Department of Statistics
The University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT No. 117
May, 1967

BAYESIAN ESTIMATION OF
MEANS FOR THE RANDOM EFFECT MODEL

by

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

This research was supported in part by the National Science Foundation under Grant No. GP-6519.
BAYESIAN ESTIMATION OF MEANS FOR THE RANDOM EFFECT MODEL

by

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

University of Wisconsin

Summary

The problem of estimating the means in the one-way random effect model $y_{jk} = \theta_j + e_{jk}$ is considered from a Bayesian viewpoint. Posterior distributions of the $\theta_j$'s are obtained under the assumption that $\theta_j$ are independently drawn from a Normal population $N(0, \sigma^2_\theta)$ and that $e_{jk}$ are independent random errors having a $N(0, \sigma^2_e)$ distribution. It is shown that the posterior distribution of the $\theta_j$'s are clustered more closely together than are the corresponding distributions for a fixed effect model. A numerical example is given.
1. Introduction

Suppose we have data \( y_{jk} \) arranged in a one-way classification with \( J \) groups and \( K \) observations per groups. For example, the following data taken from Davies (1961, p. 105) show \( K = 5 \) laboratory determinations made on samples from each of \( J = 6 \) batches.

<table>
<thead>
<tr>
<th>Batch</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>145</td>
<td>140</td>
<td>195</td>
<td>45</td>
<td>195</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>155</td>
<td>150</td>
<td>40</td>
<td>230</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>90</td>
<td>205</td>
<td>195</td>
<td>115</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>160</td>
<td>110</td>
<td>65</td>
<td>235</td>
<td>80</td>
</tr>
<tr>
<td>Mean</td>
<td>105</td>
<td>128</td>
<td>164</td>
<td>98</td>
<td>200</td>
<td>70</td>
</tr>
</tbody>
</table>

In the analysis of such data, it has sometimes been convenient to think in terms of either a fixed effect or of a random effect model.

Suppose the model is written

\[
y_{jk} = \theta_j + e_{jk} \quad j = 1, 2, \ldots, J \quad (1)
\]

with the errors distributed independently and Normally so that

\[
e_{jk} \sim N(0, \sigma^2)
\]

With the fixed effect model we concentrate attention on the estimation of the batch mean \( \theta_j \) regarded as separate entities of interest in themselves.

With the random effect model, on the other
hand, we think of the $\theta_j$'s as independently distributed observations from a distribution and it is about this distribution rather than about the individual $\theta$'s that we wish to learn. More specifically, it would usually be assumed that the distribution was Normal so that

$$\theta_j \sim N(0, \sigma_j^2)$$  \hspace{1cm} (2)

The main item of interest would then be $\sigma^2$, the "batch to batch" variance, and not the value of the individual batch means. The problem of estimating $\sigma^2$ and $\sigma_1^2$ in this latter situation is often called the problem of variance components, the components in this case being $\sigma_1^2$ and $\sigma^2$.

Although the analysis of both fixed effect and random effect models was originally done from a sampling theory viewpoint, it can also be tackled from the Bayesian point of view. (See for example Jeffreys (1961), Lindley (1965), Box and Tiao (1965), Tiao and Tan (1965, 1966), and Tiao and Box (1967)). Recently Lindley, in his discussion of the work of Stein (1962), has pointed out that an approach from this angle underlines the necessity for consideration of an intermediate category of problems which we now discuss.

In the usual Bayesian approach to the fixed effect analysis, the $\theta_j$ are given disperse locally uniform reference priors. Such a model seems appropriate if the means $\theta_j$ can be expected to bear no strong relationship one to another. Certainly, problems do occur where this assumption is reasonable. Thus we might be interested in comparing in the laboratory yields from six different methods of making a particular chemical product. While we know that yields of less than zero or greater than 100% of the theoretical maximum are impossible, we could be in a situation where for any one of the
processes yields from, say 10% to 90%, might be found.

Now let us return to the Davies example. It is perfectly clear that, in fact, in most if not all examples were concentrated on the estimation of \( \sigma^2 \) and \( \sigma^2 \) and not on the \( \theta_j \)'s. It is nevertheless true that interest could have attached to the estimation of the individual batch means \( \theta_j \) even in this situation where the \( \theta_j \) may be properly thought of as coming from a distribution. Thus although we might be concerned with the estimation of the \( \theta \)'s, we might nevertheless not be in a situation where we could assume that \( \theta_j \)'s had unrelated disperse priors. Instead, it might be more realistic to adopt the random effect model referred to above and consider its consequences for the estimation of the \( \theta_j \)'s.

Intuitively we would expect that because the \( \theta_j \)'s themselves came from a population, information about \( \theta_j \) will in some way be supplied by observations other than those in the \( j^{th} \) group. For instance, if in the Davies’ example we were told the sample mean for all but one of the groups, then we should expect to be able to say something about the mean of the remaining group without any observations from that group at all. One might further expect that the net effect of this additional information would be to draw the estimates of the means more closely together.

2. Estimation of Means Using the Random Effect Model

From now on we assume the random effect model defined in equations (1) and (2). The usual analysis of variance table associated with this model is given below.
\[ p(\theta, \sigma_1^2, z, y) = \frac{(K+1)J+1}{2} \frac{J-1}{z} \frac{1}{z-1} \]
\[ \times \exp \left[-\frac{1}{2\sigma_1^2} (\nu_1 m_1 + k (\sum_j (\theta_j - \bar{y}_j)^2 + \frac{z}{1-z} \sum_j (\theta_j - \bar{y})^2)) \right]. \]
(6)

Now the expression in the exponent
\[ Q(\theta) = \sum_j (\theta_j - \bar{y}_j)^2 + \frac{z}{1-z} \sum_j (\theta_j - \bar{y})^2 \]
may be written
\[ Q(\theta) = (\theta - \bar{y})' I (\theta - \bar{y}) + \frac{z}{1-z} \theta' (I - J^{-1} \frac{1}{1} \frac{1}{1}) \theta \]
where \( I \) is a column vector with each of its \( J \) elements unity. A useful identity in the reduction of sums of quadratic forms is
\[ (x-a)' A (x-a) + (x-b)' B (x-b) \]
\[ = (x-c)' (A+B) (x-c) + (a-b)' A (A+B)^{-1} B (a-b) \]
with \( c = (A+B)^{-1} (Aa+Bb) \), and provided of course that the inverse \((A+B)^{-1}\) exists.

After making the necessary substitutions, we find that
\[ Q(\theta) = k^{-1} z \nu_2 m_2 + (1-z)^{-1} (\theta - \hat{\theta}(z))' (I - z J^{-1} \frac{1}{1} \frac{1}{1}) (\theta - \hat{\theta}(z)) \]
(9)
where
\[ \hat{\theta}'(z) = (\hat{\theta}_1(z), \ldots, \hat{\theta}_J(z)), \hat{\theta}_j(z) = \bar{y}_j - z(\bar{y}_j - \bar{y}). \]

Thus (6) becomes
\[ p(\theta, \sigma_1^2, z, y) = \frac{(K+1)J+1}{2} \frac{J-1}{z} \frac{1}{z-1} \]
\[ \times \exp \left[-\frac{1}{2\sigma_1^2} (\nu_1 m_1 + z \nu_2 m_2 + (1-z)^{-1} k (\theta - \hat{\theta}(z))' (I - z J^{-1} \frac{1}{1} \frac{1}{1}) (\theta - \hat{\theta}(z))) \right] \]
(10)
The distribution of $\theta_j$ conditional on $z$

On integrating (7) with respect to $\sigma^2_j$, we have

$$p(\theta_j | z) = \int \frac{1}{\sigma^2_j} \left(\frac{1}{\sigma^2_j} + z \frac{v_{2m_2}}{\nu_1 + v_{2m_2}} + \frac{\nu_1 m_1 + z v_{2m_2}}{\nu_1 + v_{2m_2}} \frac{1}{I - \frac{K(\Theta - \hat{\Theta}(z))^\prime(I - zJ^{-1}11^\prime)(\Theta - \hat{\Theta}(z))}{s^2(z)}} \right)^{(\nu_1 + v_{2m_2} + J)}$$

(11)

It follows at once that conditional on $z$

(i) the distribution of $\theta_j$ is the $J$-dimensional $t$ distribution

having $\nu_1 + v_{2m_2}$ degrees of freedom defined by

$$p(\theta_j | z, y) \propto \left\{ 1 + \frac{K(\Theta - \hat{\Theta}(z))^\prime(I - zJ^{-1}11^\prime)(\Theta - \hat{\Theta}(z))}{(\nu_1 + v_{2m_2}) s^2(z)} \right\}^{(\nu_1 + v_{2m_2} + J)}$$

(12)

where

$$s^2(z) = \frac{v_{1m_1} + z v_{2m_2}}{\nu_1 + v_{2m_2}} (1 - z)$$

or equivalently,

$$p(\theta_j | z, y) \propto \left\{ 1 + \frac{\nu_1 + v_{2m_2}}{(\nu_1 + v_{2m_2}) s^2(z)} \right\}^{(\nu_1 + v_{2m_2} + J)}$$

(13)

(ii) the $\theta_j$'s have means $\hat{\theta}_j(z)$, $j = 1, \ldots, J$, common variance,

and common covariance with one another

$$\text{Var} (\theta_j | z) = \frac{\nu_1 + v_{2m_2}}{K(\nu_1 + v_{2m_2})} \left[ 1 + \frac{z}{1 - \frac{K(\Theta - \hat{\Theta}(z))^\prime(I - zJ^{-1}11^\prime)(\Theta - \hat{\Theta}(z))}{s^2(z)} \right] s^2(z),$$

(14)

$$\text{Cov} (\theta_i, \theta_j | z) = \frac{\nu_1 + v_{2m_2}}{K(\nu_1 + v_{2m_2})} \left[ 1 + \frac{z}{1 - \frac{K(\Theta - \hat{\Theta}(z))^\prime(I - zJ^{-1}11^\prime)(\Theta - \hat{\Theta}(z))}{s^2(z)} \right] s^2(z)$$

(iii) the marginal distribution of $\theta_j$ is a $t$-distribution having

$\nu_1 + v_{2m_2}$ degrees of freedom

$$\frac{\nu_1 + v_{2m_2}}{\nu_1 + v_{2m_2} - 2} \left[ \frac{z}{1 - \frac{K(\Theta - \hat{\Theta}(z))^\prime(I - zJ^{-1}11^\prime)(\Theta - \hat{\Theta}(z))}{s^2(z)}} \right] \frac{1}{(\nu_1 + v_{2m_2}) \text{Var} (\theta_j | z)^2}$$

(15)
(iv) the marginal distribution of a particular difference 

\[ \theta_i - \bar{\theta} \] is a t distribution having \( v_1 + v_2 \) degrees of freedom

\[ t_{v_1 + v_2} \]

\[ (\bar{y}_1 - y_i) - (1-z)(\bar{y}_1 - \bar{y}) \]

\[ (v) \] the joint distribution of the deviations \( \theta_i - \bar{\theta} \), \( \theta_2 - \bar{\theta}, \ldots, \theta_J - \bar{\theta} \) is given by the \( J-1 \) dimensional t distribution having \( v_1 + v_2 \) degrees of freedom defined by

\[ p(\theta - \bar{\theta} \mid y) \propto \left\{ 1 + \frac{K}{(v_1 + v_2) s^2} \left[ (\theta - \bar{\theta}) - (1-z)(\bar{y}_1 - \bar{y}) \right] \right\}^{2 - \frac{(v_1 + v_2 + J - 1)}{2}} \]

\[ (v) \]

The marginal distribution of \( z \)

Now the marginal distribution of \( z = \frac{\sigma^2}{\sigma^2_{12}} \) (see, for example, Tiao and Box, [1967]), is given by

\[ p(z \mid y) = \frac{m_2}{m_1} \frac{p(F_{v_2, v_1} = \frac{m_2}{m_1} \mid z)}{\text{Pr}(F_{v_2, v_1} < \frac{m_2}{m_1})} \quad 0 < z < 1 \]

where \( p(F_{v_2, v_1}) \) is the density function of an F variable with \((v_2, v_1)\) degrees of freedom. The \( r^{th} \) moment of \( z \) is, therefore,

\[ \nu_r = E(z^r) = \frac{B(\frac{v_2}{2} + r, \frac{v_1}{2} - r)}{B(\frac{v_2}{2}, \frac{v_1}{2})} \frac{I_{v_2, \frac{v_1}{2}}(\frac{v_2}{2} + r, \frac{v_1}{2} - r)}{I_{v_2, \frac{v_1}{2}}(\frac{v_2}{2}, \frac{v_1}{2})} \left( \frac{m_2 m_1}{m_2} \right)^r \]

\[ (19) \]

where \( x = \frac{v_2 m_1}{v_1 m_1 + v_2 m_2} \) and \( I_x(p, q) \) is the usual incomplete beta function.
The marginal distribution of $\theta_j$

The unconditional distribution of the $\theta_j$'s may now be obtained by integrating the conditional distribution for given $z$ with $p(z|y)$ as weight function. Thus

$$p(\theta_j|y) = \int_0^1 p(\theta|x, y) p(z|y) dz$$

(20)

Also to obtain the moments of the $\theta_j$'s we can take the expected values over $z$ of the conditional moments. In particular, we find

$$E(\theta_j|y) = \hat{\theta}_j = \bar{y}_j - \mu_1 \bar{y}_j - \bar{y}_j$$

(21)

$$\text{Var}(\theta_j|y) = E \text{Var}(\theta_j|z) + E \left( \hat{\theta}_j(z) - \hat{\theta}_j \right)^2$$

(22)

$$= \frac{J(v_1m_1 - v_1m_2) - (J-1)(\mu_1^2 v_1m_1 + \mu_2^2 v_2m_2)}{J K (v_1 + v_2 - 2)} + (\bar{y}_j - \bar{y})^2 \mu_2$$

$$u_2 = \mu_2 - \mu_1^2.$$

$$\text{Cov}(\theta_i, \theta_j|y) = E \text{Cov}(\theta_i, \theta_j|z) + E(\hat{\theta}_i(z) - \hat{\theta}_i)(\hat{\theta}_j(z) - \hat{\theta}_j)$$

(23)

$$= \frac{\mu_1^2 v_1m_1 + \mu_2^2 v_2m_2}{JK(v_1 + v_2 - 2)} + (\bar{y}_j - \bar{y})(\bar{y}_j - \bar{y}) \mu_2$$

These may be compared with the corresponding results which would be obtained if the fixed effect model (supposing disperse priors for log $\sigma$ and the $\theta_j$'s independently) were appropriate. These are

$$E(\theta_i|y) = \bar{y}_j, \quad \text{Var}(\theta_j|y) = \frac{m_1 - v_1}{K v_1 - 2}, \quad \text{Cov}(\theta_i, \theta_j|y) = 0.$$  

(24)

A numerical example

We now illustrate these results for the Davies' example quoted earlier. For these data, we have
\[ J = 6, \ K = 5, \ \nu_1 = 24, \ \nu_2 = 5, \ \bar{m}_1 = 2,451, \ \bar{m}_2 = 11, \ 271 \]
\[ \bar{y} = 15.275, \ \nu_1' = 0.233, \ \nu_2' = 0.02619. \]

In the table below, to facilitate comparison we have arranged and numbered the groups in order of magnitude of the group means and we have shown the corresponding results for the fixed effect model.

**Table 3**

| Group | Random Effect | Fixed Effect | Standard Deviations (\(\text{Var}(\theta_j | \bar{y})^{1/2}\)) |
|-------|---------------|--------------|-------------------------------------------------|
|       |               | \(\bar{y}_j\) | \(\bar{y}_j\) | Random Effect | Fixed Effect | \(420.4 + (\bar{y}_j - \bar{y})^2 0.02619\) | \((534.8)^{1/2}\) |
| 1     | 83            | 70           | 22.5                                           | 23.1         |
| 2     | 108           | 98           | 21.1                                           | 23.1         |
| 3     | 110           | 105          | 20.8                                           | 23.1         |
| 4     | 128           | 128          | 20.5                                           | 23.1         |
| 5     | 156           | 164          | 21.3                                           | 23.1         |
| 6     | 186           | 200          | 23.6                                           | 23.1         |

**Correlation Matrix:**

\[
P_{ij} = \frac{\text{Cov}(\theta_i, \theta_j | \bar{y})}{\sqrt{\text{Var}(\theta_i | \bar{y}) \text{Var}(\theta_j | \bar{y})}}^{1/2}
\]

<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.14</td>
<td>0.17</td>
<td>0.05</td>
<td>-0.07</td>
<td>-0.16</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.09</td>
<td>0.05</td>
<td>-0.01</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.05</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For the fixed effect model, the $\theta_j$'s are uncorrelated. Figure 1(a) shows the actual posterior distributions for $\theta_1, \theta_2, \ldots, \theta_6$ obtained by numerical integration of the conditional $t$ distribution over the marginal distribution of $z$. Shown for comparison in Figure 1(b) are the corresponding distributions (centered about the sample means) for the fixed effects model. The numerical values in the table and in the figure clearly show the greater clustering of the distributions about $\bar{y}$ that occurs with the random model.

Figure 2 shows the distribution of $\theta_6 - \theta_1$ for the random effect model together with the corresponding distribution appropriate to the fixed effect model. In this extreme case, a very large difference is seen between the two distributions.
References


Figure 1

Posterior distributions of $\theta_j$

1 (a) Random effect model.

1 (b) Fixed effect model.

- Sample mean.
- Posterior mean
Figure 2.

Posterior distribution of $\theta_6 - \theta_1$