{1}{2}n_{k-1}(1+\beta)}(1 + T_1 + \cdots + T_{k-1})^{-\frac{1}{2}N(1+\beta)}$$

where $N = \sum_{j=1}^{k} n_j$ and $T_i = [n_i s_i(\beta, \theta_i)/n_k s_k(\beta, \theta_k)]e^{\phi_i}$, $i = 1, \cdots, (k - 1)$. In deciding if a particular point $\boldsymbol{\phi}_0$ is included in the H.P.D. region of content $(1 - \alpha)$, we calculate the probability

$$(7.7)\qquad \Pr\{p(\boldsymbol{\phi} \mid \boldsymbol{\theta}, \beta, Y) > p(\boldsymbol{\phi}_0 \mid \boldsymbol{\theta}, \beta, Y)\}.$$

Now, $p(\boldsymbol{\phi} \mid \boldsymbol{\theta}, \beta, Y)$ is a monotonic decreasing function of the quantity M where

$$(7.8)\qquad M = -2 \log W$$

with

$$(7.9)\qquad W = [N^{\frac{1}{2}N(1+\beta)}/\prod_{i=1}^{k} n_i^{\frac{1}{2}n_i(1+\beta)}](\prod_{i=1}^{k-1} T_i^{\frac{1}{2}n_i(1+\beta)}) \cdot (1 + T_1 + \cdots + T_{k-1})^{-\frac{1}{2}N(1+\beta)}$$

Making use of the properties of the Dirichlet integral, we find the characteristic

function of M,

$$(7.10)\quad E(e^{itM}) = \frac{N^{itN(1+\beta)/2}}{\prod_{s=1}^{k} n_s^{itn_s(1+\beta)/2}} \frac{\Gamma\left[\frac{N}{2}(1+\beta)\right]}{\prod_{s=1}^{k}\Gamma\left[\frac{n_s}{2}(1+\beta)\right]} \frac{\prod_{s=1}^{k}\Gamma\left[\frac{n_s}{2}(1+\beta)(1-2it)\right]}{\Gamma\left[\frac{N}{2}(1+\beta)(1-2it)\right]}$$

Taking logarithms and employing Stirling's series, we obtain for $\beta \neq -1$ the cumulant generating function of M,

$$(7.11)\quad \kappa_M(it) = a - \tfrac{1}{2}(k-1)\log(1-2it) + \textstyle\sum_{r=1}^{\infty} \alpha_r(1-2\mathrm{it})^{-(2r-1)}$$

where $\alpha_r = [B_{2r}/2r(2r-1)][2/(1+\beta)]^{2r-1}\{\sum_{i=1}^{\infty} n_i^{-(2r-1)} - N^{-(2r-1)}\}$. As in Sections 3 and 4, once again we are able to obtain an asymptotic χ^2 series for the distribution of the criterion. To decide if a point ϕ_0 is included in the $(1-\alpha)$ H.P.D. region, Expression (7.11) then allows us to evaluate the probability

$$(7.12)\qquad \Pr\{M < -2\log W_0\}$$

where W_0 is obtained by inserting ϕ_0 in (7.9). In particular, we may be interested in the point $\phi_0 = \mathbf{0}$ which corresponds to the situation $\sigma_1 = \sigma_2 = \cdots = \sigma_k$. In this case, $-2\log W_0$ reduces to

$$(7.13)\quad -2\log W_0 = -\textstyle\sum_{i=1}^{k} n_i(1+\beta)[\log s_i(\beta, \theta_i) - \log \bar{s}(\beta, \mathbf{\theta})]$$

where $\bar{s}(\beta, \mathbf{\theta}) = N^{-1}\sum_j n_j s_j(\beta, \theta_j)$.

It is interesting to note that the results in (7.11) and (7.13) correspond exactly to the results of the likelihood ratio test in the Neyman-Pearson framework. As mentioned in our earlier work (1964), in testing the hypothesis $H_0 : \sigma_1 = \cdots = \sigma_k$ against the alternative H_1 that they are not all equal, the likelihood ratio criterion is

$$(7.14)\qquad \lambda(\beta) = \textstyle\prod_{i=1}^{k} [s_i(\beta, \theta_i)/\bar{s}(\beta, \mathbf{\theta})]^{\frac{1}{2}n_i(1+\beta)}.$$

It is readily shown that the cumulant generating function of the sampling distribution of the quantity $-2\log\lambda(\beta)$ is precisely that given by the right hand side of (7.11). It follows that the complement of the probability (7.12) is numerically equivalent to the significance level associated with the *observed* likelihood ratio statistic $\lambda(\beta)$.

We should perhaps point out once more that this result does depend upon the particular parametrization we have used which, as explained in Section 5, seems to be reasonable. Alternative parametrizations, for example, the use of the $(k-1)$ ratios $\sigma_1/\sigma_k, \cdots, \sigma_{k-1}/\sigma_k$, would lead to a slightly different result. For other than very small samples, the difference would be negligible in practice.

8. The situation when $\theta_1, \cdots, \theta_k$ are not known. We now discuss the more common situation in which the location parameters $\theta_1, \cdots, \theta_k$ are not known. Including these parameters in our model, we shall make the usual assumption that they are distributed locally uniform *a priori*,

(8.1) $$p(\theta_i) \propto 1 \qquad i = 1, \cdots, k.$$

Following the results given in our earlier papers, the posterior distribution of $\boldsymbol{\theta}$ is

(8.2) $$p(\boldsymbol{\theta} \mid \beta, Y) = \prod_{i=1}^{k} p(\theta_i \mid \mathbf{y}_i)$$

with

$$p(\theta_i \mid \mathbf{y}_i) = \{\sum |y_{ij} - \theta_i|^{2/(1+\beta)}\}^{-\frac{1}{2}n_i(1+\beta)} / \int_{-\infty}^{\infty} \{\sum |y_{ij} - \theta_i|^{2/1+\beta}\}^{-\frac{1}{2}n_i(1+\beta)} d\theta_i .$$

Consequently, the posterior distribution of the $(k - 1)$ contrasts $\boldsymbol{\phi}$ defined in (7.5) becomes

(8.3) $$p(\boldsymbol{\phi} \mid \beta, Y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\boldsymbol{\phi} \mid \boldsymbol{\theta}, \beta, Y) p(\boldsymbol{\theta} \mid \beta, Y)\, d\theta_1 \cdots d\theta_k ,$$

where the first factor in the integrand is given by (7.6). In the special case $\beta = 0$, i.e., the parent populations are normal, it is readily verified that the above integral can be evaluated exactly yielding

(8.4) $$p(\boldsymbol{\phi} \mid \beta = 0, Y) = [\Gamma(\tfrac{1}{2}\nu) / \prod_{i=1}^{k} \Gamma(\tfrac{1}{2}\nu_i)] \alpha_1^{\frac{1}{2}\nu_1} \cdots \alpha_{k-1}^{\frac{1}{2}\nu_{k-1}} (1 + \alpha_1 + \cdots + \alpha_{k-1})^{-\frac{1}{2}\nu}$$

where $\nu_i = n_i - 1$, $\nu = N - k$, and $\alpha_i = [\sum^{n_i} (y_{ij} - \bar{y}_i)^2 / \sum^{n_k} (y_{kj} - \bar{y}_k)^2] e^{\phi_i}$.

Adopting the same argument as given in the preceding section, the cumulant generating function of the quantity

(8.5) $$M^*(0) = -2 \log W^*(0)$$

where

$$W^*(0) = [\nu^{\frac{1}{2}\nu} / \prod_{i=1}^{k} \nu_i^{\frac{1}{2}\nu_i}] \alpha_1^{\frac{1}{2}\nu_1} \cdots \alpha_{k-1}^{\frac{1}{2}\nu_{k-1}} (1 + \alpha_1 + \cdots + \alpha_{k-1})^{-\frac{1}{2}\nu}$$

is

(8.6) $$\kappa_{M^*(0)}(it) = a - \tfrac{1}{2}(k - 1) \log (1 - 2it) + \sum_{r=1}^{\infty} \alpha_r (1 - 2it)^{-(2r-1)}$$

where

$$\alpha_r = [B_{2r} / 2r(2r - 1)]\, 2^{2r-1} \{\sum_{i=1}^{k} \nu_i^{-(2r-1)} - \nu^{-(2r-1)}\}.$$

To decide if, say, $\phi_0 = \mathbf{0}$ is included in the $(1 - \alpha)$ H.P.D. region, we then calculate the probability

(8.7) $$\Pr \{M^*(0) < -2 \log W_0^*(0)\}$$

where

$$-2 \log W_0^*(0) = -\sum_{i=1}^{k} \nu_i [\log s_i^2 - \log s^2]$$

with $s_i^2 = \nu_i^{-1} \sum (y_{ij} - \bar{y}_i)^2$, $s^2 = \nu^{-1} \sum \nu_i s_i^2$. We recognize that the results in (8.6) and (8.7) correspond exactly to Bartlett's modified form of the likelihood ratio test of the equality of k normal variances when the means are not assumed known.

In the more general situation when the parent populations are not necessarily

normal, $\beta \neq 0$, it does not seem possible to express the integral in (8.3) exactly in terms of simple functions. However, it was demonstrated in our paper (1964) in dealing with the ratio of two scale parameters that the effect of integrating over the posterior distribution of (θ_1, θ_2) is essentially to replace the θ's by their corresponding modal values and to reduce the "degrees of freedom" $n_i(1 + \beta)$ by one unit. This is of course exact in the case $\beta = 0$. Extending this argument, the posterior distribution $p(\phi \mid \beta, Y)$ is then approximately

$$(8.8)\quad p(\phi \mid \beta, Y) \cong [\Gamma(\tfrac{1}{2}m)/\textstyle\prod_{i=1}^{k} \Gamma(\tfrac{1}{2}m_i)]\gamma_1^{\frac{1}{2}m_1} \cdots \gamma_{k-1}^{\frac{1}{2}m_{k-1}}(1 + \gamma_1 + \cdots + \gamma_{k-1})^{-\frac{1}{2}m}$$

where $m_i = n_i(1 + \beta) - 1$, $m = N(1 + \beta) - k$, $\gamma_i = [n_i s_i(\beta, \hat{\theta}_i)/n_k s_k(\beta, \hat{\theta}_k)]e^{\phi_i}$ and $\hat{\theta}_i$ is the mode of the posterior distribution $p(\theta_i \mid \mathbf{y}_i)$ in (8.2). Consequently to this degree of approximation, the cumulant generating function of

$$(8.9)\qquad M^*(\beta) = -2 \log W^*(\beta)$$

where

$$W^*(\beta) = (m^{\frac{1}{2}m}/\textstyle\prod_{i=1}^{k} m_i^{\frac{1}{2}m_i})\gamma_1^{\frac{1}{2}m_1} \cdots \gamma_{k-1}^{\frac{1}{2}m_{k-1}}(1 + \gamma_1 + \cdots + \gamma_{k-1})^{-\frac{1}{2}m}$$

is given by

$$(8.10)\quad \kappa_{M^*(\beta)}(it) = a - \tfrac{1}{2}(k - 1) \log (1 - 2it) + \textstyle\sum_{r=1}^{\infty} \alpha_r (1 - 2it)^{-(2r-1)}$$

where $\alpha_r = [B_{2r}/2r(2r - 1)]\, 2^{2r-1}\{\sum_{i=1}^{k} m_i^{-(2r-1)} - m^{-(2r-1)}\}$.

Hence the distribution of $M^*(\beta)$ can, as before, be expressed as a χ^2 series. The required probability in deciding whether $\phi_0 = \mathbf{0}$ is included in the H.P.D. region of content $(1 - \alpha)$ is

$$(8.11)\qquad \Pr\{M^*(\beta) > -2 \log W_0^*(\beta)\}$$

where

$$-2 \log W_0^*(\beta) = -\textstyle\sum_{i=1}^{k} m_i[\log (n_i s_i(\beta, \hat{\theta}_i)/m_i) - \log \bar{s}(\beta, \hat{\boldsymbol{\theta}})]$$

with $\bar{s}(\beta, \hat{\boldsymbol{\theta}}) = m^{-1}\sum_{i=1}^{k} n_i s_i(\beta, \hat{\theta}_i)$.

In the case $\beta = 0$, expressions (8.10) and (8.11) reduce to (8.6) and (8.7) respectively. The methods discussed in Section 3 can be employed to approximate the distribution of $M^*(\beta)$. Thus, we arrive at the somewhat remarkable result that for *any* known value of β (not close to -1), the decision as to whether the point corresponding to $\sigma_1 = \cdots = \sigma_k$ lies inside or outside the $(1 - \alpha)$ H.P.D. region is made by referring $M_0^*(\beta)$ to a scaled χ^2 distribution. The quantity $M_0^*(\beta)$ is in exactly the same form as Bartlett's modified form of the likelihood ratio statistic for the case $\beta = 0$, except that n_i is replaced by $n_i(1 + \beta)$ and the sample variances s_i^2, by the quantity $[\sum |y_{ij} - \hat{\theta}_i|^{2/(1+\beta)}]/(n_i(1 + \beta) - 1)$.

In this and the preceding sections, we are able to obtain, for each β, a criterion for the inferential problem of comparing the spread of k distributions, both for known and unknown $(\theta_1, \cdots, \theta_k)$. For a given set of data, this class of criteria

could then allow us to study how our inference about equality of the scale parameters may be affected by the departure from normality in the parent populations. In some cases, $\beta = 0$ and/or $\boldsymbol{\theta}$ known, corresponding results can be obtained from the sampling theory point of view. But in the more general situation when $\beta \neq 0$ and the $\boldsymbol{\theta}$ unknown, no sampling result is available.

Finally, when the parameter β is included in the model as a variable parameter, we can define the linear contrasts ϕ as

$$\phi_i = \log \sigma_k - \log \sigma_i, \qquad i = 1, \cdots, k-1,$$

and obtain their joint posterior distribution after eliminating $\boldsymbol{\theta}$ and β. The resulting distribution is complicated and it is hoped in later work to consider this problem further.

9. Acknowledgment. We are grateful to Professor E. S. Pearson for remarks which led to much of the work described in this paper.

REFERENCES

BARTLETT, M. S. (1937). Properties of sufficiency and statistical test. *Proc. Roy. Soc. Ser. A* **160** 268–282.

BOX, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36** 317–346.

BOX, G. E. P. and TIAO, G. C. (1962). A further look at robustness via Bayes' theorem. *Biometrika* **49** 419–432.

BOX, G. E. P. and TIAO, G. C. (1964). A Bayesian approach to the importance of assumptions applied to the comparison of variances. *Biometrika* **51** 153–167.

CORNISH, E. A. (1954). The multivariate t-distribution associated with a set of normal sample deviates. *Austral. J. Phys.* **7** 531–542.

DUNNETT, C. W. and SOBEL, M. (1954). A bivariate generalization of Student's t-distribution, with tables for certain special cases. *Biometrika* **41** 153–169.

HARTLEY, H. O. (1940). Testing the homogeneity of a set of variances. *Biometrika* **31** 249–255.

HILDRETH, C. (1963). Bayesian statisticians and remote clients. *Econometrika* **32** 422–438.

JEFFREYS, H. (1961). *Theory of Probability* (3rd ed.). Clarendon Press, Oxford.

LINDLEY, D. V. (1964). The Bayesian analysis of contingency tables. *Ann. Math. Statist.* **35** 1622–1643.