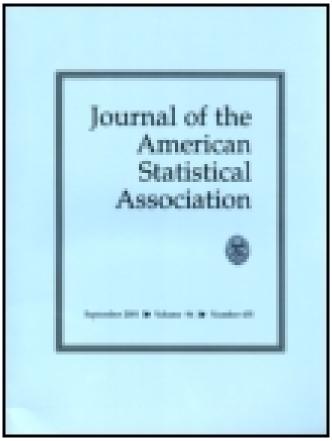

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/uasa20

Inference about Variance Components in the One-Way Model

Bruce M. Hill [a b]
[a] The University of Michigan , USA
[b] Harvard University , USA
Published online: 10 Apr 2012.

To cite this article: Bruce M. Hill (1965) Inference about Variance Components in the One-Way Model, Journal of the American Statistical Association, 60:311, 806-825

To link to this article: http://dx.doi.org/10.1080/01621459.1965.10480829

INFERENCE ABOUT VARIANCE COMPONENTS IN THE ONE-WAY MODEL*

BRUCE M. HILL
The University of Michigan and Harvard University

The estimation of variance components in the one-way model is considered from a subjective Bayesian point of view. The situation in which the classical unbiased estimate of the between variance component is negative is explored in some detail. Exact and approximate posterior distributions are obtained in both the balanced and unbalanced case. Common sense aspects of the problem are emphasized, and some contrasts with other approaches. For example, Bayesianly speaking, a large negative unbiased estimate of the between variance component indicates an uninformative experiment in which the effective likelihood for that variance component is extremely flat, instead of strong evidence that the variance component is nearly zero.

1. INTRODUCTION

WE SHALL discuss some aspects of the estimation of variance components in a random model (Eisenhart Model II) analysis of variance from a subjective Bayesian point of view. Particular attention will be given to the situation in which the classical unbiased estimate is negative. In this paper we shall be concerned with the one-way model, balanced and unbalanced, extending the results to regression models and to higher dimensional arrays in a later report.

The analysis of variance opens a Pandora's box of problems which constitute a real challenge to any and all statisticians and theories of statistical inference. Although an enormous variety of models have been proposed for even the most simple arrays of observations, there are serious gaps in our understanding of the *analysis* of even the most simple of these models, not to mention design. Here the approach will be to focus attention on an extremely simple model, a model that can at best be regarded only as a rough approximation in situations of real interest. Nonetheless, if only because the real situations we are confronted with are so very much more complex, it seems essential to look closely at the simpler ones.

We work with the model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, I, \quad j = 1, \cdots, J_i, \quad N = \sum_i J_i,$$

where $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\epsilon_{ij} \sim N(0, \sigma^2)$, and the α_i and ϵ_{ij} are mutually independent. Although we shall return to this general unbalanced model it is convenient for the present to restrict attention to the balance case, $J_i = J$ for all i. In this case the statistics

$$\bar{y}_{..} = (IJ)^{-1} \sum_{i,j} y_{ij}, \qquad SSW = \sum_{i,j} (y_{ij} - \bar{y}_{i.})^2,$$

* This work was partially sponsored by the National Science Foundation, Grant #NSF-GP-12.

and

$$SSB = \sum_{i,j} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

are sufficient for $\theta = (\mu, \sigma^2, \sigma_\alpha^2)$, where

$$\bar{y}_{i\cdot} = (J)^{-1} \sum_{j=1}^{J} y_{ij}.$$

It is easily verified that $\bar{y}_{\cdot\cdot}$ is normally distributed about μ with variance $(IJ)^{-1}\sigma^2 + (I)^{-1}\sigma_\alpha^2$, and that SSW and SSB are distributed as multiples of chi-square random variables,

$$\sigma^2 \chi^2_{I(J-1)} \quad \text{and} \quad (\sigma^2 + J\sigma_\alpha^2)\chi^2_{(I-1)},$$

respectively, the three being mutually independent given μ, σ^2 and σ_α^2.

The customary approach to estimation is to view $\bar{y}_{\cdot\cdot}$ as an estimate of μ, $MSW = [I(J-1)]^{-1}\ SSW$ as an estimate of σ^2, $MSB = [I-1]^{-1}\ SSB$ as an estimate of $\sigma^2 + J\sigma_\alpha^2$, and $(J)^{-1}\ [MSB - MSW]$ as an estimate of σ_α^2, all being unbiased. However, the possibility of negative estimates of σ_α^2 (which occur with substantial probability when σ_α^2 is small) and the lack of exact and confusion about proposed approximate confidence intervals for σ_α^2 (even in the balanced case) raise serious questions [3]–[7], [9]–[15], [18], [20], [24]–[26], [29]–[30].

The aim of this paper is to illustrate Bayesian inference in this problem, and to use the Bayesian formulation as a means to discuss certain common sense aspects of the situation. Since there is some disagreement as to just what is common sense here, such discussion does not seem out of place.

2. LIKELIHOOD FUNCTION AND POSTERIOR DISTRIBUTIONS

Based upon the above model, and writing Pr for what is in fact a density function, the likelihood function is

$$L(\mu, \sigma^2, \sigma_\alpha^2) = L(\mu, \sigma^2, \sigma_\alpha^2 \mid \text{data } y_{ij}) = \Pr\{y_{ij} \mid \mu, \sigma^2, \sigma_\alpha^2\}$$

$$= \int \cdots \int \Pr\{y_{ij} \mid \mu, \sigma^2, \sigma_\alpha^2, \alpha_i\} \Pr\{\alpha_i \mid \mu, \sigma^2, \sigma_\alpha^2\} \prod d\alpha_i$$

$$\propto \int \cdots \int (\sigma^2)^{-N/2} (\sigma_\alpha^2)^{-I/2}$$

$$\cdot \exp\left[-\frac{1}{2} \left\{ \sum_{i,j} (y_{ij} - \mu - \alpha_i)^2/\sigma^2 + \sum_i \alpha_i^2/\sigma \right\}_\alpha^2 \right] \prod d\alpha_i \tag{1}$$

$$\propto (\sigma^2)^{-(N-I)/2} \exp\left[-\frac{1}{2} \frac{SSW}{\sigma^2} \right] \prod_i (\sigma^2 + J_i\sigma_\alpha^2)^{-1/2} \exp\left[-\frac{1}{2} \sum_i \frac{J_i(\bar{y}_i - \mu)^2}{\sigma^2 + J_i\sigma_\alpha^2} \right].$$

Letting $\rho(\mu, \sigma^2, \sigma_\alpha^2)$ be a subjectively chosen prior density, the posterior density becomes $\rho''(\mu, \sigma^2, \sigma_\alpha^2) \propto L(\mu, \sigma^2, \sigma_\alpha^2)\ \rho(\mu, \sigma^2, \sigma_\alpha^2)$.

We shall be interested both in the situation where the likelihood function is sharp or highly concentrated relative to the prior distribution, in which case (some) Bayesian credible intervals typically differ but little from (some) classical confidence intervals, and also in those situations where although the likelihood function is expected to be relatively sharp (on the basis, say, of Fisherian information) before the experiment, it in actual fact is not, as for example when $MSB \ll MSW$. The following references discuss and illustrate Bayesian inference [1], [2], [8], [16], [17], [19], [21], [22], [23], [28].

All of our analysis will be based on the assumption of diffuse prior opinion for μ effectively independent of that for $(\sigma^2, \sigma_\alpha^2)$, or roughly $p(\mu, \sigma^2, \sigma_\alpha^2) = p(\sigma^2, \sigma_\alpha^2)$. This is not especially restrictive, and leads to the approximate marginal posterior density

$$
\begin{aligned}
p''(\sigma^2, \sigma_\alpha^2) &= \int_{-\infty}^{\infty} p''(\mu, \sigma^2, \sigma_\alpha^2)\, d\mu \\
&\propto p(\sigma^2, \sigma_\alpha^2)(\sigma^2)^{-(N-I)/2} \\
&\quad \cdot \exp\left[-\frac{SSW}{2\sigma^2}\right] \prod_{i=1}^{I} (\sigma^2 + J_i\sigma_\alpha^2)^{-1/2} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\sum_i \frac{J_i(\bar{y}_{i\cdot} - \mu)^2}{\sigma^2 + J_i\sigma_\alpha^2}\right] d\mu \\
&\propto p(\sigma^2, \sigma_\alpha^2)(\sigma^2)^{-(N-I)/2} \\
&\quad \cdot \exp\left[-\frac{SSW}{2\sigma^2}\right] \prod_{i=1}^{I} (\sigma^2 + J_i\sigma_\alpha^2)^{-1/2} \left\{\sum_{i=1}^{I} \frac{J_i}{\sigma^2 + J_i\sigma_\alpha}\right\}^{-1/2} \\
&\quad \cdot \exp\left[-\frac{1}{2}\sum_i \frac{J_i(\bar{y}_{i\cdot} - \hat{\mu})^2}{\sigma^2 + J_i\sigma_\alpha^2}\right]
\end{aligned} \tag{2}
$$

where

$$
\hat{\mu} = \frac{\sum_i J_i \bar{y}_{i\cdot}/(\sigma^2 + J_i\sigma_\alpha^2)}{\sum_i J_i/(\sigma^2 + J_i\sigma_\alpha^2)}
$$

is a function of σ_α^2/σ^2. In the balanced case $J_i = J$ this simplifies to $\hat{\mu} = \bar{y}_{\cdot\cdot}$ and

$$
\begin{aligned}
p''(\sigma^2, \sigma_\alpha^2) \propto p(\sigma^2, \sigma_\alpha^2)(\sigma^2)^{-(N-I)/2} \exp\left[-\frac{SSW}{2\sigma^2}\right] (\sigma^2 + J\sigma_\alpha^2)^{-(I-1)/2} \\
\cdot \exp\left[-\frac{SSB}{2(\sigma^2 + J\sigma_\alpha^2)}\right].
\end{aligned} \tag{3}
$$

It is to be noted that equation (3) expresses the fact that independent measurements have been made of σ^2 and $\sigma^2 + J\sigma_\alpha^2$, and that the problem of inference about σ^2 and σ_α^2 is thus closely related to other problems, for example, inference about background and source radioactivity based upon independent readings with and without the source present. Thus if Y_1 = counts without source, Y_2 = counts with source, $E(Y_1) = \lambda_1$, $E(Y_2) = \lambda_1 + \lambda_2$, and Y_1 and Y_2 have independent Poisson distributions, then

$$\rho''(\lambda_1, \lambda_2)/\rho(\lambda_1, \lambda_2)$$
$$= \Pr\{Y_1, Y_2 \mid \lambda_1, \lambda_2\} \propto \lambda_1^{Y_1} \exp[-\lambda_1](\lambda_1 + \lambda_2)^{Y_2} \exp[-(\lambda_1 + \lambda_2)].$$

This is essentially equivalent to (3) as far as the likelihood function is concerned, although there is an interesting difference between inference about λ_2 and inference about σ_α^2 which we will not go into here.

3. BALANCED CASE

A. *Inference on* σ^2

The marginal posterior density for σ^2 is

$$\rho''(\sigma^2) = \int_0^\infty \rho''(\sigma^2, \sigma_\alpha^2) d\sigma_\alpha^2$$
$$\propto (\sigma^2)^{-I(J-1)/2} \exp\left[-\frac{SSW}{2\sigma^2}\right] \int_0^\infty \rho(\sigma^2, \sigma_\alpha^2)(\sigma^2 + J\alpha_\alpha^2)^{-(I-1)/2}$$
$$\cdot \exp\left[-\frac{SSB}{2(\sigma^2 + J\sigma_\alpha^2)}\right] d\sigma_\alpha^2 .$$

Now it is instructive to view the factor

$$\rho'(\sigma^2) \equiv \int_0^\infty \rho(\sigma^2, \sigma_\alpha^2)(\sigma^2 + J\sigma_\alpha^2)^{-(I-1)/2} \exp\left[-\frac{SSB}{2(\sigma^2 + J\sigma_\alpha)}\right] d\sigma_\alpha^2 \tag{4}$$

as a partial posterior distribution based solely on the data $\bar{y}_{i\cdot}$, and the factor

$$L(\sigma^2 \mid SSW) \equiv (\sigma^2)^{-I(J-1)/2} \exp\left[-\frac{SSW}{2\sigma^2}\right]$$

as a likelihood function based solely on the deviations $y_{ij} - \bar{y}_{i\cdot}$. The over-all posterior density is thus $\rho''(\sigma^2) \propto \rho'(\sigma^2) L(\sigma^2 \mid SSW)$. The reason for this factorization is to focus attention on the information about σ^2 contained in the $\bar{y}_{i,}$, which seems often to be overlooked.

The subtleties involved in inference on σ^2 stem from the factor $\rho'(\sigma^2)$. When $\rho'(\sigma^2)$ can be regarded as gentle relative to $L(\sigma^2 \mid SSW)$ the principle of stable measurement applies and $\rho''(\sigma^2)$ will be nearly proportional to $L(\sigma^2 \mid SSW)$ [8, p. 15], [16], [19, Ch. 3], [22, p. 20], [23, Ch. 4]. For example, if $\rho'(\sigma^2) \alpha \sigma^{-2}$ (strictly speaking this is impossible, but when stable estimation occurs the exact choice of $\rho'(\sigma^2)$ is immaterial, as is discussed further at the end of this section) then

$$\rho''(\sigma^2) \propto (\sigma^2)^{-(I(J-1)/2)-1} \exp\left[-\frac{SSW}{2\sigma^2}\right] \tag{5}$$

so that σ^2 is distributed approximately like $SSW/\chi^2_{I(J-1)}$ where χ^2_f denotes a random variable having the chi-square distribution with f degrees of freedom. This result is in harmony with the conventional approach under which SSW/σ^2 has the chi-square distribution with $I(J-1)$ degrees of freedom for given σ^2. From a Bayesian viewpoint, however, this "stable measurement posterior" is merely one (rather simple) posterior distribution that may arise, and we cannot regard it as having any innate significance, except at best as a reasonable approximation in some circumstances. Indeed, the factor $\rho'(\sigma^2)$ may not be gentle relative to $L(\sigma^2|SSW)$ even though sample sizes are large and $L(\sigma^2|SSW)$ is quite sharp, and we must be extremely cautious in invoking the stable measurement approximation.

Recall now that $\rho'(\sigma^2)$ represents the posterior density of σ^2 arising from measurement of $h^2=\sigma^2+J\sigma_\alpha^2$ (based only on the $\bar{y}_i$) in conjunction with the prior joint density $\rho(\sigma^2, \sigma_\alpha^2)$. It is clear that precise measurement of h^2 can lead to sharp upper bounds for tail probabilities of the form $\Pr'\{\sigma^2>C\}$ and $\Pr'\{\sigma_\alpha^2>C\}$ (where the single prime indicates posterior probability based only on the data $\bar{y}_{i.}$) since each is no larger than $\Pr'\{h^2>C\}$. On the other hand, within the set of $(\sigma^2, \sigma_\alpha^2)$ for which $\sigma^2+J\sigma_\alpha^2$ is in reasonable accord with our assumed precise measurement of h^2, the partial posterior density $\rho'(\sigma^2, \sigma_\alpha^2)$ based solely on the $\bar{y}_{i.}$ will be gentle *unless* our initial prior opinion as measured by $\rho(\sigma^2, \sigma_\alpha^2)$ is sharp. It is as though the precise measurement of h^2 tends to sharply restrict the space of $(\sigma^2, \sigma_\alpha^2)$ to a set in which $\sigma^2+J\sigma_\alpha^2$ is nearly constant and within that set preference is effectively governed by $\rho(\sigma^2, \sigma_\alpha^2)$.

To illustrate, let us suppose that measurement of h^2 is so sharp that we effectively determine a line $\sigma^2+J\sigma_\alpha^2=h^2=\hat{h}^2$ having probability unity in the $(\sigma^2, \sigma_\alpha^2)$ plane. Then the joint density of $(\sigma^2, \sigma_\alpha^2)$ given the data $\bar{y}_{i.}$ is

$$\rho'(\sigma^2, \sigma_\alpha^2) \propto \begin{cases} \rho(\sigma^2, \sigma_\alpha) & \text{if } \sigma^2 + J\sigma_\alpha^2 = \hat{h}^2, \\ 0 & \text{if not} \end{cases}$$

and

$$\rho'(\sigma^2) \propto \begin{cases} \rho(\sigma^2, J^{-1}(\hat{h}^2 - \sigma^2)) & \text{if } 0 \le \sigma^2 \le \hat{h}^2, \\ 0 & \text{if } \sigma^2 > \hat{h}^2 \end{cases}$$

Unless there is some rather special prior opinion, such as for example that it is highly probable that σ_α^2 is close to some specified value or that $\sigma_\alpha^2<\sigma^2$, etc., one would anticipate that in many situations a rather typical prior opinion would have $\rho(\sigma^2, \sigma_\alpha^2)$ varying gently along the line $\sigma^2+J\sigma_\alpha^2=\hat{h}^2$. Thus it is natural to expect $\rho'(\sigma^2)$ to be nearly constant in the interval $0\le\sigma^2\le\hat{h}^2$, and similarly with respect to $\rho'(\sigma_\alpha^2)\propto\rho(\hat{h}^2-J\sigma_\alpha^2, \sigma_\alpha^2)$ for $0\le\sigma_\alpha^2\le\hat{h}^2/J$. Except in the face of rather special prior knowledge, the data $\bar{y}_{i.}$ for large I will lead one to rather precise opinions about $h^2=\sigma^2+J\sigma_\alpha^2$, but one will be vague about σ^2 and σ_α^2 outside of the fact that one will be nearly certain the first is not much larger than $\hat{h}^2$ and the second than $\hat{h}^2/J$.

Our discussion so far has been quite symmetric with respect to σ^2 and $J\sigma_\alpha^2$. An asymmetry is now introduced by bringing in the likelihood $L(\sigma^2|SSW)$ based on the deviations $y_{ij}-\bar{y}_{i.}$. This function tends to point sharply to a value $\hat{\sigma}^2$. If $\hat{\sigma}^2\ll\hat{h}^2$ and if $\rho(\sigma^2, \sigma_\alpha^2)$ is gentle along the line $\sigma^2+J\sigma_\alpha^2=\hat{h}^2$ then the over-all posterior $\rho''(\sigma^2)\propto\rho'(\sigma^2)L(\sigma^2|SSW)$ will be nearly proportional to $L(\sigma^2|SSW)$ within $[0, \hat{h}^2]$, and except for the truncation effect beyond $\hat{h}^2$ [which will be far out in the upper tail of $L(\sigma^2|SSW)$] we will have essentially the stable estimation posterior above (5). On the other hand, if $\hat{\sigma}^2\gg\hat{h}^2$, then

$\rho'(\sigma^2)$ and $L(\sigma^2 | SSW)$ tend to cancel one another in the sense that each is large where the other is small. Since $L(\sigma^2 | SSW)$ points roughly to MSW, while the measurement of h^2 points roughly to MSB, we see that the condition for such "cancellation" is roughly that $MSW \gg MSB$.

We shall now assume $\rho(\sigma^2, \sigma_\alpha^2) = \rho_1(\sigma^2)\rho_2(\sigma_\alpha^2)$. This certainly need not be the case, but, for example, if σ^2 reflects measurement error on an individual, while σ_α^2 reflects variability of "true values" between individuals then it is not unnatural that opinions about σ^2 and σ_α^2 should be roughly independent. We find

$$\begin{aligned}
\rho''(\sigma^2) &\equiv \int_0^\infty \rho''(\sigma^2, \sigma_\alpha^2)\, d\sigma_\alpha^2 \\
&\propto \int_0^\infty \rho(\sigma^2, \sigma_\alpha^2) L(\sigma^2 | SSW)(h^2)^{-(I-1)/2} \exp\left[-\frac{SSB}{2h^2}\right] d\sigma_\alpha^2 \\
&\propto \rho_1(\sigma^2) L(\sigma^2 | SSW) \int_0^\infty \rho_2(\sigma_\alpha^2)(\sigma^2 + J\sigma_\alpha^2)^{-(I-1)/2} \exp\left[-\frac{SSB}{2(\sigma^2 + J\sigma_\alpha^2)}\right] d\sigma_\alpha^2.
\end{aligned}$$

It is tempting at this point to take $\rho_2(\sigma_\alpha^2) \propto (\sigma_\alpha^2)^{-1}$ along the lines of Savage [23, p. 5.7] and others; however, the integral is then infinite because of the pole at zero. This is no real problem since the prior density $\rho_2(\sigma_\alpha^2) \propto (\sigma_\alpha^2)^{-1}$ would always be regarded as unrealistic near zero, and only be employed when this feature was not crucial in determining the posterior distribution. To avoid trouble near zero we shall give σ_α^2 a proper prior distribution, taking

$$\rho_2(\sigma_\alpha^2) \propto (\sigma_\alpha^2)^{-((\lambda_\alpha/2)+1)} \exp\left[-\frac{C_\alpha}{2\sigma_\alpha}\right],$$

so that $(\sigma_\alpha^2)^{-1}$ has a gamma distribution with parameters $\lambda_\alpha/2$ and $C_\alpha/2$ chosen subjectively. Then

$$\begin{aligned}
\rho''(\sigma^2) &\propto \rho_1(\sigma^2) L(\sigma^2 | SSW) \int_0^\infty (\sigma_\alpha^2)^{-(\lambda_\alpha/2+1)} \exp\left[-\frac{C_\alpha}{2\sigma_\alpha^2}\right] (h^2)^{-((I-1)/2)} \\
&\qquad \cdot \exp\left[-\frac{SSB}{2h^2}\right] d\sigma_\alpha^2 \\
&\propto \rho_1(\sigma^2) L(\sigma^2 | SSW) \int_0^\infty y^{-((\lambda_\alpha/2)+1)} \exp\left[-\frac{JC_\alpha}{2y}\right] (\sigma^2 + y)^{-((I-1)/2)} \\
&\qquad \cdot \exp\left[-\frac{SSB}{2(\sigma^2 + y)}\right] dy, \qquad (6)
\end{aligned}$$

and the last integral may be regarded as proportional to the density of a random variable.

$$Z = X - Y = \left\{(\Gamma_{(I-3)/2, SSB/2})^{-1} - (\Gamma_{\lambda_\alpha/2, JC_\alpha/2})^{-1}\right\},$$

at least on the positive half-line, where $\Gamma_{\lambda,\theta}$ is a random variable having the density

$$f(t) \propto t^{\lambda-1}e^{-\theta t}, \qquad t \geq 0,$$

and X and Y are independent. If $C_\alpha = K_{\alpha\lambda\alpha}$, then

$$\lim_{\lambda_\alpha \to 0} (\sigma_\alpha^2)^{-(\lambda_\alpha/2+1)} \exp\left[-\frac{C_\alpha}{2\sigma_\alpha^2}\right] = (\sigma_\alpha^2)^{-1};$$

so we may represent diffuse prior opinion by a choice of λ_α near zero. However, the expression for $\rho(\sigma_\alpha^2)$, with suitable λ_α and C_α, can be employed whether or not prior opinion about σ_α^2 is diffuse.

The over-all posterior distribution of σ^2 thus may be regarded as a weighting of $\rho_1(\sigma^2)L(\sigma^2 | SSW)$ by the density of Z, the resulting function being normalized to have unit area on the positive half-line. Although Z can clearly be negative, only the portion of its density for $\sigma^2 \geq 0$ is relevant to the over-all posterior distribution of σ^2. Since both the mean and mode of Z are roughly $MSB - JC_\alpha/\lambda_\alpha$, this quantity locates the center of the distribution of Z. When it is negative, then the density of Z will be monotonically decreasing for $\sigma^2 \geq 0$, and the effect of the information supplied by the $\bar{y}_{i.}$ is to modify the partial posterior $\rho_1(\sigma^2)$ $L(\sigma^2 | SSW)$ based upon the $y_{ij} - \bar{y}_{i.}$ alone by giving the greatest weight to small σ^2, and thus tending to move opinions downward. When $MSB - JC_\alpha/\lambda_\alpha > 0$ then the density of Z will have only a portion of its left tail below zero, and there is a possibility of obtaining an extremely sharp posterior $\rho''(\sigma^2)$.

Now it is clear that a great variety of situations are possible, and that the posterior distributions are in general analytically cumbersome. Although there are some natural approximations, and one such will be described in Section 3B where the distributional form of Z again arises in connection with σ_α^2, the point of view taken here is that such approximations need only play a minor role in the *analysis* of data. For it is ordinarily a minor computing task to plot the posterior distributions corresponding to one or perhaps several choices of prior. After such plotting an approximation may be chosen in order to simplify the description of the posterior, but this is not essential. The real need for approximations in Bayesian theory is in the planning of an experiment, for example in the choice of sample sizes, which we shall not discuss in this paper.

We shall consider one other form of diffuse or gentle prior, namely $\rho(\sigma^2, \sigma_\alpha^2) \propto (\sigma^2)^{-1}(\sigma^2 + J\sigma_\alpha^2)^{-1}$ for $\sigma^2 \geq 0$, $\sigma_\alpha^2 \geq 0$, which as pointed out by Tiao and Tan [28] can be obtained by the method of Jeffreys. Although the J in this prior depends upon the sample size and would thus seem inappropriate, our point of view is that this prior is simply a convenient analytical form only to be used when in fact it makes little difference what the exact choice of prior is because of the sharpness of the likelihood function. We find for this prior

$$\begin{aligned}
\rho''(\sigma^2) &\propto \int_0^\infty (\sigma^2)^{-1}(\sigma^2 + J\sigma_\alpha^2)^{-1} L(\sigma^2 \mid SSW)(h^2)^{-(I-1)/2} \exp\left[-\frac{SSB}{2h^2}\right] d\sigma_\alpha^2 \\
&\propto (\sigma^2)^{-1} L(\sigma^2 \mid SSW) \int_0^\infty (h^2)^{-((I-1)/2)-1} \exp\left[-\frac{SSB}{2h^2}\right] d\sigma_\alpha^2 \\
&\propto (\sigma^2)^{-1} L(\sigma^2 \mid SSW) F_{I-1}(SSB/\sigma^2), \qquad (7)
\end{aligned}$$

where $F_n(t)$ is the cumulative distribution function of chi-square with n degrees of freedom. Note that this differs from the earlier stable estimation posterior only in the factor $F_{I-1}(SSB/\sigma^2)$, which is a monotonically decreasing function of σ^2, concave for $\sigma^2 < SSB/(I+1)$ and convex for $\sigma^2 > SSB/(I+1)$.

If in (7) we put $SSB=0$, then

$$\rho''(\sigma^2) \propto (\sigma^2)^{-(IJ-1)/2-1} \exp\left\{-\frac{SSW}{2\sigma^2}\right\},$$

so that posteriorly $\sigma^2 \sim SSW/\chi^2_{(IJ-1)}$, as was noted by Tiao and Tan [28]. This is a disturbing result. For if I is large then an SSB near zero would, based upon the $\bar{y}_{i.}$ alone, lead to a strong opinion that σ^2 is "small." On the other hand, a large SSW would, based upon the $y_{ij}-\bar{y}_{i.}$ alone, lead to a strong opinion that σ^2 is "large." These two separate sources of information can thus lead to very different opinions about σ^2. Surely in the extreme case where SSB is very small and SSW is very large one is reluctant to accept $\sigma^2 \sim SSW/\chi^2_{(IJ-1)}$ as describing over-all posterior opinion about σ^2. It should be pointed out that it is not simply this result (which after all was based upon an improper prior) that is in question, but really any way of combining the two divergent sources of information, if we keep to the original model. We shall consider this question further in connection with the example of Section 3D.

A remark about the choice of prior may be appropriate here. In a rather abstract sense Bayesian theory supposes that an individual has a unique set of opinions at a given time, and thus a unique prior distribution of the parameters. Nonetheless the prior is never known with complete accuracy, and there is always a good deal of flexibility in the exact choice. Furthermore, when an informative experiment has been performed in the sense that the likelihood function is quite sharp relative to the prior, then the posterior distribution is essentially proportional to the likelihood function (i.e., we have stable estimation), and within wide limits the exact analytical specification of the prior is immaterial. In such circumstances certain conventional "diffuse" priors are convenient (like that of Jeffreys) although largely arbitrary. When the posterior distribution depends crucially upon the prior, as we shall see it does in inference upon σ_α^2 when $\hat{\sigma}_\alpha^2 < 0$, then no merely conventional prior can be used. In a certain sense this makes the prior seem to depend upon the data. In fact, however, it is only the degree of care we take in approximating our prior, not the prior itself, that depends upon the data.

B. *Inference on σ_α^2*

The posterior density for σ_α^2 is

$$\rho''(\sigma_\alpha^2) \propto \int_0^\infty \rho(\sigma^2, \sigma_\alpha^2)(\sigma^2)^{-I(J-1)/2} \exp\left\{-\frac{SSW}{2\sigma^2}\right\} (h^2)^{-(I-1)/2} \exp\left\{-\frac{SSB}{2h^2}\right\} d\sigma^2.$$

Along the lines of the previous discussion we first take

$$\rho(\sigma^2, \sigma_\alpha^2) = \rho_1(\sigma^2)\rho_2(\sigma_\alpha^2)$$

so that

$$\rho''(\sigma_\alpha^2) \propto \rho_2(\sigma_\alpha^2) \int_0^\infty \rho_1(\sigma^2) L(\sigma^2 \mid SSW)(h^2)^{-(I-1)/2} \exp\left\{-\frac{SSB}{2h^2}\right\} d\sigma^2.$$

Further, taking $\rho_1(\sigma^2) \propto (\sigma^2)^{-1}$ (we could just as well have chosen a proper gamma density) we have

$$\rho''(\sigma_\alpha^2) \propto \rho_2(\sigma_\alpha^2)\rho'(\sigma_\alpha^2)$$

where

$$\rho'(\sigma_\alpha^2) \propto \int_0^\infty (\sigma^2)^{-(I(J-1)/2)-1} \exp\left\{-\frac{SSW}{2\sigma^2}\right\} (h^2)^{-(I-1)/2} \exp\left\{-\frac{SSB}{2h^2}\right\} d\sigma^2. \quad (8)$$

Let now $X = \Gamma_{I-3/2, SSB/2}$ and $Y = \Gamma_{I(J-1)/2, SSW/2}$ have independent gamma distributions. Then since $\rho'(\sigma_\alpha^2)$ is proportional to the density of $Z = J^{-1}(X^{-1} - Y^{-1})$ on the positive half-line, the posterior density $\rho''(\sigma_\alpha^2) \propto \rho_2(\sigma_\alpha^2)\rho'(\sigma_\alpha^2)$ may be regarded as a truncation from below at zero of the distribution of Z together with a weighting by $\rho_2(\sigma_\alpha^2)$. We shall henceforth regard $\rho'(\sigma_\alpha^2)$ as proportional to the density of Z on the whole real line, using the fact $\rho_2(\sigma_\alpha^2) = 0$ for $\sigma_\alpha^2 < 0$ to yield the required truncation at zero.

It is $\rho'(\sigma_\alpha^2)$ that plays the role of a likelihood function (actually a marginal likelihood) in the formation of the marginal posterior density $\rho''(\sigma_\alpha^2)$. Since the distribution of Z depends upon I, J, SSB and SSW, these four quantities determined by the data are parameters for $\rho'(\sigma_\alpha^2)$. The important thing to see is that the form of the posterior $\rho''(\sigma_\alpha^2)$ depends crucially upon the place where the truncation at zero occurs in the function $\rho'(\sigma_\alpha^2)$, and this in turn depends upon the above four parameters. In particular, since the random variable Z has a unimodal density $\rho'(\sigma_\alpha^2)$ with mode roughly at $\hat{\sigma}_\alpha^2 = J^{-1}(MSB - MSW)$, it follows that when $\hat{\sigma}_\alpha^2 \ll 0$ the function $\rho'(\sigma_\alpha^2)$ will be monotonically decreasing in the interval $\sigma_\alpha^2 \geq 0$, with maximum at $\sigma_\alpha^2 = 0$. Since $\rho''(\sigma_\alpha^2) = 0$ for $\sigma_\alpha^2 < 0$ only this interval is of interest.

Thus we see that a negative $\hat{\sigma}_\alpha^2$ leads to a likelihood factor with maximum at $\sigma_\alpha^2 = 0$ and decreasing monotonically in σ_α^2. The over-all import of such data is thus to give relatively more weight to small σ_α^2 in the posterior than in the prior, and this is presumably in accord with some frequentistic interpretations of negative $\hat{\sigma}_\alpha^2$. However, when we ask further how the degree of sharpness with which $\rho'(\sigma_\alpha^2)$ is peaked at zero (which determines how strongly the data suggest that σ_α^2 is small) depends upon the magnitude of the assumed negative $\hat{\sigma}_\alpha^2$, we come to a divergence of viewpoint. For when $\hat{\sigma}_\alpha^2 = J^{-1}(MSB - MSW) < 0$, then the larger is SSW (and hence keeping SSB, I and J fixed, the more negative is $\hat{\sigma}_\alpha^2$) then the flatter is the function $\rho'(\sigma_\alpha^2)$. In fact, if all other quantities are held fixed,

$$\lim_{SSW \to \infty} \rho''(\sigma_\alpha^2) = \rho_2(\sigma_\alpha^2)$$

so that the posterior is the same as the prior, and in this sense the experiment has been completely uninformative about σ_α^2!

To prove this we first note that

$$\left.\frac{\partial \ln \rho'(\sigma_\alpha^2)}{\partial \sigma_\alpha^2}\right|_{\sigma_\alpha^2=0} \approx \frac{(I-1)(J-1)}{2(MST)^2}\,[[1+2/I(J-1)]MSB - MSW] \quad (9)$$

where

$$MST = \sum_{i,j} (y_{ij} - \bar{y}_{..})^2/(IJ-1). \text{ Thus } \left.\frac{\partial \rho'(\sigma_\alpha^2)}{\partial \sigma_\alpha^2}\right|_0 < 0$$

if and only if $MSW > [1+2/I(J-1)]MSB$ or $\hat{\sigma}_\alpha^2 < -2MSB/IJ(J-1)$, and when this occurs $\rho'(\sigma_\alpha^2)$ decreases monotonically from its value at zero. From (8) it follows easily that

$$\lim_{\sigma_\alpha^2 \uparrow \infty} \frac{\partial \ln \rho'(\sigma_\alpha^2)}{\partial \sigma_\alpha^2} = 0,$$

the limit being approached monotonically from below. Further, an asymptotic expansion as

$$SSW \uparrow \infty \text{ yields } \frac{\partial^2 \ln \rho'(\sigma_\alpha^2)}{\partial^2 \sigma_\alpha^2} > 0$$

simultaneously for all σ_α^2 in any finite interval for sufficiently large SSW, and

$$\frac{\partial \ln \rho'(\sigma_\alpha^2)}{\partial \sigma_\alpha^2}$$

is then monotonically increasing within such intervals from its negative value (9) at $\sigma_\alpha^2 = 0$. Since also

$$\lim_{SSW \uparrow \infty} \left.\frac{\partial \ln \rho'(\sigma_\alpha^2)}{\partial \sigma_\alpha^2}\right|_0 = 0,$$

it follows that $\rho'(\sigma_\alpha^2)$ becomes "flat" (measured by logarithmic derivative) as $SSW \uparrow \infty$, and this in turn implies

$$\lim_{SSW \uparrow \infty} \rho''(\sigma_\alpha^2) = \rho_2(\sigma_\alpha^2), \text{ for all } \sigma_\alpha^2,$$

where both $\rho''(\sigma_\alpha^2)$ and $\rho_2(\sigma_\alpha^2)$ are here assumed normalized to have unit area. Convergence in distribution is then easily shown. It is worth observing that as $SSB \downarrow 0$,

$$\left.\frac{\partial \ln \rho'(\sigma_\alpha^2)}{\partial \sigma_\alpha^2}\right|_0$$

decreases to

$$-\frac{IJ}{2MSW}\,\frac{\left(1-\frac{1}{I}\right)\left(1-\frac{1}{IJ}\right)}{\left(1-\frac{1}{J}\right)} \quad \text{or} \quad \approx -\frac{IJ}{2MSW}$$

if both I and J are large, which indicates that the more negative $\hat{\sigma}_\alpha^2$ is for *fixed* SSW, then the stronger is the indication that σ_α^2 is small.

One way of seeing the implications of these results is to compare $\rho'(\sigma_\alpha^2)$ with an exponential exp $[-f\sigma_\alpha^2]$. For fixed SSW the decreasing function $\rho'(\sigma_\alpha^2)$ is most concentrated near zero when $SSB=0$, and in this case

$$\left.\frac{\partial \ln \rho'(\sigma_\alpha^2)}{\partial \sigma_\alpha^2}\right|_0 \approx -IJ/2MSW.$$

For large SSW,

$$\frac{\partial \ln \rho'(\sigma_\alpha^2)}{\partial \sigma_\alpha^2}$$

is an increasing function of σ_α^2, so $\rho'(\sigma_\alpha^2)$ is "flatter" than

$$\exp\left[-\frac{IJ}{2MSW}\,\sigma_\alpha^2\right]$$

in the sense that

$$\rho'(0)/\rho'(\sigma_\alpha^2) \le \exp\left[\frac{IJ}{2MSW}\,\sigma_\alpha^2\right].$$

When $IJ/2MSW$ is small we may thus be able to quickly appraise the over-all influence of the data upon our prior opinion about σ_α^2, and see that the posterior is essentially the same as the prior.

We have so far discussed only the case $\hat{\sigma}_\alpha^2 \ll 0$, when the truncation leaves only the monotonically decreasing portion of $\rho'(\sigma_\alpha^2)$. Since $E(Z)\approx\hat{\sigma}_\alpha^2$, Var $(Z)\approx[2/IJ^2][MSB^2+MSW^2/J]$ it follows that a positive $\hat{\sigma}_\alpha^2$ may lead to a situation in which only a small portion of the area under $\rho'(\sigma_\alpha^2)$ lies to the left of zero and is cut off by the truncation. In this case it will sometimes suffice to approximate $\rho'(\sigma_\alpha^2)$ by a normal density with the same mean and variance. Even when the truncation occurs precisely at the mode of $\rho'(\sigma_\alpha^2)$ a normal approximation may be suitable, although now only the half of the normal density above its mean will be used. As remarked earlier in connection with σ^2, such approximations are primarily useful in planning the experiment, since the likelihood factor $\rho'(\sigma_\alpha^2)$ and the posterior $\rho''(\sigma_\alpha^2)$ can always be obtained numerically.

We note that Var (Z) increases with MSW, so that when $MSB > MSW$ (and thus $\hat{\sigma}_\alpha^2 > 0$), a large MSW leads to a relatively spread out posterior distribution centered near $\hat{\sigma}_\alpha^2$. Both this and our earlier result that large SSW

leads to a flat $\rho'(\sigma_\alpha^2)$ when $\hat{\sigma}_\alpha^2<0$ are reflections of the fact that before the data are taken, the larger σ^2 is, the greater the variance of the random variable $\hat{\sigma}_\alpha^2$. Bayesianly, after the data are taken, MSW plays much the same role as σ^2 did prior to the data.

Finally, if we use the Jeffreys prior $\rho(\sigma^2, \sigma_\alpha^2)=(\sigma^2)^{-1}(\sigma^2+J\sigma_\alpha^2)^{-1}$, then the posterior distribution of σ_α^2 is that of the random variable

$$J^{-1}(SSB/\chi^2_{I-1} - SSW/\chi^2_{I(J-1)})$$

truncated from below at zero. This distribution, censored instead of truncated from below at zero, has also been proposed as a fiducial solution [11], [14]. Our viewpoint is that when $\hat{\sigma}_\alpha^2<0$, and particularly when MSW is large, this distribution (even truncated, and certainly when censored) is inappropriate as a measure of posterior opinion. For we have seen that in this case the data are extremely uninformative, that $\rho''(\sigma_\alpha^2)\approx\rho_2(\sigma_\alpha^2)$, and since one is left with his prior density it is inappropriate to use a merely formal prior such as that of Jeffreys. From our point of view such a formal prior is only useful as a convenient but more or less arbitrary choice made from a collection of priors all of which essentially yield the likelihood function (normalized in some way) as posterior distribution. When, as in the present case with $\hat{\sigma}_\alpha^2\ll 0$, the likelihood factor is flat relative to the prior (the reverse of "stable estimation"), it becomes crucial to carefully assess the prior.

C. *Inference on σ_α^2/σ^2 and μ*

We consider now inference on $\tau^2=\sigma_\alpha^2/\sigma^2$ and μ based upon the prior

$$\rho(\mu, \sigma^2, \sigma_\alpha^2) \propto (\sigma^2)^{-1}(\sigma_\alpha^2)^{(-\lambda_\alpha/2)-1}\exp\left[-\frac{C_\alpha}{2\sigma_\alpha^2}\right] \tag{10}$$

We find (see Section 4 for details) the posterior density

$$g''(\tau^2) \propto \frac{(\tau^2)^{(-\lambda_\alpha/2)-1}\left(\tau^2+\dfrac{1}{J}\right)^{-(I-1)/2}}{\left[SSW+\dfrac{C_\alpha}{\tau^2}+\dfrac{SSB}{1+J\tau^2}\right]^{(N+\lambda_\alpha-1)/2}} \quad \text{for} \quad \tau^2\geq 0.$$

Note that if we had chosen $\lambda_\alpha=0$, $C_\alpha=0$, corresponding to the prior $(\sigma_\alpha^2)^{-1}$, then

$$g''(\tau^2) \propto \frac{(\tau^2)^{-1}\left(\tau^2+\dfrac{1}{J}\right)^{-(I-1)/2}}{\left[SSW+\dfrac{SSB}{1+J\tau^2}\right]^{(IJ-1)/2}}$$

$$\propto (\tau^2)^{-1}\frac{\left(\tau^2+\dfrac{1}{J}\right)^{(I(J-1))/2}}{[SSB+SSW(1+J\tau^2)]^{(IJ-1)/2}}$$

would be improper. However, the difficulty arises only at τ^2 near zero, where the prior $(\sigma_\alpha^2)^{-1}$ cannot be viewed as realistic, and sometimes can be avoided by the simple device of truncating $g''(\tau^2)$ at some $\epsilon>0$, or alternatively by approximating

$$(\tau^2)^{-1} \text{ by } \left(\tau^2+\frac{1}{J}\right)^{-1},$$

which yields the same result as if we had chosen the Jeffreys prior to begin with. In the latter case we find

$$g''(\tau^2) \propto \frac{\left(\tau^2+\frac{1}{J}\right)^{(I(J-1)/2)-1}}{[SSB+SSW(1+J\tau^2)]^{(IJ-1)/2}}$$

which states that

$$\frac{SSW}{SSB}(1+J\tau^2) = \frac{\chi^2_{I(J-1)}}{\chi^2_{I-1}}$$

truncated from below at SSW/SSB, or

$$(1+J\tau^2) \sim \frac{MSB}{MSW} F_{I(J-1),I-1}, \quad \text{with } F \tag{11}$$

truncated from below at MSW/MSB, in some but not complete harmony with tradition. Here $F_{I(J-1),I-1}$ denotes a random variable having the truncated F distribution with the indicated degrees of freedom. We also note that in the balanced case with $\lambda_\alpha=C_\alpha=0$, that given τ^2,

$$\left[SSW+\frac{SSB}{1+J\tau^2}\right]\Big/\sigma^2$$

has the chi-square distribution with $IJ-1$ degrees of freedom.

Now let us consider the situation $MSB \le MSW$. To gain insight we shall make the approximation

$$1+J\tau^2 \sim \frac{MSB}{MSW} F_{I(J-1),I-1}$$

with F truncated from below at MSW/MSB (the approximation of course may sometimes be poor) and find

$$\Pr''\{t_0 \le \tau^2 \le t_1\} = \frac{\Pr\left\{\frac{MSW}{MSB}(1+Jt_0) \le F^* \le \frac{MSW}{MSB}(1+Jt_1)\right\}}{\Pr\left\{\frac{MSW}{MSB} \le F^* < \infty\right\}},$$

where here F^* has an ordinary F distribution with the above degrees of freedom. Now, depending on the degrees of freedom, unity may be either more or less than the 50th percentile of F^*. However, in the present case, with numerator degrees of freedom larger than denominator, unity is always less than the 50th percentile, the latter approaching unity as both degrees of freedom grow large. If $MSB \leq MSW$ then the probability $\Pr''\{t_0 \leq \tau^2 \leq t_1\}$ for intervals of given length will be maximal when $t_0 = 0$, and

$$\Pr''\{0 \leq \tau^2 \leq t_1\} = \frac{\Pr\left\{\frac{MSW}{MSB} \leq F^* \leq \frac{MSW}{MSB}(1 + Jt_1)\right\}}{\Pr\left\{\frac{MSW}{MSB} \leq F^* < \infty\right\}}.$$

Further, if we take the limit as MSW/MSB goes to ∞, then

$$\lim_{MSW/MSB \uparrow \infty} \Pr''\{0 \leq \tau^2 \leq t_1\} = 1 - (1 + Jt_1)^{-(I-1)/2} \tag{12}$$

which is a proper probability distribution. When I and J are large it is thus clear that a large ratio MSW/MSB leads to a strong opinion that τ^2 is very small. Since in this case the truncation occurs far out in the tail of the F distribution, in practice it would not be clear how seriously to take this limiting distribution in view of the strong dependence upon the tail of F.

When $MSW/MSB \approx 1$, $\Pr''\{0 \leq \tau^2 \leq t_1\}$ may be quite large for even small t_1. For example, if both degrees of freedom are large, then

$$\Pr''\{0 \leq \tau^2 \leq t_1\} \approx 2 \Pr\{1 \leq F^* \leq 1 + Jt_1\}$$

and if $t_1 = J^{-1}$ this is $2 \Pr\{1 \leq F^* \leq 2\}$ which will be very near unity.

Inference on μ is based on the result (see Section 4 for details) that, given τ^2, $\sqrt{n}H(\mu - \bar{y}_{..})$ has the Student t distribution with $n = N + \lambda_\alpha - 1$ degrees of freedom, and

$$\sqrt{n}H = [N + \lambda_\alpha - 1]^{1/2} \left[\frac{IJ/(1 + J\tau^2)}{SSW + \frac{C_\alpha}{\tau^2} + \frac{SSB}{1 + J\tau^2}}\right]^{1/2};$$

while, if further $\lambda_\alpha = C_\alpha = 0$, then

$$\mu - \bar{y}_{..} \sim \left[\frac{SST + J\tau^2 SSW}{N(N-1)}\right]^{1/2} t_{(N-1)} \tag{13}$$

given τ^2, where here

$$SST = SSB + SSW = \sum_{i,j} (y_{ij} - \bar{y}_{..})^2,$$

and $t_{(N-1)}$ is a random variable having the t distribution with $N-1$ degrees of freedom. Thus, we obtain simple conditional credible intervals on μ, given τ^2, and obtain over-all posterior probabilities by integrating with respect to $g''(\tau^2)$. In any event, since for each given τ^2 it is clear that the posterior expectation of μ is $\bar{y}_{..}$, this is also true unconditionally, and in fact the posterior distribution of μ is symmetric about $\bar{y}_{..}$.

D. *An Example and Discussion*

We shall consider a rather extreme example in order to bring out most clearly the nature of our results. Suppose that $I=5$, $J=10$, $MSW=2500$, $MSB=.01$, and thus $\hat{\sigma}_\alpha^2=1/10\ (.01-2500)\approx -250$. Then

(a) Inference upon $\tau^2=\sigma_\alpha^2/\sigma^2$ can be based upon the limiting distribution (12) as MSW/MSB goes to ∞, so that $\Pr''\{0\leq\tau^2\leq c|\,\text{data}\}\approx 1-(1+10c)^{-2}$, which is .99 if $c=.9$, and the data provide a strong indication that $\tau^2<1$.

(b) Since $-IJ/2MSW=-.01$ and SSW is sufficiently large, the likelihood factor $\rho'(\sigma_\alpha^2)$ is "flatter" than the exponential $\exp[-.01\ \sigma_\alpha^2]$, and thus within the inverval $0\leq\sigma_\alpha^2\leq 10$ the likelihood ratio $\rho'(0)/\rho'(\sigma_\alpha^2)\leq\exp\ [.1]$ ≈ 1.1. Within this interval the likelihood factor and hence the data modify the prior $\rho_2(\sigma_\alpha^2)$ only slightly. Generally speaking in this kind of situation the prior density $\rho_2(\sigma_\alpha^2)$ must be carefully evaluated if the model is not rejected and decisions are to be based upon knowledge of σ_α^2.

(c) Based upon (7) we find $\sigma^2\sim SSW/\chi_{49}^2=112500/\chi_{49}^2$ and the posterior expectation of σ^2 is about 2300. As discussed earlier the discrepancy between the inference based upon the $\bar{y}_{i.}$ alone and that based upon the $y_{ij}-\bar{y}_{i.}$ alone is disturbing. Thus $h^2\sim SSB/\chi_4^2=.04/\chi_4^2$, so that

$$\Pr'\{\sigma^2<c\,|\,\bar{y}_{i.}\}\geq\Pr'\{h^2<c\,|\,\bar{y}_{i.}\}=\Pr\{\chi_4^2>.04/c\}$$

which is quite near unity if $c>10$. But

$$\Pr'\{\sigma^2>c\,|\,y_{ij}-\bar{y}_{i.}\}=\Pr\{SSW/\chi_{45}^2>c\}=\Pr\{\chi_{45}^2<112500/c\}$$

which is also quite near unity if say $c<1000$. In a situation such as this one is forced to face up to the real uncertainties concerning the model. This is not at all inconsistent with Bayesian philosophy since real opinions about data never precisely correspond to any model, and even a very small prior doubt of the model can be enormously magnified by extreme data. When such data occur there is no alternative but to broaden the model, often along the lines suggested by the data. In this example it is natural to consider constraints, either deterministic or probabilistic, that tend to keep the $\bar{y}_{i.}$ nearly equal. It is important to see that extreme data cannot in themselves cause rejection of the model (except of course if the observations are literally impossible), but rather that they may give heavy weight to alternative models which were never before the experiment regarded as totally implausible. Although the Bayesian theory of models is in a primitive form, some useful ideas have been presented [1], [2], [17, Ch. V and VI].

(d) Using (13) we have, given τ^2,

$$\mu\sim\bar{y}_{..}+\left[\frac{SST+J\tau^2SSW}{N(N-1)}\right]^{1/2}t_{(N-1)}$$

$\approx \bar{y}_{..} + \sqrt{45}(1+10\tau^2)^{1/2} t_{49}$, and the posterior distribution of μ is symmetric about $\bar{y}_{..}$. Depending upon the amount of work we wish to do we can alternatively either integrate out τ^2, replace τ^2 by its posterior expectation making some allowance for uncertainty about τ^2, or simply make use of rough bounds such as $45 \leq \text{Var}(\mu | \text{data}) \leq 11 \times 45$.

We emphasize that in this artificial example there can be no prior opinions about the parameters, since neither the data nor the parameters have any meaning. Our purpose was merely to indicate the form which inference will tend to take. In any real problem the prior can always be assessed, although perhaps with difficulty.

4. THE UNBALANCED CASE

We shall now deal with the estimation of σ^2, σ_α^2, $\tau^2 = \sigma_\alpha^2/\sigma^2$ and μ in the general unbalanced case. Our approach will be to work directly with the bivariate posterior distribution of (σ^2, τ^2) and (μ, τ^2), since the conditional posterior distributions of σ^2, σ_α^2 and μ, given τ^2, take on standard and relatively simple forms.

Based upon the approximate posterior density (2), and letting

$$\rho(\sigma^2, \sigma_\alpha^2) \propto (\sigma^2)^{-1}(\sigma_\alpha^2)^{-\lambda_\alpha/2-1} \exp[-C_\alpha/2\sigma_\alpha^2],$$

we obtain the posterior density

$$\begin{aligned} g''(\sigma^2, \tau^2) &\propto \sigma^2 \rho''(\sigma^2, \sigma^2\tau^2) \\ &\propto \sigma^2 \rho(\sigma^2, \sigma^2\tau^2)(\sigma^2)^{-(N-1)/2} \exp\left[-\frac{SSW}{2\sigma^2}\right] \prod_i (1 + J_i\tau^2)^{-1/2} \\ &\quad \times \left[\sum_i \frac{J_i}{1 + J_i\tau^2}\right]^{-1/2} \exp\left[-\frac{1}{2\sigma^2} \sum_i \frac{J_i(\bar{y}_{i.} - \hat{\mu})^2}{1 + J_i\tau^2}\right] \\ &\propto (\sigma^2)^{-((N-1+\lambda_\alpha)/2)-1} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{\tau^2 SSW + C_\alpha}{\tau^2} + \sum_i \frac{J_i(\bar{y}_{i.} - \hat{\mu})^2}{1 + J_i\tau^2}\right)\right] \\ &\quad \times (\tau^2)^{-(\lambda_\alpha/2)-1} \prod_i (1 + J_i\tau^2)^{-1/2} \left[\sum_i \frac{J_i}{1 + J_i\tau^2}\right]^{-1/2}. \end{aligned}$$

The marginal posterior density of τ^2 is then

$$g''(\tau^2) \propto \frac{(\tau^2)^{-\lambda_\alpha/2-1} \prod_i (1 + J_i\tau^2)^{-1/2} \left[\sum_i \frac{J_i}{1 + J_i\tau^2}\right]^{-1/2}}{\left\{SSW + \frac{C_\alpha}{\tau^2} + \sum_i \frac{J_i(\bar{y}_{i.} - \hat{\mu})^2}{1 + J_i\tau^2}\right\}^{(N-1+\lambda_\alpha)/2}},$$

while the conditional posterior density of σ^2 given τ^2 is

$$\frac{g''(\sigma^2, \tau^2)}{g''(\tau^2)} \propto (\sigma^2)^{-((N-1+\lambda_\alpha)/2)-1} \exp\left\{-\frac{1}{2\sigma^2}\left(SSW + \frac{C_\alpha}{\tau^2} + \sum_i \frac{J_i(\bar{y}_{i.} - \hat{\mu})^2}{1 + J_i\tau^2}\right)\right\}.$$

Hence, given τ^2, the variable

$$\left(SSW + \frac{C_\alpha}{\tau^2} + \sum_i \frac{J_i(\bar{y}_{i.} - \hat{\mu})^2}{1 + J_i\tau^2}\right) \Big/ \sigma^2$$

is distributed like chi-square on $N+\lambda_\alpha - 1$ degrees of freedom

$$\Pr''\{t_0 \le \sigma^2 \le t_1\} = \int_0^\infty \Pr''\{t_0 \le \sigma^2 \le t_1 \mid \tau^2\} g''(\tau^2) d\tau^2$$

$$= \int_0^\infty \Pr\left\{\frac{SSW + \frac{C_\alpha}{\tau^2} + \sum_i \frac{J_i(\bar{y}_{i.} - \hat{\mu})^2}{1 + J_i\tau^2}}{t_1} \le \chi^2_{(N+\lambda_\chi - 1)} \le \frac{SSW + \frac{C_\alpha}{\tau^2} + \sum_i \frac{J_i(\bar{y}_{i.} - \hat{\mu})^2}{1 + J_i\tau^2}}{t_0}\right\} g''(\tau^2) d\tau^2$$

and

$$\Pr''\{t_0 \le \sigma_\alpha^2 \le t_1\} = \int_0^\infty \Pr''\{t_0 \le \sigma_\alpha^2 \le t_1 \mid \tau^2\} g''(\tau^2) d\tau^2$$

$$= \int_0^\infty \Pr''\left\{\frac{t_0}{\tau^2} \le \sigma^2 \le \frac{t_1}{\tau^2} \,\middle|\, \tau^2\right\} g''(\tau^2) d\tau^2$$

so that each of these posterior probabilities is the integral of the chi-square probability of an interval whose endpoints are functions of τ^2, with respect to the posterior distribution of τ^2. Various approximations are available, the simplest, which is appropriate when $g''(\tau^2)$ points sharply to a value $\hat{\tau}^2 > 0$, being

$$\Pr''\{t_0 \le \sigma^2 \le t_1\} \approx \Pr''\{t_0 \le \sigma^2 \le t_1 \mid \tau^2 = \hat{\tau}^2\}$$

and

$$\Pr''\{t_0 \le \sigma_\alpha^2 \le t_1\} \approx \Pr''\left\{\frac{t_0}{\hat{\tau}^2} \le \sigma^2 \le \frac{t_1}{\hat{\tau}^2} \mid \tau^2 = \hat{\tau}^2\right\},$$

the latter probabilities being evaluated from the chi-square distribution. Even when $g''(\tau^2)$ is not particularly sharp these approximations will be appropriate if the endpoints of the interval for $\chi^2(_{N-\lambda_\alpha - 1})$ are relatively gentle functions of τ^2. In general, however, we must be prepared to perform the integration, perhaps numerically, and towards this end we shall now examine $g''(\tau^2)$ more closely.

Write

$$g''(\tau^2) \propto \frac{(\tau^2)^{-(\lambda_\alpha/2)-1} \prod (\tau^2 + L_i)^{-1/2} \left[\sum \frac{1}{\tau^2 + L_i}\right]^{-1/2}}{\left[SSW + \frac{C_\alpha}{\tau^2} + \sum \frac{(\bar{y}_{i.} - \hat{\mu})^2}{\tau^2 + L_i}\right]^{(N+\lambda_\alpha - 1)/2}} = g_1(\tau^2) g_2(\tau^2)$$

where $g_1(\tau^2)$ and $[g_2(\tau^2)]^{-1}$ are the numerator and denominator of $g''(\tau^2)$, and $L_i = J_i^{-1}$. Verify that

$$g_1(\tau^2) = (\tau^2)^{-((I+\lambda_\alpha-1)/2)-1} \prod (1 + L_i/\tau^2)^{-1/2} \left[\sum_i \frac{1}{1 + L_i/\tau^2} \right]^{-1/2},$$

$$\frac{\partial \ln g_1(\tau^2)}{\partial \tau^2} < 0, \quad \text{and} \quad \lim_{\tau^2 \to \infty} (\tau^2)^{((I+\lambda_\alpha-1)/2)+1} g_1(\tau^2) = I^{-1/2},$$

so that $g_1(\tau^2)$ decreases montonically from $g_1(0) = \infty$ to $g_1(\infty) = 0$.

Next we find (see also Wald [27])

$$\frac{\partial}{\partial \tau^2} \left\{ \sum \frac{(\bar{y}_{i\cdot} - \hat{\mu})^2}{\tau^2 + L_i} \right\} = - \sum \frac{(\bar{y}_{i\cdot} - \hat{\mu})^2}{(\tau^2 + L_i)^2}$$

so that

$$g_2(\tau^2) = \left\{ SSW + \frac{C_\alpha}{\tau^2} + \sum_i \frac{(\bar{y}_{i\cdot} - \hat{\mu})^2}{\tau^2 + L_i} \right\}^{-(N+\lambda_\alpha-1)/2}$$

is monotonically increasing from

$$g_2(0) = \begin{cases} 0 & \text{if } C_\alpha > 0, \\ [SSW + \sum J_i(\bar{y}_{i\cdot} - \bar{y})^2]^{-(N-1)/2} & \text{if } C_\alpha = 0,\ \lambda_\alpha = 0, \end{cases}$$

to

$$g_2(\infty) = [SSW]^{-(N+\lambda_\alpha-1)/2},$$

where here

$$\bar{y} = (\sum J_i \bar{y}_{i\cdot}) / \sum J_i.$$

Thus $g''(\tau^2)$ is the product of monotonically increasing and decreasing factors, typically goes to zero as τ^2 goes to zero or infinity, and has a unique mode say at $\hat{\tau}^2$. By various approximations we can judge the posterior probability of intervals about $\hat{\tau}^2$ (or better centered at the posterior mean or median), and thus determine approximate credible intervals. It is also natural to approximate $g''(\tau^2)$ by an F density (with degrees of freedom to be fitted), since $g''(\tau^2)$ is of this form in the balanced case.

Now we turn to inference about μ. If $g''(\mu, \sigma^2, \tau^2)$ is the joint posterior density of (μ, σ^2, τ^2) then

$$g''(\mu, \tau^2) = \int_0^\infty g''(\mu, \sigma^2, \tau^2) d\sigma^2$$

$$\propto (\tau^2)^{-\lambda_\alpha/2-1} \prod (1 + J_i\tau^2)^{-1/2} \int_0^\infty (\sigma^2)^{-((N+\lambda_\alpha)/2)-1}$$

$$\times \exp \left[- \frac{1}{2\sigma^2} \left\{ SSW + \frac{C_\alpha}{\tau^2} + \sum \frac{J_i(\bar{y}_{i\cdot} - \mu)^2}{1 + J_i\tau^2} \right\} \right] d\sigma^2$$

$$\propto \frac{(\tau^2)^{-(\lambda_\alpha/2)-1} \prod (1 + J_i\tau^2)^{-1/2}}{\left\{SSW + \frac{C_\alpha}{\tau^2} + \sum \frac{J_i(\bar{y}_{i\cdot} - \mu)^2}{1 + J_i\tau^2}\right\}^{(N+\lambda_\alpha)/2}}.$$

Since

$$\sum \frac{J_i(\bar{y}_{i\cdot} - \mu)^2}{1 + J_i\tau^2} = \sum \frac{J_i(\bar{y}_{i\cdot} - \hat{\mu})^2}{1 + J_i\tau^2} + (\mu - \hat{\mu})^2 \sum \frac{J_i}{1 + J_i\tau^2},$$

then

$$g''(\mu \mid \tau^2) \propto \left\{1 + H^2(\mu - \hat{\mu})^2\right\}^{-(N+\lambda_\alpha)/2},$$

where

$$H = \left[\frac{\sum \frac{J_i}{1 + J_i\tau^2}}{SSW + \frac{C_\alpha}{\tau^2} + \sum \frac{J_i(\bar{y}_{i\cdot} - \hat{\mu})^2}{1 + J_i\tau^2}}\right]^{1/2}$$

Now let $\chi = \sqrt{n}H(\mu - \hat{\mu})$, $n = N + \lambda_\alpha - 1$. Then, given τ^2, the conditional posterior density of χ is

$$\rho''(\chi \mid \tau^2) \propto \left\{1 + \frac{\chi^2}{n}\right\}^{-(n+1)/2},$$

so that $\sqrt{n}H(\mu - \hat{\mu})$ has (conditional on τ^2) Student's distribution with $n = N + \lambda_\alpha - 1$ degrees of freedom.

By numerical integration or other approximation methods we can obtain the unconditional posterior distribution of μ; for example, if $g''(\tau^2)$ is sharply concentrated at say $\hat{\tau}^2$, then we can act as though it were known that $\tau^2 = \hat{\tau}^2$ making some rough allowance for the residual uncertainty about τ^2.

We note that the unbalanced case offers only more complexity in the form of the posterior distributions, and no fundamental difficulties.

REFERENCES

[1] Anscombe, F. J., "Tests of Goodness of Fit," *Journal of the Royal Statistical Society, Series B*, 25 (1963), 81–94.

[2] Box, G. E. P., Cox, D. R., "An Analysis of Transformations," *Journal of the Royal Statistical Society, Series B*, 26 (1964), 211–52.

[3] Bross, I., "Fiducial Intervals for Variance Component Analysis," *Biometrics* 6 (1950), 136–44.

[4] Bulmer, M. G., "Approximate Confidence Limits for Components of Variance," *Biometrika* 44 (1957), 159–67.

[5] Crump, S. L., "Estimation of Variance Components in Analysis of Variance", *Biometrics Bulletin,* Vol 2 (1946), p. 7.

[6] Crump, S. L., "The Present Status of Variance Component Analysis," *Biometrics* 7 (1951), 1–15.

[7] Daniels, H. E., "The Estimation of Components of Variance," *Journal of the Royal Statistical Society, Supplement* 6 (1939), 186–97.

[8] Edwards, W., Lindman, H., and Savage, L. J., "Bayesian Statistical Inference for Psychological Research," *Psychological Review* 70 (1963), 193–242.

[9] Eisenhart, C., "The Assumptions Underlying Analysis of Variance," *Biometrics* 3 (1947), 1–21.

[10] Fisher, R. A., *Statistical Method for Research Workers*. Thirteenth Edition. New York: Hafner Publishing Company, Inc., 1958.

[11] Fisher, R. A. "The Fiducial Argument in Statistical Inference," *Annals of Eugenics* 6 (1935), 391–8.

[12] Fisher, R. A., *The Design of Experiments*. Seventh Edition. New York: Hafner Publishing Company, Inc., 1960.

[13] Green, J. R., "A Confidence Interval for Variance Components," *Annals of Mathematical Statistics* 25 (1954), 671–85.

[14] Healy, M. J. R., "Fiducial Limits for a Variance Component," *Journal of the Royal Statistical Society, Series B*, 25 (1963), 128–30.

[15] Herbach, L. H., "Properties of Model II—Type Analysis of Variance Tests . . . ," *Annals of Mathematical Statistics*, 30 (1959), 939–59.

[16] Hill, B. M., "The Three Parameter Log-Normal Distribution and Bayesian Analysis of a Point-Source Epidemic," *Journal of the American Statistical Association*, 58 (1963), 72–84.

[17] Jeffreys, H., *Theory of Probability*. Third Edition, Oxford: Clarendon Press, 1961.

[18] Nelder, J. A., "The interpretation of negative components of variance," *Biometrika*, 41 (1954), 544–8.

[19] Raiffe H., and Schlaifer, R., *Applied Statistical Decision Theory*. Boston: Harvard Business School, 1961.

[20] Sattherthwaite, F. E., "An Approximate Distribution of Estimates of Variance Components," *Biometrics* 2 (1946), 110–14.

[21] Savage, L. J., *The Foundations of Statistics*. New York: John Wiley and Sons, Inc., 1954.

[22] Savage, L. J., *The Foundations of Statistical Inference, A Discussion*. London: Methuen and Co., Ltd., New York: John Wiley and Sons, Inc., 1962.

[23] Savage, L. J., "The Subjective Basis of Statistical Practice," University of Michigan Technical Report, 1961.

[24] Scheffé, H., *The Analysis of Variance*, New York: John Wiley and Sons, Inc., 1961.

[25] Searle, S. R., "Matrix Methods in Components of Variance and Covariance Analysis," *Annals of Mathematical Statistics*, 27 (1956), 737–48.

[26] Thompson, W. A., Jr., "The Problem of Negative Estimates of Variance Components," *Annals of Mathematical Statistics*, 33 (1962), 273–89.

[27] Wald, A., "A Note on the Analysis of Variance with Unequal Class Frequencies," *Annals of Mathematical Statistics*, 11 (1940), 96–100.

[28] Tiao, G. C., and Tan, W. Y., "Bayesian Analysis of Random-Effect Models . . . ," Technical Report No. 30, University of Wisconsin (1964).

[29] Yates, F., "The Analysis of Multiple Classifications with Unequal Numbers in the Different Classes," *Journal of the American Statistical Association*, 29 (1934), 51–66.

[30] Yates, F., Zacopanay, I., "The Estimation of the Efficiency of Sampling with Special Reference to Sampling for Yield in Cereal Experiments," *Journal of Agricultural Science*, 25 (1935), 545–77.