DEPARTMENT OF STATISTICS

The University of Wisconsin
Madison, Wisconsin

Technical Report #84
August 1965

BAYESIAN ANALYSIS OF A THREE-COMPONENT
HIERARCHICAL DESIGN MODEL

by

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

* This research was supported in part by National Science Foundation under grant G.S. - 549 and was completed when the authors were on leave of absence at Harvard University, Graduate School of Business Administration.
BAYESIAN ANALYSIS OF A THREE-COMPONENT
HIERARCHICAL DESIGN MODEL

G. C. Tiao and G. E. P. Box *
University of Wisconsin

SUMMARY

The problem of making inference about the parameters in the three-
component hierarchical design model $y_{ijk} = \mu + a_i + b_{ij} + e_{ijk}$ is con-
sidered from a Bayesian viewpoint. Under the usual normality and inde-
pendence assumptions and adopting a non-informative reference prior, va-
rious features of the posterior distribution of the variance components
$\sigma_1^2 = \text{Var}(e_{ijk})$, $\sigma_2^2 = \text{Var}(b_{ij})$ and $\sigma_3^2 = \text{Var}(a_i)$ are discussed, including
inferences about a variance ratio, the relative contributions of the com-
ponents and the magnitude of the individual components. A scaled $\chi^2$
ap-proximation technique is developed for the marginal distributions of the
components which can be applied to general q-component model. In addition,
a Bayesian solution to the problem of pooling variance estimates is given.

1. Introduction

We consider in this paper the problem of making inferences about the
parameters of the three-component hierarchical design random effect model

$$y_{ijk} = \mu + a_i + b_{ij} + e_{ijk} \tag{1.1}$$

$$i = 1, \ldots, I; \quad j = 1, \ldots, J; \quad k = 1, \ldots, K$$

where $y_{ijk}$ are the observations, $\mu$ is a common location parameter, $a_i$ and
$b_{ij}$ are two different kinds of random effects and $e_{ijk}$ are random errors
or disturbances. We shall make the usual assumptions that the effects $a_i,$
b_{ij} and the errors $e_{ijk}$ are all independent and that

$$a_i \sim N(0, \sigma_3^2), \quad b_{ij} \sim N(0, \sigma_2^2) \quad \text{and} \quad e_{ijk} \sim N(0, \sigma_1^2). \tag{1.2}$$

It follows in particular that $\text{Var}(y_{ijk}) = \sigma_1^2 + \sigma_2^2 + \sigma_3^2$ so that the param-
eters ($\sigma_1^2, \sigma_2^2, \sigma_3^2$) are the variance components. Hierarchical designs of
this type are often used in industrial experiments. For example, to in-

* This research was supported in part by National Science Foundation under
grant G.S. - 549 and was completed when the authors were on leave of
absence at Harvard University, Graduate School of Business Administration.
appears to contain more information about $\sigma_1^2$ than the single estimate $m_1$ but the practice of pooling presents obvious difficulties. For example, $\sigma_2^2$ might be non-zero but nevertheless fails to be detected by the significance test so that the pooled estimate is biased. The difficulties mentioned above all occur when the underlying assumptions of normality and independence are precisely justified. Additional complications arise if one attempts to study the effect of departures from these assumptions, see for example Scheffé (1959).

In the problem of variance components, the traditional sampling theory has thus led to a number of difficulties with no generally accepted set of solutions. Recently, the problem of the two-component model has been reconsidered from a Bayesian viewpoint by Tiao and Tan (1965) and Hill (1965). More recently, this approach has been employed by Tiao and Tan (1966) to show the effect of departures from independence of the errors.

In the present paper, we analyze the three-component model on the usual normal and independence assumptions and give a Bayesian discussion of the "pooling" dilemma.

2. Prior and Posterior Distributions of $(\mu, \sigma_1^2, \sigma_2^2, \sigma_3^2)$

From (1.1) and (1.2) and using the notations introduced in Table 1 the likelihood function can be written

$$L(\mu, \sigma_1^2, \sigma_{12}, \sigma_{123} | y) \propto \frac{\nu_1}{2} \left( \frac{\nu_1}{\sigma_1^2} \right)^{-\frac{\nu_1}{2}} \frac{\nu_2}{2} \left( \frac{\nu_2}{\sigma_{12}^2} \right)^{-\frac{\nu_2}{2}} \frac{\nu_3+1}{2} \left( \frac{\nu_3+1}{\sigma_{123}^2} \right)^{-\frac{\nu_3+1}{2}}$$

$$\times \exp \left[ \frac{1}{2} \left( \frac{\nu_3+1}{\sigma_{123}^2} \right) \left( \frac{\nu_1}{\sigma_1^2} \right)^{-\nu_1} \left( \frac{\nu_2}{\sigma_{12}^2} \right)^{-\nu_2} \left( \frac{\nu_3+1}{\sigma_{123}^2} \right)^{-\nu_3+1} \right]$$

(2.1)

We see in (2.1) that the information contained in the likelihood function can be regarded as coming from $(\nu_3+1)$ independent observations from a normal population $N(\mu, \sigma_{123}^2/JK)$, $\nu_2$ independent observations from another normal population $N(0, \sigma_{12}^2)$ and $\nu_1$ further independent observations from a third normal population $N(0, \sigma_1^2)$.

So far as prior information is concerned, the situation of most general interest is that where little is known about the value of the variance components initially. We shall therefore consider the inferences
and \( p(F_{\nu_1, \nu_2}) \) is the density function of an \( F \) variable with \( \nu_1 \) and \( \nu_2 \) degrees of freedom. The first factor \( \frac{m_1}{m_2} \cdot p\left(F_{\nu_1, \nu_2} = \frac{\sigma_{12}^2}{\sigma_1^2} \right) \) is the unconstrained distribution of \( \sigma_{12}^2 / \sigma_1^2 \) and the factor \( H_1(\sigma_{12}^2 / \sigma_1^2) \) represents the effect of the constraint \( C \). Now, we can regard \( C \) as the union of the two constraints \( C_1 : \sigma_1^2 < \sigma_{12}^2 \) and \( C_2 : \sigma_{12}^2 < \sigma_{123}^2 \). Thus, we can write

\[
\frac{\Pr^*\left( \frac{\sigma_{12}^2}{\sigma_1^2} , \gamma \right)}{\Pr^*(C|\gamma)} = \frac{\Pr^*\left( C_1, \frac{\sigma_{12}^2}{\sigma_1^2} , \gamma \right)}{\Pr^*(C_1|\gamma)} \frac{\Pr^*\left( C_2|C_1, \frac{\sigma_{12}^2}{\sigma_1^2} , \gamma \right)}{\Pr^*(C_2|C_1, \gamma)}
\]

(3.3)

from which it follows that

\[
p\left( \frac{\sigma_{12}^2}{\sigma_1^2} | \gamma \right) = \left( \frac{m_1}{m_2} \right) p\left(F_{\nu_1, \nu_2} = \frac{\sigma_{12}^2}{\sigma_1^2} \right) \cdot \frac{m_1}{m_2} \cdot H_1\left( \frac{\sigma_{12}^2}{\sigma_1^2} \right) \cdot H_{2.1}\left( \frac{\sigma_{12}^2}{\sigma_1^2} \right)
\]

(3.4)

with

\[
H_1\left( \frac{\sigma_{12}^2}{\sigma_1^2} \right) = \frac{1}{\Pr\left( F_{\nu_1, \nu_2} < \frac{m_2}{m_1} \right)}
\]

\[
H_{2.1}\left( \frac{\sigma_{12}^2}{\sigma_1^2} \right) = \frac{\Pr\left( F_{\nu_2, \nu_1} < \frac{m_2}{m_1} \right) \cdot \Pr\left( F_{\nu_3, \nu_1 + \nu_2} < \frac{m_3}{\nu_2 m_1 + \frac{\sigma_{12}^2}{\sigma_1^2} \nu_1 m_1} \right)}{\Pr\left( F_{\nu_2, \nu_1} < \frac{m_2}{m_1}, F_{\nu_3, \nu_2} < \frac{m_3}{m_2} \right)}
\]

Clearly, the effect of the constraint \( C_1 \) is to truncate the unconstrained distribution of \( \sigma_{12}^2 / \sigma_1^2 \) from below at the point \( \sigma_{12}^2 / \sigma_1^2 = 1 \) and \( H_1(\sigma_{12}^2 / \sigma_1^2) \) is precisely the normalizing constant induced by the truncation. The quantity \( H_{2.1}(\sigma_{12}^2 / \sigma_1^2) \) is a monotonic decreasing function of the ratio \( \sigma_{12}^2 / \sigma_1^2 \). Thus, the "additional" effect of the second constraint \( C_2 \) that \( \sigma_{12}^2 < \sigma_{123}^2 \)
We can now relate the above result to the sampling theory solution to this problem. It is well known that the sample quantity \( \frac{m_1}{m_2} \) is distributed as \( \frac{\sigma_1^2}{\sigma_2^2} F_{\nu_2, \nu_1} \). Thus, the unconstrained posterior distribution \( \frac{m_1}{m_2} \propto F_{\nu_2, \nu_1} \left( \frac{\sigma_1^2 m_1}{\sigma_2^2 m_2} \right) \) is also the "confidence distribution" of \( \frac{\sigma_2^2}{\sigma_1^2} \). The inference procedures based upon this confidence distribution are not satisfactory. In the first place the distribution extends from the origin to infinity so that the lower confidence limit for \( \frac{\sigma_2^2}{\sigma_1^2} = 1 + K \) can be smaller than unity and the corresponding limit for \( \frac{\sigma_2^2}{\sigma_1^2} \) less than zero. In addition, the sampling procedure fails to take into account the information coming from the distribution of the mean square \( m_3 \), which, in our Bayesian approach, is included through the constraint \( \sigma_{123}^2 > \sigma_{12}^2 \).

4. Relative Contribution of Variance Components

Features of the posterior distribution of \( \sigma_1^2, \sigma_2^2, \sigma_3^2 \), often of interest to the investigator, are the relative contributions of the variance components to the total variance of \( y_{1jk} \). In this connection inferences can be drawn by considering the joint distribution of the ratios

\[
\gamma_1 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}, \quad \gamma_2 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}
\]

(4.1)

In (2.6), we make the transformation from \( \sigma_1^2, \sigma_2^2, \sigma_3^2 \) to \( \sigma_1^2, \gamma_1, \gamma_2 \) and integrate out \( \sigma_1^2 \) to obtain

\[
p(\gamma_1, \gamma_2 | y) = N[x_1 \Phi_2 (J-1) + x_2 \Phi_1, I(K-1)]^3 x_1 (\nu_1/2)^{-2} x_2 (\nu_2/2)^{-2} (1 + x_1 + x_2)^{-\nu_1/2 - \nu_2/2}/2 \]  

(4.2)

where

\[
x_1 = \Phi_1 \frac{(1-\gamma_2) + (K-1)\gamma_1}{1 - \gamma_1 - \gamma_2}, \quad x_2 = \Phi_2 \frac{(1-\gamma_2) + (K-1)\gamma_1}{1 + (K-1)\gamma_1 + (JK-1)\gamma_2}
\]
\[ \varphi_1 = \frac{\nu_1 m_1}{\nu_2 m_2}, \quad \varphi_2 = \frac{\nu_3 m_3}{\nu_2 m_2} \quad (\gamma_1 > 0, \gamma_2 > 0, \gamma_1 + \gamma_2 < 1) \]

and \( N \) is the normalizing constant

\[ N^{-1} = \prod \left\{ \frac{x_1^2}{\nu_1 m_1}, \frac{x_2^2}{\nu_2 m_2}, \frac{x_3^2}{\nu_3 m_3} \right\} B \left( \frac{\nu_1}{2}, \frac{\nu_2}{2}, \frac{\nu_3}{2} \right) J_2 \kappa \psi_1 \psi_2 . \]

This distribution is analytically complicated. However, for a given set of data, contours of the density function can be plotted. Using the example introduced in the previous section, three contours of the posterior distribution of \((\gamma_1, \gamma_2)\) are shown in Figure 2. The mode of the distribution is approximately at the point \( P : (\gamma_{10} = 0.425, \gamma_{20} = 0.5) \) also shown. The regions contained by the contours have the property that the density of every point included exceeds that of every point excluded. Such regions are called H.P.D. regions by Box and Tiao (1965). The three contours A, B, C, in Figure 2 were drawn such that

\[ \begin{align*}
A: & \quad \log p(\gamma_{10}, \gamma_{20} | y) - \log p(\gamma_1, \gamma_2 | y) = 0.69 = \frac{1}{2} X_1^2(.5) \\
B: & \quad \log p(\gamma_{10}, \gamma_{20} | y) - \log p(\gamma_1, \gamma_2 | y) = 1.20 = \frac{3}{2} X_2^2(.3) \\
C: & \quad \log p(\gamma_{10}, \gamma_{20} | y) - \log p(\gamma_1, \gamma_2 | y) = 2.30 = \frac{3}{2} X_2^2(.1)
\end{align*} \]

That is, they roughly contain 50%, 70% and 90% of the probability mass respectively.

Figure 2 provides us with a very illuminating picture of the inferential situation for this example. A striking feature illustrated by the contours is that the two ratios \( \gamma_1 \) and \( \gamma_2 \) are strongly negatively correlated. Precise inference about the total percentage contribution of \( \sigma_2^2 \) and \( \sigma_3^2 \), i.e., \( \gamma_1 + \gamma_2 \) can be drawn by noting that nearly 90% of the probability mass is contained in the interval \( .80 < \gamma_1 + \gamma_2 < .95 \). On
the other hand, the evidence from the data are not at all sufficient to allow us to distinguish between the contribution of $\sigma_2^2$ and that of $\sigma_3^2$. This points to the need of additional data before any conclusions on this point can be properly drawn. In examples such as the one illustrated here, too much emphasis on point estimates such as the mode could be misleading.

Within the sampling theory framework one could, of course, calculate the relative contributions of $(\sigma_2^2, \sigma_3^2)$ by taking the ratios of the unbiased estimators $(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2)$,

$$\hat{\gamma}_1 = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 + \hat{\sigma}_3^2}, \quad \hat{\gamma}_2 = \frac{\hat{\sigma}_3^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 + \hat{\sigma}_3^2}$$ (4.4)

The resulting ratios correspond to the point $P_1$ on the diagram. For large sample, $(\hat{\gamma}_1, \hat{\gamma}_2)$ tends to the maximum likelihood estimates of $(\gamma_1, \gamma_2)$ and the asymptotic confidence regions obtained from normal theory would correspond to the H.P.D. regions in the Bayesian framework. Small sample properties of $(\hat{\gamma}_1, \hat{\gamma}_2)$ are, however, far from clear.

5. Distributions of $\sigma_1^2$ and $\sigma_2^2$

We now begin to study the marginal posterior distributions of the variance components $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ from which inferences about the individual components can be drawn. We shall develop an approximation method by which all the distributions involved here can be reduced to the $\chi^2$ forms obtained by Tiao and Tan (1965) for the one way random effect model. This method not only gives an approximate solution to the present three-component hierarchical model but also can be applied to the general q-component hierarchical models where q is any positive integer.
so that to this degree of approximation, $\sigma_1^2$ is distributed \textit{a posteriori} as $\nu_1^{m_1} \chi_{\nu_1}^{-2}$ where

\begin{equation}
\nu_1^* = b, \quad m_1^* = \frac{\nu_2 m_2}{\sigma b}
\end{equation}

The posterior distribution of the component $\sigma_2^2$ is

\begin{equation}
p(\sigma_2^2 | y) = \frac{K(\nu_1 m_1)^{-1}(\nu_2 m_2)^{-1}}{\Pr\left\{ \nu_2, \nu_1 < \frac{m_2}{m_1} \right\}} \int_0^\infty p(\chi_{\nu_2}^{-2} = \frac{\sigma_1^2 + K \sigma_2^2}{\nu_2 m_2}) p(\chi_{\nu_1}^{-2} = \frac{\sigma_1^2}{\nu_1 m_1}) d\sigma_1^2
\end{equation}

This distribution is more complicated than that of $\sigma_1^2$, but for moderately large value of $\nu_1$ a simple and satisfactory approximation is given by

\begin{equation}
p(\sigma_2^2 | y) \approx \frac{K \frac{m_1}{\nu_2 m_2} p\left(\chi_{\nu_2}^{-2} = \frac{m_1 + K \sigma_2^2}{\nu_2 m_2}\right)}{\Pr\left\{ \chi_{\nu_2}^{-2} < \frac{\nu_2 m_2}{m_1} \right\}}
\end{equation}

Closer approximations can be obtained using the asymptotic formulae derived by these authors.

We now proceed to obtain the marginal posterior distributions of the components ($\sigma_1^2$, $\sigma_2^2$, $\sigma_3^2$) in our model.

Distributions of $\sigma_1^2$

For the posterior distribution of $\sigma_1^2$, we have from (2.5) or (2.6),
\[ p(\sigma_1^2 | \chi) = \omega(v_1 m_1)^{-1} p\left(\chi_{v_1}^{-2} = \frac{\sigma_1^2}{v_1 m_1}\right) \int_{\sigma_1^2}^{\infty} (v_2 m_2)^{-1} p\left(\chi_{v_2}^{-2} = \frac{\sigma_2^2}{v_2 m_2}\right) \times \Pr \left\{ \frac{\chi_{v_3}^2}{\sigma_{12}^2} < \frac{v_{2m_2}}{v_{3m_2}} \right\} d\sigma_{12}^{2} \] 

where

\[ \omega^{-1} = \Pr \left\{ \frac{\chi_{v_1}^2}{v_{2m_2}} > \frac{v_{1m_1}}{v_{2m_2}}, \frac{\chi_{v_3}^2}{\chi_{v_2}^2} < \frac{v_{3m_3}}{v_{2m_2}} \right\} \]

The distribution of \( \sigma_1^2 \) is thus proportional to the product of two factors, the first being the density of an inverted \( \chi^2 \) with \( v_1 \) degrees of freedom and the other a double integral of \( \chi^2 \) variables. The first factor \((v_1 m_1)^{-1} p\chi_{v_1}^{-2} = \sigma_1^2/v_1 m_1\) is the unconstrained distribution of \( \sigma_1^2 \) and the integral on the right of (5.7) is induced by the constraint \( \chi \).

Because this integral is a monotonic decreasing function of \( \sigma_1^2 \), the effect of the constraint is, as is to be expected, to pull the distribution towards the left of the unconstrained inverted \( \chi^2 \) distribution. Exact evaluation of the distribution is tedious even on an electronic computer. To approximate the distribution, one may employ the scaled \( \chi^2 \) approach by equating the first two moments of \( v_1 m_1/\sigma_1^2 \) to that of a \( \chi_b^2 \) variable. It can be verified that the r'th moment of \( v_1 m_1/\sigma_1^2 \) is

\[ \mathbb{E}[\left(\frac{v_1 m_1}{\sigma_1^2}\right)^r | \chi] = 2^r \Gamma\left(\frac{v_1}{2} + r\right) \frac{\Pr\left\{ \frac{\chi_{v_1}^2 + 2r}{\chi_{v_2}^2} > \frac{v_1 m_1}{v_{2m_2}}, \frac{\chi_{v_3}^2}{\chi_{v_2}^2} < \frac{v_{3m_3}}{v_{2m_2}} \right\}}{\Pr\left\{ \frac{\chi_{v_1}^2}{v_{2m_2}} > \frac{v_1 m_1}{v_{2m_2}}, \frac{\chi_{v_3}^2}{\chi_{v_2}^2} < \frac{v_{3m_3}}{v_{2m_2}} \right\}} \]
Evaluation of \((a, b)\) would thus involve calculating bivariate Dirichlet integrals. Although this approach simplifies the problem somewhat, existing methods for approximating such integrals still seem to be too complicated for routine practical use.

We now describe a two-stage scaled \(\chi^2\) approximation method. First, consider the joint distribution of \((\sigma_1^2, \sigma_{12}^2)\). Integrating out \(\sigma_{123}^2\) in (2.5) we get

\[
p(\sigma_1^2, \sigma_{12}^2 | y) = \omega(v_{1m_1})^{-1} \ p \left( \chi_{v_1}^{-2} = \frac{\sigma_1^2}{v_{1m_1}} \right) \ p \left( \chi_{v_2}^{-2} = \frac{\sigma_{12}^2}{v_{2m_2}} \right) \ Pr \left\{ \chi_{v_3}^2 < \frac{v_{3m_3}}{\sigma_{12}^2} \right\}
\]

\[
= \omega'(v_{1m_1})^{-1} \ p \left( \chi_{v_1}^{-2} = \frac{\sigma_1^2}{v_{1m_1}} \right) \ G(\sigma_{12}^2)
\]

where

\[
G(\sigma_{12}^2) = \frac{(v_{2m_2})^{-1} \ p \left( \chi_{v_2}^{-2} = \frac{\sigma_{12}^2}{v_{2m_2}} \right) \ Pr \left\{ \chi_{v_3}^2 < \frac{v_{3m_3}}{\sigma_{12}^2} \right\}}{\Pr \left\{ F_{v_3, v_2} < \frac{m_3}{m_2} \right\}}
\]

\[
\omega' = \omega \ Pr \left\{ F_{v_3, v_2} < \frac{m_3}{m_2} \right\}
\]

The quantity \(G(\sigma_{12}^2)\) is in precisely the same form as the posterior distribution in (5.2). Regarding the function \(G\) as if it were the distribution of \(\sigma_{12}^2\), it can therefore be approximated by the distribution of

\[
\chi_{v_2}^{-2} = \frac{\sigma_{12}^{m_3}}{v_{2m_2}}, \quad \text{where}
\]

\[
v_2' = \frac{v_2}{a_1 \ I_{\theta_1} \left( v_3/2, v_2/2 \right)}, \quad m_2' = \frac{v_2 m_3}{a_1 v_2}
\]

(5.10)
so that $\sigma_1^2$ is approximately distributed as $\nu_{m_1} \chi^2$ where

\[
v'_1 = \frac{\nu_1}{a_2} \frac{I_{\theta_2} \left( \frac{\nu_1'}{2}, \frac{\nu_1}{2} + 1 \right)}{I_{\theta_2} \left( \frac{\nu_1'}{2}, \frac{\nu_1}{2} \right)}, \quad m'_1 = \frac{\nu_{m_1}}{a_2 \nu'_1}
\]

(5.13)

\[
a_2 = \left( \frac{\nu_1}{2} + 1 \right) \frac{I_{\theta_2} \left( \frac{\nu_1'}{2}, \frac{\nu_1}{2} + 2 \right)}{I_{\theta_2} \left( \frac{\nu_1'}{2}, \frac{\nu_1}{2} + 1 \right)} - \frac{\nu_1}{2} \frac{I_{\theta_2} \left( \frac{\nu_1'}{2}, \frac{\nu_1}{2} + 1 \right)}{I_{\theta_2} \left( \frac{\nu_1'}{2}, \frac{\nu_1}{2} \right)}, \quad \theta_2 = \frac{\nu_{m_1}' \nu_1}{\nu_{m_1}' \nu_1 + \nu_{m_0}}
\]

Then making use of an incomplete beta function table, the quantities $(a_1, \nu'_2, m'_2, a_2, \nu_1, m_1)$ can be conveniently calculated from which the posterior distribution of $\sigma_1^2$ is approximately determined. For the example introduced in Section 3, we find $(a_1 = 0.865, \nu'_2 = 12.35, m'_2 = 11.51; a_2 = 1.0, \nu_1 = 20.0, m_1 = 1.04)$ so that a posteriori the variance $\sigma_1^2$ is approximately distributed as $(20.87)\chi^2$ with $20$ degrees of freedom.

In this example the effect of the constraint $C = \sigma_1^2 < \sigma_{12}^2 < \sigma_{123}^2$ is seen to be negligible and inferences about $\sigma_1^2$ can be based upon the unconstrained distribution $(\nu_{m_1})^{-1} p(\chi^2_{\nu_1} = \sigma_1^2 / \nu_{m_1})$. This is to be expected because for this example the mean square $m_2$ is much larger than the mean square $m_1$.

**Distribution of $\sigma_2^2$**

Using the result in (5.6), it follows from (5.12) that the posterior distribution of $\sigma_2^2$ is approximately

\[
(\sigma_2^2 | \mathbf{y}) \sim \frac{K}{\nu_{m_2}'} \frac{\nu_{m_2}' \nu_{m_1}'}{\nu_{m_2}' \nu_{m_1}'} \frac{\nu_{m_1}'}{\nu_{m_0}'} p\left( \chi^2_{\nu_2} = \frac{m_1 + K \sigma_2^2}{\nu_{m_0}'} \right) \frac{1}{\text{Pr}\left\{ \chi^2_{\nu_2} < \frac{\nu_{m_1}'}{\nu_{m_0}' m_1} \right\}}
\]

(5.14)
To this degree of approximation, then, the quantity \( \frac{m_1 + K \sigma_2^2}{\nu_2' m_2'} \) has the inverted \( \chi^2 \) distribution with \( \nu_2' \) degrees of freedom truncated from below at \( m_1 / \nu_2' \). For our example, the quantity \( 0.007 + 0.014 \sigma_2^2 \) thus behaves like an inverted \( \chi^2 \) variable with 12.35 degrees of freedom truncated from below at .007. Posterior intervals about \( \sigma_2^2 \) can then be calculated from a table of \( \chi^2 \) probabilities.

6. A Bayesian Solution to the Pooling Problem

When standard sampling theory is used in the analysis of hierarchical design models such as the one in (1.1), a problem arises when the mean square \( m_2 \) is of about the same size as the mean squares \( m_1 \). A test of significance having failed to show a real component \( \sigma_2^2 \), it is sometimes argued that the two mean squares should be pooled to give a better estimate of \( \sigma_1^2 \). If such a policy were adopted, then the estimate of \( \sigma_1^2 \) would be \( m_1 \) if \( m_2 / m_1 \) is large and \( (\nu_1 m_1 + \nu_2 m_2) / (\nu_1 + \nu_2) \) if \( m_2 / m_1 \) is "close to 1". We now discuss a Bayesian result in the light of this practice.

Two-Component Model

To simplify the analysis, we shall first consider the special case that \( I = 1 \) and \( \sigma_3^2 = 0 \), i.e., a two-component random effect model. In this case, the quantity \( \nu_1 m_1 / \sigma_1^2 \) is approximately distributed as \( a \chi^2 \) where \( (a, b) \) are given in (5.4). This approximation can alternatively be written as

\[
\frac{\nu_1 m_1 + \lambda \nu_2 m_2}{\sigma_1^2} \approx \chi^2_{\nu_1 + 8\nu_2}
\]

(6.1)

where
\[ \frac{m_1}{\sigma_1^2} \rightarrow \frac{\chi^2}{20} \text{ i.e. } \frac{\sigma_1^2}{m_1} \rightarrow 20 \chi^2_{20} \]

so that

\[ E\left(\frac{\sigma_1}{m_1}\right) = 1.111 \text{ and } \text{Var}\left(\frac{\sigma_1}{m_1}\right) = 0.154 \]

Thus, the additional evidence about \( \sigma_1^2 \) coming from \( m_2 \) and \( m_3 \) indicates that \( \sigma_1^2 \) is considerably smaller than would have been expected using \( m_1 \) alone. In addition, the variance of the distribution of \( \sigma_1^2/m_1 \) is seen to be only about half of that it would have been had only the evidence from \( m_1 \) been used.

7. Posterior Distribution of \( \sigma_3^2 \)

We consider in this section the posterior distribution of \( \sigma_3^2 \) and a method which can be used to approximate this distribution. To facilitate the derivation, we first return to the joint distribution in (5.1) for the two-component model and discuss the marginal posterior distribution of \( \sigma_{12}^2 = \sigma_1^2 + K\sigma_2^2 \). We obtain

\[ p(\sigma_{12}^2 | \mathbf{y}) = \frac{(\nu_2 \nu_2)^{-1} p(\chi^2_{\nu_2} = \frac{\sigma_{12}^2}{\nu_2 \nu_2}) \text{Pr}\{\chi^2_{\nu_1} > \frac{v_{1, m_1}}{\frac{v_{1, m_1}}{\sigma_{12}^2}}\}}{\text{Pr}\{F_{\nu_2, \nu_1} < \frac{m_1}{m_1}\}} \quad (7.1) \]

This distribution is proportional to the product of two factors, the first the density of an inverted \( \chi^2 \) with \( \nu_2 \) degrees of freedom and the other the left tail probability of a \( \chi^2 \) variable. This is in contrast to the distribution of \( \sigma_1^2 \) in (5.2) for which the probability integral involved is a right tail probability of a \( \chi^2 \) variable. By straightforward integration, we find that the \( r^{th} \) moment of \( x = v_{2, m_2}/\sigma_{12}^2 \) is
\[ E(x^r) = 2^r \frac{\Gamma\left(\frac{\nu_2}{2} + r\right)}{\Gamma\left(\frac{\nu_2}{2}\right)} \frac{I_\theta\left(\frac{\nu_2}{2} + r, \frac{\nu_1}{2}\right)}{I_\theta\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right)} \quad \Rightarrow \quad \theta = \frac{\nu_2^m \nu_2}{\nu_1^m + \nu_2^m} \]  

(7.2)

and that the moment generating function is

\[ M_x(t) = (1 - 2t)^{-\frac{\nu_2}{2}} \frac{\text{Pr}\left\{ \frac{\nu_2}{\nu_1} < \frac{m_2}{m_1} (1 - 2t) \right\}}{\text{Pr}\left\{ \frac{\nu_2}{\nu_1} < \frac{m_2}{m_1} \right\}} \]  

(7.3)

It is clear that when \( m_2/m_1 \) is large, \( M_x(t) \) tends to \((1 - 2t)^{-\frac{\nu_2}{2}}\) so that the distribution of \( x \) tends to the \( \chi^2 \) distribution with \( \nu_2 \) degrees of freedom. Further, even for intermediate value of \( m_2/m_1 \) (not close to zero), the moment generating function \( M_x(t) \) still behaves nearly like that of a \( \chi^2 \) with \( \nu_2 \) degrees of freedom provided \( \nu_1 \) is large. This is because when \( \nu_1 \) is large, both \( \text{Pr}\{\nu_2, \nu_1 < (m_2/m_1)(1 - 2t)\} \) for \( t \) in some interval \((-\delta, \delta)\) and \( \text{Pr}\{\nu_2, \nu_1 < (m_2/m_1)\} \) are close to unity so that \( M_x(t) \) is again close to \((1 - 2t)^{-\frac{\nu_2}{2}}\).

This suggests that the distribution of \( x \) might be approximated by that of a scaled \( \chi^2 \) variable, say, \( c \chi^2_d \). By equating the first two moments, we find

\[ c = \left(\frac{\nu_2}{2} + 1\right) \frac{I_\theta\left(\frac{\nu_2}{2} + 2, \frac{\nu_1}{2}\right)}{I_\theta\left(\frac{\nu_2}{2} + 1, \frac{\nu_1}{2}\right)} - \frac{\nu_2}{2} \frac{I_\theta\left(\frac{\nu_2}{2} + 1, \frac{\nu_1}{2}\right)}{I_\theta\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right)} \]  

(7.4)

\[ d = \frac{\nu_2}{c} \frac{I_\theta\left(\frac{\nu_2}{2} + 1, \frac{\nu_1}{2}\right)}{I_\theta\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right)}. \]

One check of the adequacy of the approximation when \( m_2/m_1 \) is not large is supplied by comparing the exact higher moments of \( x \) with those of
\(cx_d^2\). They seem to agree quite closely. For example, with \(\nu_1 = 20\), 
\(\nu_2 = 10\) and \(m_2/m_1 = 1\) the exact third moment is 537 as compared with 541 by the approximation.

\textbf{Three Three-Component Model}

Now from the joint posterior distribution of \((\sigma_{12}^2, \sigma_{23}^2, \sigma_3^2)\) in (2.6), we may integrate out \((\sigma_{12}^2, \sigma_{23}^2)\) to get the posterior distribution of \(\sigma_3^2\).

\[
p(\sigma_3^2 | y) = \int_0^\infty \int_0^\infty p(X_{\nu_1}^{-2} = \frac{\sigma_2^2}{\nu_1 m_1}) p(X_{\nu_2}^{-2} = \frac{\sigma_2^2 + K\sigma_3^2}{\nu_2 m_2})
\times p(X_{\nu_3}^{-2} = \frac{\sigma_2^2 + K\sigma_2^2 + JK\sigma_3^2}{\nu_3 m_3}) d\sigma_1^2 d\sigma_2^2
\]

(7.5)

The density function is a double integral which does not seem expressible in terms of simple forms. To obtain an approximation, we shall first consider the joint posterior distribution of \((\sigma_{12}^2, \sigma_{123}^2)\) and then deduce the result by making use of the fact that \(\sigma_{123}^2 = \sigma_{12}^2 + JK\sigma_3^2\).

From expression (2.5), we obtain

\[
p(\sigma_{12}^2, \sigma_{123}^2 | y) = \omega (v_3 m_3)^{-1} p(X_{\nu_3}^{-2} = \frac{\sigma_{123}^2}{v_3 m_3}) p(X_{\nu_2}^{-2} = \frac{\sigma_{12}^2}{v_2 m_2})
\times \Pr\left\{X_{\nu_1}^2 > \frac{v_1 m_1}{\sigma_{12}^2}\right\}
\]

\[
= \omega' (v_3 m_3)^{-1} p(X_{\nu_3}^{-2} = \frac{\sigma_{123}^2}{v_3 m_3}) A'(\sigma_{12}^2)
\]

(7.6)

where

\[
A'(\sigma_{12}^2) = \frac{(v_2 m_2)^{-1} \Pr\left\{X_{\nu_2}^{-2} = \frac{\sigma_{12}^2}{v_2 m_2} \right\} \Pr\left\{X_{\nu_1}^{-2} > \frac{v_1 m_1}{\sigma_{12}^2}\right\}}{\Pr\left\{F_{v_2, v_1} < \frac{m_2}{m_1}\right\}}
\]
\[ \omega = \omega \Pr \left( \frac{v_2}{v_1} < \frac{m_2}{m_1} \right) \]

and \(\omega\) is given in (5.7). The function \(G'(\sigma^2_{12})\) is in exactly the same form as the distribution of \(\sigma^2_{12}\) in (7.1) for the two-component model. Making use of the result in (7.4), we get

\[
p(\sigma^2_{12}, \sigma^2_{123} | y) \propto \frac{(v_3m_3)^{-1} p(x_2^{-2} = \frac{\sigma^2_{123}}{v_3m_3}) (v_2m_2^{-1} p(x_2^{-2} = \frac{\sigma^2_{12}}{v_2m_2})}{\Pr \left\{ \frac{v_3}{v_2} < \frac{m_3}{m_2} \right\}}
\]

(7.7)

where

\[
v_2'' = \frac{\nu_2''}{a_3} \frac{I_{\theta_3}(\frac{v_2''}{2} + 1, \frac{v_1}{2})}{I_{\theta_3}(\frac{v_2''}{2}, \frac{v_1}{2})}, \quad m_2 = \frac{v_2m_2''}{a_3 v_2''}
\]

(7.8)

\[
a_3 = \left( \frac{\nu_2}{2} + 1 \right) \frac{I_{\theta_3}(\frac{v_2}{2} + 2, \frac{v_1}{2})}{I_{\theta_3}(\frac{v_2}{2} + 1, \frac{v_1}{2})} - \frac{\nu_2 I_{\theta_3}(\frac{v_2}{2} + 1, \frac{v_1}{2})}{2 I_{\theta_3}(\frac{v_2}{2}, \frac{v_1}{2})}, \quad \theta_3 = \frac{v_2m_2''}{v_1m_2'' + v_2m_2''}
\]

from which it follows that the posterior distribution of \(\sigma^2_3\) is approximately

\[
p(\sigma^2_3 | y) \propto \frac{JK(v_3m_3)^{-1}(v_2m_2'')^{-1}}{\Pr \left\{ \frac{v_3}{v_2} < \frac{m_3}{m_2} \right\}} \int_0^\infty p\left( x_3^{-2} = \frac{\sigma^2_{12} + JK_3}{v_3m_3} \right)
\]

(7.9)

\[
\times p\left( x_2'' = \frac{\sigma^2_{12}}{v_2m_2''} \right) d\sigma^2_{12}
\]

The reader will note that the distribution in (7.9) is in precisely the same form as the posterior distribution of \(\sigma^2_2\) in (5.5) for the two-component model. Provided \(v_2''\) is moderately large, we can thus
express the distribution of $\sigma_3^2$ approximately in the form of a truncated inverted $\chi^2$ distribution,

$$p(\sigma_3^2 | y) \simeq \frac{1}{\text{Pr}\left\{ \chi^2_{\nu_3 m_3} > \frac{m_2''}{\nu_3 m_3} \right\}} \frac{\text{JK}(m_2'', \nu_3 m_3)}{\chi^2_{\nu_3 m_3}} p\left( \chi^2_{\nu_3 m_3} = \frac{m_2'' + \text{JK}\sigma_3^2}{\nu_3 m_3} \right)$$

(7.10)

To this degree of approximation, the quantity $(m_2'' + \text{JK}\sigma_3^2)/\nu_3 m_3$ is distributed as an inverted $\chi^2$ variable with $\nu_3$ degrees of freedom truncated from below at $m_2''/\nu_3 m_3$, from which Bayesian intervals for $\sigma_3^2$ can be determined. For the set of data considered in Section 3, we find $(a_3 = 1.0, \nu_2'' = 10.0, m_2'' = 12.29)$ so that the quantity $0.051 + 0.017 \sigma_3^2$ is approximately distributed as $\chi^2$ with 9 degrees of freedom truncated from below at the point 0.051.

9. A Summary of the Approximating Posterior Distribution of $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$

For convenience in calculation, Table 4 provides a short summary of the quantities (in addition to those appeared in the usual analysis of variance table) needed for approximating the individual posterior distributions of the variance components $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. The numerical values shown are those for the set of data introduced in Section 3.

9. Generalization to q-component hierarchical design model

The preceding analysis of the three component model and in particular the approximation methods for the posterior distributions of the individual variance components can be readily extended to the general q-component hierarchical model

$$y_{i_1 \ldots i_q} = \mu + a_{i_1} + b_{i_2 i_1} + \ldots + c_{i_q i_{q-1}} + \ldots + e_{i_q \ldots i_1}$$

$$i_q = 1, \ldots, I_q ; \ldots ; i_1 = 1, \ldots, I_1$$

(9.1)
<table>
<thead>
<tr>
<th></th>
<th>( m_1 = 1.04 )</th>
<th>( m_2 = 12.29 )</th>
<th>( m_3 = 26.70 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( v_1 = 20 )</td>
<td>( v_2 = 10 )</td>
<td>( v_3 = 9 )</td>
</tr>
<tr>
<td></td>
<td>( K = 2 )</td>
<td>( JK = 4 )</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( \theta_1 = 0.662 )</td>
<td>( a_1 = 0.865 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \nu_1' = 12.35 )</td>
<td>( m_1' = 11.51 )</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( \theta_2 = 0.872 )</td>
<td>( a_2 = 1.0 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( v_1' = 20.0 )</td>
<td>( m_1' = 1.04 )</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>( \sigma_1 \sim \nu_1' m_1' )</td>
<td>( \chi^{-2}_{v_1'} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \sigma_1 \sim (20.8) )</td>
<td>( \chi^{-2}_{20} )</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>( \frac{m_1 + JK \sigma_2^2}{\nu_2' m_2} \sim \chi^{-2}_{v_2'} )</td>
<td>truncated from below at ( \frac{m_1}{\nu_2' m_2} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( .007 + .014 \sigma_2^2 \sim \chi^{-2}_{12.35} )</td>
<td>truncated from below at 0.007</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>( \theta_3 = 0.854 )</td>
<td>( a_3 = 1.0 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( v_2'' = 10.0 )</td>
<td>( m_2'' = 12.29 )</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>( \frac{m_2'' + JK \sigma_3^2}{\nu_3' m_3} \sim \chi^{-2}_{v_3'} )</td>
<td>truncated from below at ( \frac{m_2''}{\nu_3' m_3} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( 0.051 + 0.017 \sigma_3^2 \sim \chi^{-2}_9 )</td>
<td>truncated from below at 0.051</td>
<td></td>
</tr>
</tbody>
</table>
where \( y_{i_1 \ldots i_q} \) are the observations, \( \mu \) is a common location parameter, and \( a_{i_1 \ldots i_q}, \ldots, e_{i_1 \ldots i_q} \) are \( q \) different kinds of random effects. Assuming that these effects are normally and independently distributed with zero means and variances \( (\sigma^2_q, \sigma^2_{q-1}, \ldots, \sigma^2_t, \ldots, \sigma^2_1) \) and following the argument in Section 2, it is readily shown that the joint posterior distribution of the variance components is

\[
p(\sigma^2_1, \ldots, \sigma^2_q | y) = (I_{q-1} I^{q-2}, \ldots, I_{q-1}) p(\sigma^2_1, \sigma^2_{12}, \ldots, \sigma^2_{12 \ldots t}, \ldots, \sigma^2_{12 \ldots q | y})
\]

with

\[
\sigma^2_{12} = \left( \sigma^2_1 + I_{1} \sigma^2_2 \right) \ldots \sigma^2_{12 \ldots t} = \left( \sigma^2_{12 \ldots (t-1)} + I_{1} \ldots I_{t-1} \sigma^2_t \right) \ldots
\]

\[
\sigma^2_{12 \ldots q} = \left( \sigma^2_{12 \ldots (q-1)} + I_{1} \ldots I_{q-1} \sigma^2_q \right).
\]

\[
p(\sigma^2_1, \ldots, \sigma^2_{12 \ldots t}, \ldots, \sigma^2_{12 \ldots q | y}) = 
\]

\[
(v_{1m_1})^{-1} p \left( x_{v_1}^{-2} = \frac{\sigma^2_1}{v_{1m_1}} \right) \ldots (v_{tm_t})^{-1} p \left( x_{v_t}^{-2} = \frac{\sigma^2_{12 \ldots t}}{v_{tm_t}} \right) \ldots (v_{qm_q})^{-1} p \left( x_{v_q}^{-2} = \frac{\sigma^2_{12 \ldots q}}{v_{qm_q}} \right)
\]

\[
Pr \left\{ \frac{x_{v_1}}{\sigma^2_{12 \ldots t}} > \frac{v_{1m_1}}{v^2_m_2}, \ldots, \frac{x_{v_t}}{\sigma^2_{12 \ldots t}} > \frac{v_{tm_t}}{v^2_m_t}, \ldots, \frac{x_{v_q}}{\sigma^2_{12 \ldots q}} > \frac{v_{qm_q}}{v^2_m_q} \right\}
\]

\[
\sigma^2_1 < \sigma^2_{12} < \ldots < \sigma^2_{12 \ldots t} < \ldots < \sigma^2_{12 \ldots q}
\]

where the \( m \)'s and the \( v \)'s are the corresponding mean squares and the degrees of freedom. The distributions in (9.3) and (9.2) parallel exactly those in (2.6) and (2.7).

In principle, the marginal posterior distribution of a particular variance component, say, \( \sigma^2_t \), can be obtained from (9.2) simply by integrating out \( \sigma^2_1, \ldots, \sigma^2_{t-1}, \sigma^2_{t+1}, \ldots, \sigma^2_q \). In practice, however, this involves calculation of \( (q-1) \)-dimensional integral for each value of \( \sigma^2_t \).
and would be difficult even on a fast computer. A simple approximation to the distribution of $\sigma_t^2$ can be obtained by first considering the joint distribution $p(\sigma_{l2\ldots(t-1)}^2, \sigma_{l2\ldots t}^2 | y)$. The latter distribution is obtained by integrating (9.3) over $(\sigma_{l2\ldots q}^2, \ldots, \sigma_{l2\ldots(t+1)}^2)$ and $(\sigma_{l2\ldots(t-2)}^2)$. It seems clear that the set $(\sigma_{l2\ldots q}^2, \ldots, \sigma_{l2\ldots(t+1)}^2)$ can be eliminated by repeated applications of the $\chi^2_b$ approximation in (5.4). The effect of each integration is merely to change the values of the mean square $m$ and the degrees of freedom $\nu$ of the succeeding variable. Similarly, the set $(\sigma_{l2\ldots(t-2)}^2)$ can be approximately integrated out upon repeated applications of the $\chi^2_d$ method in (7.4). Thus, the distribution of $(\sigma_{l2\ldots(t-1)}^2, \sigma_{l2\ldots t}^2)$ takes the approximate form

$$p(\sigma_{l2\ldots(t-1)}^2, \sigma_{l2\ldots t}^2 | y) \\ \approx \frac{(\nu_{t-1} m_{t-1}^l)^{-1} (\nu_{t-1}^l m_{t-1}^l)^{-1} \Pr \left\{ \frac{V_{l1} \nu_{t-1}}{m_{t-1}^l} < \frac{m_{t}^l}{m_{t-1}^l} \right\} (\nu_{t} m_{t}^l)^{-1} (\nu_{t}^l m_{t}^l)^{-1}}{(\nu_{t} m_{t}^l)^{-1} (\nu_{t}^l m_{t}^l)^{-1}}$$

(9.4)

Noting that $\sigma_{l2\ldots t}^2 = \sigma_{l2\ldots(t-1)}^2 + I_{t-1} \sigma_t^2$, the corresponding posterior distribution of $(\sigma_{l2\ldots(t-1)}^2, \sigma_t^2)$ is then of exactly the same form as the posterior distribution of $(\sigma_1^2, \sigma_2^2)$ in (5.1) for the two component model. The marginal distribution of $\sigma_t^2$ can then be approximated using the truncated inverted $\chi^2$ form (5.6) or the asymptotic formulae previously cited. Finally, in the case $t = 2$, so that $\nu_t = \nu_1^l$, $m_t = m_1^l$ and the posterior distribution of the error variance $\sigma_1^2$ can be approximated by that of an inverted $\chi^2$ distribution as in the one-way model.
References


Box, G.E.P. and Tiao, G.C. (1965) Multiparameter problems from a 

Hill, B.M. (1965). Inference about variance components in the one-way 


models in the analysis of variance. I. posterior distribution of 

models in the analysis of variance. II. effect of autocorrelated 
errors. To appear in Biometrika, December.
Figure 1

Posterior distribution of $\frac{\sigma_2^2}{\sigma_1^2}$

exact distribution

unconstrained distribution
Figure 3

Values of $(\Lambda, \delta)$ as functions of $m_2^2$

$(l = 7, k = 5)$