Posteriors in Balancing One-Way Random Effects

By Jiannong Liu and James S. Hodges

Model

Correspondence author: Jiannong Liu

Division of Biostatistics, School of Public Health, University of Minnesota

email: jiannong@biostat.umn.edu

November 16, 2001

University of Minnesota, 2221 University Ave SE, Suite 200, Minneapolis, Minnesota 55414 U.S.A.
SUMMARY

Poster or Bimodality in the Balanced One-Way Random Effects Multilevel Model

Key words: Hierarchical model, variance components, random effects, Bayesian analysis, prior distribution, inference of modality for these models.

We examine an example in detail, present a graphical display for examining bimodality, and offer an intuitive explanation of why random effects models with a half prior on the mean, this apparently simple model has bimodality of the joint and marginal posteriors for a conjugate analysis of the balanced one-way random effects models, multimodality is not well-characterized for any such model. The paper characterizes although some authors have examined posterior multimodality for specific nonflat parameterized
Introduction

Recent advances have moved Bayesian analyses of complex hierarchical models out of seminar rooms and into the larger world, even though our collective understanding of these models lags far behind. For example, bimodal posteriors are possible (Firat and Thompson 1996; Harville and Zimmermann 1996; Hoschele 1988; Mardia and Watkins 1989; O'Hagan 1996; Wakefield 1998), though so far has identified but not characterized the posterior modalities for the balanced one-way random effects model (BOWREM). For the BOWREM, both the marginal likelihood and the restricted likelihood have multiple maxima (Seber). Section 2 characterizes the posterior modalities for the BOWREM and some marginal posteriors for the BOWREM, with the main proofs deferred to an appendix. Section 3 gives intuition about how the data and prior determine modality, while Section 4 explores more elaborate models. Section 5 characterizes the modality of the joint posterior and some marginal posteriors for the BOWREM, with the main proofs deferred to an appendix. Section 6 provides intuition about how the data and prior determine modality, while Section 7 explores more elaborate models.
The balanced one-way random-effects model can be written as

\[ y_{ij} = \mu + \alpha_i + \epsilon_{ij} \]

where \( \epsilon \) is a constant, \( \alpha \) is the identity matrix of order \( N \) and \( \mu \) is an \( N \)-vector of \( 1 \)'s and \( \times \) is the Kronecker product. Unfortunately, the results to follow

\[ \left[(N+\ell I - \Theta)I^T(N+\ell I - \Theta) + q \overline{\zeta} \right] \frac{1}{\ell} \left( \Theta \ell - \log \left( N \overline{\zeta} + N \right) \right. - \left[ (\Theta \ell - \Lambda)(\Theta \ell - \Lambda) + \sqrt{\overline{\zeta}} \right] \frac{q \overline{\zeta}}{\ell} \left( \log N \right) - \log \left( (\lambda \ell \cdot \epsilon^{\ell \cdot \ell} \Theta) \right) = \left( (\lambda \ell \cdot \epsilon^{\ell \cdot \ell} \Theta) \right) \]

The log joint posterior of \( (\ell, \Theta) \) is

\[ \text{1. The modality of the joint posterior} \]

where

\[ (q \cdot a) I \sim \epsilon_a \text{ and } (Y \cdot a) I \sim \epsilon_a \text{ and } (n) \sim \text{I} \]

and we use the conjugate priors

\[ N^a \cdot \ell \cdot \ell = i \left( \epsilon_a \right) N \sim \epsilon_a \text{ and } \text{I} \]

where \( \{ \theta \} = \Theta \) and \( \{ \epsilon \} = \ell \)
Finally, to simplify notation in the theorem, define

\( \phi \) the local maximin and minimin of 

\[
\begin{align*}
\frac{\partial S}{\partial \phi} v^T + \frac{\partial S}{\partial \psi} \psi^T + \mu^T = & \quad d \\
\frac{\partial S}{\partial \phi} v + \frac{\partial S}{\partial \psi} \psi + \mu = & \quad b
\end{align*}
\]

where is the usual F-statistic. If \( \frac{1 - u}{N + 4S} = f \), then there are two real solutions

\[
\begin{align*}
\frac{(f + p)g}{\nabla^T + (f + p \psi) \psi} = \phi \\
\frac{(f + p)g}{\nabla^T - (f + p \psi) \psi} = \phi
\end{align*}
\]

The key to solving \( g(\phi) = 0 \) is

\[
\begin{align*}
\phi = (f + p)g(f + p \psi) \psi = \nabla \\
\psi = (\phi) \psi
\end{align*}
\]

and are given by the solutions of cubic equations of the form

\[
(\phi) = (f + p)g(f + p \psi) \psi = \nabla
\]

where

\[
\phi = (f + p)g(f + p \psi) \psi = \nabla
\]

and

\[
\psi = (\phi) \psi
\]

are notationally dense and we must begin with several definitions.
where \( K \neq d + f \)

\[
\begin{align*}
\{ (f + p) & e \over f + p e + \frac{e}{b} \frac{f + p e}{f + p e} + \frac{e}{b} \frac{f + p e}{f -} \} \\
\end{align*}
\]

(11)

Theorem 2.1 For the BOWREM defined by (1) and (1) has two modes otherwise it has one mode. Specifically:

If \( (\phi) > (\phi) \phi > (\phi) \phi \) has three critical points corresponding to \( 0 < \phi < (\phi) \phi \) and \( 0 < \phi < (\phi) \phi \) and \( 0 < \phi < (\phi) \phi \) then the joint posterior \( P = (\phi) \phi \) has two modes otherwise it has one mode. Specifically:

The sample sizes \( n, n \) and \( n \) that is, the sample sizes and the implicit prior

\[
\begin{align*}
\frac{f + p}{2n} &= \frac{e}{2n} \\
\frac{f + p}{2n} &= \frac{e}{2n} \\
\frac{(f + p)e}{f + p e + \frac{e}{b} \frac{f + p e}{f + p e}} &= \frac{e}{2n} \\
\end{align*}
\]

(6)

and

\[
\begin{align*}
\frac{(f + p)e}{p + (f + p) e} &= \frac{e}{2n} \\
\frac{(f + p)e}{f + p e + \frac{e}{b} \frac{f + p e}{f + p e}} &= \frac{e}{2n} \\
\end{align*}
\]

(8)

where
THE modal posterior of \( \eta \)

The modal posterior for two common marginal posteriors.

Joint and marginal posteriors can have different modality for a given prior (O'Hagan 1976). We

2.2.1. The modal posterior of \( \eta \)

\[ \phi \text{ maximizes } P \]

\[ \text{if } \nabla P(\phi) = 0 \]

\[ \phi \text{ is a mode of the joint posterior and all its modes have this form.} \]

\[ \nabla P(\phi) = 0 \]

\[ \phi \text{ is a mode of the joint posterior and all its modes have this form.} \]

\[ P(\phi) = \sum_{i=1}^{n} \delta(\phi - \phi_i) \]

\[ \text{which maximizes } \phi \]

\[ P(\phi) \]

\[ \phi \text{ is a mode of the joint posterior and all its modes have this form.} \]

\[ \nabla P(\phi) = 0 \]

\[ \phi \text{ is a mode of the joint posterior and all its modes have this form.} \]

\[ P(\phi) \]

\[ \phi \text{ is a mode of the joint posterior and all its modes have this form.} \]
for a derivation like the one in the Appendix produces necessary and sufficient conditions for

\[(81)\]

\[
\left\{ \frac{1}{\varepsilon^{2}} + \frac{1}{1 - \varepsilon^{2}} \right\} \frac{1}{1 - \varepsilon^{2}} \int \frac{1}{(1-u)\varepsilon^{2}(\varepsilon^{2} + \varepsilon^{2})} \ dx \approx (x, \varepsilon^{2}, \varepsilon^{2})dI
\]

the marginal posterior of

\[(91)\]

\[
(\varepsilon^{2}, \varepsilon^{2})_{marginal\ posterior\ of}
\]

for

\[\gamma_{2} = \gamma_{2}\]

\[\frac{1}{1 - \gamma_{2}} \gamma_{2} \gamma_{2} + (\lambda_{2} + \gamma_{2}) (1 + \gamma_{2}) - = \gamma_{2}\]

\[(\lambda_{2} + \lambda_{2}) (N\ - p)u + (\lambda_{2} + \gamma_{2}) [(1 + \gamma_{2}) + f]u = \gamma_{2}\]

\[(11)\]

\[
(\lambda_{2} + \gamma_{2}) f_{u}u = \gamma_{2}\]

for

\[\gamma_{2} = \gamma_{2}\]

\[\frac{1}{1 - \gamma_{2}} \gamma_{2} \gamma_{2} + (\lambda_{2} + \gamma_{2}) (1 + \gamma_{2}) - = \gamma_{2}\]

\[\lambda_{2} + \lambda_{2} = \lambda_{2} + \lambda_{2}\]

where

\[\varepsilon^{2} = \varepsilon^{2}\]

where \(\varepsilon^{2}\) and \(\varepsilon^{2}\) are the two modes. The conditions are which (16) has two modes. The conditions are under a derivation like the one in the Appendix produces necessary and sufficient conditions for

\[(91)\]

\[
\left\{ \frac{1}{\varepsilon^{2}} + \frac{1}{1 - \varepsilon^{2}} \right\} \frac{1}{1 - \varepsilon^{2}} \int \frac{1}{(1-u)\varepsilon^{2}(\varepsilon^{2} + \varepsilon^{2})} \ dx \approx (x, \varepsilon^{2}, \varepsilon^{2})dI
\]

\[
(\varepsilon^{2}, \varepsilon^{2})_{marginal\ posterior\ of}
\]
\[ (f + p)z/(1 - p) = \phi \]
\[ f'\sqrt{\phi S}z(\phi - 1) = \epsilon' \]
\[ p/ \left[ \frac{h_S \phi + h S + \chi z}{1} \right] = \epsilon P \]
\[ \eta' \cdot \cdots \cdot \eta' = \eta' \eta'(\phi - 1) + \eta' \eta' = \eta' \]
\[ f = \eta' \]

Furthermore, \((\epsilon' \cdot \cdots \cdot \eta' \cdot \cdots \cdot \eta') = 1 \)

Theorem 2.1

If \( f \) has an improper prior \((f(1 - p))/((1 - a)N) \)

the posterior has a finite mode at \( \phi \) when 0.0 \(< \phi < 1 \).

The following results can be derived straightforwardly by methods like those in the Appendix.

and sometimes the resulting posterior has good frequency properties. The following results can be derived straightforwardly by methods like those in the Appendix.

2.3 Posterior modality when \( f \) has an improper prior

\[ (1 - p)z = \epsilon \]

\[ \mu = \left[ z/(1 + N) - p \right] u + \left( \frac{h_S + \epsilon S + \chi z}{1} \right) (1 + p) - \epsilon \]

\[ \left( q + \frac{q}{S} \right)/N - p \] \[ u + \left( \frac{h_S + \chi z}{1} \right) (1 + q + f) = \epsilon \]

\[ (\frac{h_S + \chi z}{1 - f}) u = \epsilon \]

\[ \text{for } f = \frac{z}{\chi} = \frac{S}{\xi} + \xi \]

\[ \frac{1 - (S + \xi)}{\chi + \xi} = \frac{\xi}{\chi} \]

\[ \frac{1 - (S + \xi)}{\chi + \xi} = \frac{\xi}{\chi} \]
When a \( \neq 1 \) and \( \chi = 0 \), the theorem reduces to the case in O'Hagan (1989). Similarly:

\[
\frac{1 + \nu \sigma^2}{\nu + \nu \sigma^2} \propto \nu \sigma^2
\]

for the two marginal posteriors (16) and (18).
For any \( \phi \neq 0 \), if \( \phi \) is large, \( P \) is more likely to have one mode, for which \( \phi \) is close to 1.

2. If \( \lambda \) is large and \( \phi \) is small, \( P \) is more likely to have one mode, for which \( \phi \) is close to \( 1 \).

or in the middle depends on \( \frac{\phi}{\lambda} \) and \( \frac{\lambda}{\phi} \) or \( \lambda \) and \( \phi \) or \( \frac{\phi}{\lambda} \) and \( \frac{\lambda}{\phi} \).

1. If both \( \phi \) and \( \lambda \) are large, \( P \) is more likely to have one mode. Whether \( \phi \) is close to 0 or 1.

B. Priors: Varying \( \phi \) and/or \( \lambda \) for fixed \( \alpha \), \( \beta \), \( \gamma \), \( \delta \), \( \sigma \), and \( \tau \).

1. If both \( \phi \) and \( \lambda \) are small, \( P \) is more likely to have one mode, for which \( \phi \) is close to 0.

2. If \( \phi \) is large and \( \lambda \) is small, \( P \) is more likely to have one mode, for which \( \phi \) is close to 1.

3. If \( \lambda \) is large and \( \phi \) is small, \( P \) is more likely to have one mode, for which \( \phi \) is close to 0.

1. If both \( \phi \) and \( \lambda \) are large, the joint posterior is more likely to have two modes if \( \phi \) is close to 0.\n
2. If both \( \phi \) and \( \lambda \) are small, the joint posterior is more likely to have two modes if \( \phi \) is close to 1.

3. If \( \lambda \) is large and \( \phi \) is small, the joint posterior is more likely to have two modes if \( \phi \) is close to 0.

4. If both \( \phi \) and \( \lambda \) are small, the joint posterior is more likely to have two modes if \( \phi \) is close to 1.

\[ \text{Avoid messy and unreadable spurious limits.} \]

We now consider the joint posterior's tendency toward unimodality depending on whether posteriors are unimodal.

The two unimodal posteriors, that is, when one of \( \alpha \), \( \beta \), or \( \gamma \) or \( \delta \) is made large enough, the result is no single prior guarantees a unimodal posterior. Similar results are available for any fixed \( \alpha \), \( \beta \), \( \gamma \), \( \delta \), \( \sigma \), and \( \tau \).
Groups are separated more for given $N$ and $n$. However, it does appear that as $\frac{1}{x}\frac{B}{S}$ increases, $P$ is more likely to have two modes if $N$ or $a$ increases. This is not generally true; Lucas [1993] catalogued prior and likelihood pairs that produce uni-modal posteriors when they conflict (this is not generally true; Lucas [1999, 1993]).

To see the sense of this, recall that $\mathbb{E}_\frac{S}{2}B(1) = \frac{1}{2}(1 - \frac{u}{y})$, so if $\frac{B}{S}$ is large enough, $P$ is unimodal, but if $\frac{B}{S}$ is large enough, $P$ is bimodal. Thus, in case A(1), for given $a$ and $c$, if $\frac{B}{S}$ is small enough, it is easy to show that bimodality occurs in cases $A(1)$ and $C(1)$ because the data conflict with the prior

$P$ will be unimodal

i. If $N$ is increased with $n$, $\frac{B}{S}$ and $\frac{u}{y}$ become

ii. If $\frac{B}{S}$ increases and $a$ is held fixed, $\frac{u}{y}$ will become

iii. If $u/\frac{B}{S}$ is held fixed, $\frac{1 - u}{N}$ and $\frac{N}{S}$ will become

$P$ will be unimodal

1. If $\frac{B}{S}$ is increased relative to $\frac{u}{y}$, $\frac{1 - u}{N}$ and $\frac{N}{S}$, $\frac{u}{y}$ and $a$, $\frac{1 - u}{N}$ and $\frac{N}{S}$ will be unimodal

2. If $\frac{B}{S}$ is increased and $a$ is held fixed, $\frac{1 - u}{N}$ will be unimodal

3. If $u/\frac{B}{S}$ is increased with $N$, $\frac{1 - u}{N}$ and $\frac{N}{S}$ will become

$P$ will be unimodal

4. If both $a$ and $q$ are small, $\frac{B}{S}$ is more likely to have two modes; is close to 0 for one and

$P$ will be unimodal.

Cases B(4) and C(1) can be understood with similar arguments. In the unimodal cases under $P$, if $\frac{B}{S}$ and $\frac{u}{y}$ are small and $\frac{1 - u}{N}$ and $\frac{N}{S}$ increase, the prior means and variances of $\frac{B}{S}$ and $\frac{u}{y}$ become small and conflict with the information in $\frac{S}{2}B$. But if $\frac{B}{S}$ and $\frac{u}{y}$ are fixed and $\frac{1 - u}{N}$ and $\frac{N}{S}$ increase, the bimodality occurs in cases A(1), B(4), and C(1) because the data conflict with the prior.
with every thing else held fixed, the mode near $\phi = 0.9977$ becomes vanishingly small, although we have not been able to prove this.

The foregoing describes tendencies. Section 4 gives a bimodal example with $\alpha = \nu = 1$; although $\alpha$ and $\nu$ are small, bimodality occurs because $\alpha$ is finite enough. In addition, although the foregoing describes tendencies, Section 4 gives a bimodal example with $\alpha = \nu = 1$, whereas in the same

$$\text{likelihood, i.e., the log posterior as a function of } \phi \text{ with } \alpha, \nu, \text{ and } \Theta \text{ set to values that make the posterior } \phi \text{ given in Theorem 2.1. The two modes are nearly the same}

and with every thing else held fixed, the mode near $\phi = 0.9977$ becomes vanishingly small, although we have not been able to prove this.

The foregoing describes tendencies. Section 4 gives a bimodal example with $\alpha = \nu = 1$; although $\alpha$ and $\nu$ are small, bimodality occurs because $\alpha$ is finite enough. In addition, although the foregoing describes tendencies, Section 4 gives a bimodal example with $\alpha = \nu = 1$, whereas in the same

$$\text{likelihood, i.e., the log posterior as a function of } \phi \text{ with } \alpha, \nu, \text{ and } \Theta \text{ set to values that make the posterior } \phi \text{ given in Theorem 2.1. The two modes are nearly the same}

and with every thing else held fixed, the mode near $\phi = 0.9977$ becomes vanishingly small, although we have not been able to prove this.

The foregoing describes tendencies. Section 4 gives a bimodal example with $\alpha = \nu = 1$; although $\alpha$ and $\nu$ are small, bimodality occurs because $\alpha$ is finite enough. In addition, although the foregoing describes tendencies, Section 4 gives a bimodal example with $\alpha = \nu = 1$, whereas in the same

$$\text{likelihood, i.e., the log posterior as a function of } \phi \text{ with } \alpha, \nu, \text{ and } \Theta \text{ set to values that make the posterior } \phi \text{ given in Theorem 2.1. The two modes are nearly the same}

and with every thing else held fixed, the mode near $\phi = 0.9977$ becomes vanishingly small, although we have not been able to prove this.

The foregoing describes tendencies. Section 4 gives a bimodal example with $\alpha = \nu = 1$; although $\alpha$ and $\nu$ are small, bimodality occurs because $\alpha$ is finite enough. In addition, although the foregoing describes tendencies, Section 4 gives a bimodal example with $\alpha = \nu = 1$, whereas in the same

$$\text{likelihood, i.e., the log posterior as a function of } \phi \text{ with } \alpha, \nu, \text{ and } \Theta \text{ set to values that make the posterior } \phi \text{ given in Theorem 2.1. The two modes are nearly the same}

and with every thing else held fixed, the mode near $\phi = 0.9977$ becomes vanishingly small, although we have not been able to prove this.
Figure 1: The log profile posterior, plus an unimportant constant, as a function of $\phi$, for $a = 0.002$.

Table 1: Values of $\phi$, $\beta$, and $g(\phi)$ for different values of $b$.

Table 2: Values of $\phi$, $\beta$, and $g(\phi)$ for different values of $\phi$. 

$q$
Figure 2: Panels (a), (b), (c): Log profile posterior (plus an unimportant constant) of $\phi$ for various $\lambda$ (Panel (a)), various $b$ (Panel (b)), and two values of $a$ (Panel (c)). Panel (d) is the log profile marginal posterior for $\xi$, for two values of $a$. In panels (c) and (d), the scale beside the left vertical axis is for $a = 0.8$ and the scale beside the right vertical axis is for $a = 1.85$. 
For \( q = 0.05, 0.2, 0.3, 0.4 \), with higher \( q \) corresponding to lower \( q \) on the right-hand side, the joint posterior is bimodal (Table 1, first row). The priors contain strong information that \( a \) and \( b \) are small (Table 1, second row). As \( q \) increases, \( E(\frac{1}{a}) = 0.1, \) an in the first row of Table 1, the joint posterior is bimodal. As \( q \) increases, \( SD(\frac{1}{a}) = 0.08 \), \( E(\frac{1}{b}) = 0 \), \( SD(\frac{1}{b}) = 0.03 \), and \( S^2_B \) suggests that at least one of them is quite positive. Figure 3(a) shows how the modes change with \( \lambda \). The remaining mode has the larger \( a \) and smaller \( b \) predicted by both the prior and the data, while allowing larger values, and the modes move left until the right one disappears. Figure 2 shows analogous effects from increasing \( b \). When \( a = 1, \) \( \lambda = 8, \) the modes \((\frac{1}{b})^0 \) and \((\frac{1}{a})^0 \) move right until the right one disappears. The solid lines represent \( q = 1, 2, 3, 4, 5 \), with higher \( q \) corresponding to higher \( q \) on the right-hand side. Table 2 and Figure 2(b) show analogous effects from increasing \( b \). When \( \lambda = 8, \) the modes \((\frac{1}{b})^0 \) and \((\frac{1}{a})^0 \) move right until the right one disappears. The solid lines represent \( q = 1, 2, 3, 4, 5 \), with higher \( q \) corresponding to higher \( q \) on the right-hand side. The dashed lines represent \( b = 0 \).
Figure 3 gives a different view of the effect of changing $\gamma$ or $\beta$ for fixed $N/n$ and $\sigma$, and $\alpha$. In Figure 3, each line shows, for particular $\gamma$ and $\beta$, the boundary in the error mean square and between-group mean square, where error mean square $= S_{W}/N(n-1)$ and between-group mean square $= S_{B}/N-1(\gamma S_{W}/n)$ such that datasets (points) above the line yield bimodal posteriors and datasets below the line yield unimodal posteriors. The solid lines represent different values of $\gamma$, with higher lines corresponding to higher $\gamma$. Similarly, the dashed lines represent different values of $\beta$, with higher lines corresponding to higher $\beta$. For a given choice of $\gamma$, the boundary is close to a straight line. Note also the kink in the lines for small $\gamma$ or $\beta$. Because of nonlinearity, the threshold in the between-group mean square at which bimodality occurs, $\alpha$, is a function of $S_{W}/\sqrt{n}$. The mean squares above the line represent between-group mean squares at which bimodality occurs. As $\gamma$ increases, the bimodal threshold approaches the threshold for small $\gamma$ or $\beta$ where the mean squares above the line represent between-group mean squares at which bimodality occurs. For a large enough error mean square, the threshold is close to a straight line. Note also the kink in the lines for small $\gamma$ or $\beta$. For the joint and marginal posteriors, the joint and marginal posteriors can have different modalities (O'Hagan, 1978).

Finally, the joint and marginal posteriors can have different modalities (O'Hagan, 1978). For $\gamma = 1$, $\beta = 0.1$, and $\alpha = 10$, $\gamma = 1$, $\beta = 1/0$, and $\alpha = 0.1, 1, 2$ and $\gamma = 10.28$, $\beta = 0.18$, $\gamma = 4.0$, and $\alpha = 0.5$, so the joint posterior is bimodal, with modes at 0.0.1 and 0.0.8. However, for the marginal posterior (19), $\gamma = 8.70338$, the joint posterior is bimodal with modes at 0.267 and 0.84.2. Therefore, for the given value of $\gamma$ or $\beta$ and certain values of the between-group mean square ($S_{B}/N$), a small error mean square ($S_{W}/\sqrt{n}$) implies unimodality, a somewhat larger error mean square implies bimodality, while a still larger error mean square implies unimodality. This continuity arises because of nonlinearity in the between-group mean square, with a smaller error mean square implying a somewhat larger error mean square implying bimodality with a still larger error mean square implying unimodality. This continuity also means that a small error mean square implies unimodality, a somewhat larger error mean square implies bimodality, while a still larger error mean square implies unimodality. This continuity also means that a small error mean square implies unimodality, a somewhat larger error mean square implies bimodality, while a still larger error mean square implies unimodality.
A graphical display of bimodality can be extended to show more information about the nature of posterior bimodality. Figure 4 shows four examples of this extended graphical display, considering the nature of bimodal and unimodal posteriors. The graphic introduced in Figure 3 can be extended to show more information about the nature of posterior bimodality.

Figure 4 illustrates how and affect bimodality for the prior . This prior is used in documentation for BUGS ([Spiegelhalter et al., 1995a, 1995b] and WinBUGS) and has gained some currency.

In Figure 4, the upper panels have and the lower panels have . The two upper panels have and, while the two lower panels have .

Comparing the right- and left-hand panels, changing from to depresses slightly the boundary between bimodal and unimodal posteriors; substantially compresses the region where both modes matter; and for a given value of the error mean square, pushes the contours to boundary between bimodal and unimodal posteriors. Comparing the upper and lower panels, changing from to 10 has little effect on the boundary between bimodal and unimodal posteriors; slightly compresses the upper part of the error mean square; pushes the contours to boundary between bimodal and unimodal posteriors; and for a given value of the error mean square, pushes the contours to boundary between bimodal and unimodal posteriors.
Figure 4: The difference in log height of the two modes, as affected by $N$ and $n$, for $\alpha = \lambda = a = b = 0.001$. 
Discussion

Does bimodality occur in real datasets?

From and generally rather lower than the first. It is not clear why this happens.

3.1 Discussion

Because of perhaps ever create two modes from one. Rather, it implies a second mode different

Figure 4 also shows that for the sample sizes and priors at represent, explaining the other mean.

Perhaps the contours to considered lower between-groups mean squares.
The construction of Figures 3, 4, and 5 might suggest that for a given value of $S^2_B$, the posterior changes modality only once as $S^2_B$ increases from zero. However, we have found cases in which this is false. For example, using the $IG(0.0001, 0.0001)$ prior for both variances, and $N = 12$, $V = 10$, $n = 20$, with $S^2_W$ held fixed while $S^2_B$ increases, we might nonetheless conclude that the posterior is unimodal when $S^2_B = 0$, unimodal when $S^2_B = 1$, multimodal when $S^2_B = 9$, unimodal when $S^2_B = 2$, the posterior is unimodal when $S^2_B = 9$, multimodal when $S^2_B = 1$, unimodal when $S^2_B = 0$. This is false. For example, using the $IG(0.0001, 0.0001)$ prior for both variances, we have found cases in which the posterior changes modality only once as $S^2_B$ increases from zero. However, we have found cases in which the posterior changes modality only once as $S^2_B$ increases from zero.

The construction of Figures 3, 4, and 5 might suggest that for a given value of $S^2_B$, the posterior changes modality only once as $S^2_B$ increases from zero. However, we have found cases in which this is false. For example, using the $IG(0.0001, 0.0001)$ prior for both variances, and $N = 12$, $V = 10$, $n = 20$, with $S^2_W$ held fixed while $S^2_B$ increases, we might nonetheless conclude that the posterior is unimodal when $S^2_B = 0$, unimodal when $S^2_B = 1$, multimodal when $S^2_B = 9$, unimodal when $S^2_B = 2$, the posterior is unimodal when $S^2_B = 9$, multimodal when $S^2_B = 1$, unimodal when $S^2_B = 0$. This is false. For example, using the $IG(0.0001, 0.0001)$ prior for both variances, we have found cases in which the posterior changes modality only once as $S^2_B$ increases from zero. However, we have found cases in which the posterior changes modality only once as $S^2_B$ increases from zero.

The construction of Figures 3, 4, and 5 might suggest that for a given value of $S^2_B$, the posterior changes modality only once as $S^2_B$ increases from zero. However, we have found cases in which this is false. For example, using the $IG(0.0001, 0.0001)$ prior for both variances, and $N = 12$, $V = 10$, $n = 20$, with $S^2_W$ held fixed while $S^2_B$ increases, we might nonetheless conclude that the posterior is unimodal when $S^2_B = 0$, unimodal when $S^2_B = 1$, multimodal when $S^2_B = 9$, unimodal when $S^2_B = 2$, the posterior is unimodal when $S^2_B = 9$, multimodal when $S^2_B = 1$, unimodal when $S^2_B = 0$. This is false. For example, using the $IG(0.0001, 0.0001)$ prior for both variances, we have found cases in which the posterior changes modality only once as $S^2_B$ increases from zero. However, we have found cases in which the posterior changes modality only once as $S^2_B$ increases from zero.

The construction of Figures 3, 4, and 5 might suggest that for a given value of $S^2_B$, the posterior changes modality only once as $S^2_B$ increases from zero. However, we have found cases in which this is false. For example, using the $IG(0.0001, 0.0001)$ prior for both variances, and $N = 12$, $V = 10$, $n = 20$, with $S^2_W$ held fixed while $S^2_B$ increases, we might nonetheless conclude that the posterior is unimodal when $S^2_B = 0$, unimodal when $S^2_B = 1$, multimodal when $S^2_B = 9$, unimodal when $S^2_B = 2$, the posterior is unimodal when $S^2_B = 9$, multimodal when $S^2_B = 1$, unimodal when $S^2_B = 0$. This is false. For example, using the $IG(0.0001, 0.0001)$ prior for both variances, we have found cases in which the posterior changes modality only once as $S^2_B$ increases from zero. However, we have found cases in which the posterior changes modality only once as $S^2_B$ increases from zero.

The construction of Figures 3, 4, and 5 might suggest that for a given value of $S^2_B$, the posterior changes modality only once as $S^2_B$ increases from zero. However, we have found cases in which this is false. For example, using the $IG(0.0001, 0.0001)$ prior for both variances, and $N = 12$, $V = 10$, $n = 20$, with $S^2_W$ held fixed while $S^2_B$ increases, we might nonetheless conclude that the posterior is unimodal when $S^2_B = 0$, unimodal when $S^2_B = 1$, multimodal when $S^2_B = 9$, unimodal when $S^2_B = 2$, the posterior is unimodal when $S^2_B = 9$, multimodal when $S^2_B = 1$, unimodal when $S^2_B = 0$. This is false. For example, using the $IG(0.0001, 0.0001)$ prior for both variances, we have found cases in which the posterior changes modality only once as $S^2_B$ increases from zero. However, we have found cases in which the posterior changes modality only once as $S^2_B$ increases from zero.

Harville, D. A. and Zimmerman, A. C. (1969). The posterior distribution of the fixed and


Pirtle, W. A. and Thompson, R. A. (1990). Investigation of bimodality in neighborhoods and poste-

References

University of Iowa assisted the example in Section 4.

in a future paper.


Haviland, D. A. and Zimmerman, A. C. (1969). The posterior distribution of the fixed and

Acknowledgments

Although no single prior for the variance guarantees unimodality, unimodality can be

Although we have not yet characterized posterior modality for more complicated hierarchical

Although we have not yet characterized posterior modality for more complicated hierarchical

Although we have not yet characterized posterior modality for more complicated hierarchical

We proceed either of these examples.

moderate and slight no single prior for the variance guarantees unimodality. Although we have not

In particular, it appears that posterior for these models have no more than two

POSTWRM. In particular, it appears that posterior for these models have no more than two

POSTWRM, hierarchical models with two unknown variances appear to behave similarly to the

preorder for a more complex model, there is single result with an informative prior.

The POSTWRM is complex posterior modality arises from subtle ways in which the data and

Conclusion

5.3
\[ \begin{align*}
\frac{e^{\theta + \gamma t/2}}{\sqrt{2\pi t}} + \frac{e^{\theta + \gamma t/2}}{\sqrt{2\pi t}} = \theta \\
\frac{\gamma}{t} = \gamma
\end{align*} \]

\[
I_{U}(\Theta, \gamma) = \log F(\Theta) = \left( \frac{e^{\theta + \gamma t/2}}{\sqrt{2\pi t}} \right)
\]

### Appendix: Proof of the main result

Let

\[
\frac{\partial}{\partial \theta} \left( \frac{e^{\theta + \gamma t/2}}{\sqrt{2\pi t}} + \frac{e^{\theta + \gamma t/2}}{\sqrt{2\pi t}} \right) = \theta
\]

and solve \( \frac{\partial}{\partial \theta} = 0 \) to obtain

\[
\theta = \frac{\gamma}{t}, \quad \gamma = \frac{\gamma}{t}
\]
solutions, but at most two modes. Therefore, and \( g \) have at most two real

with restriction \( g \) is continuous in \( \phi \) and at most three real

\( \bar{g} \) have the same critical points, and \( \bar{g} \) maximizes \( (\phi) \bar{g} \) with restriction \( \bar{g} \) with 

Then consider the profile log posterior \( \bar{g} \), where \( \bar{g} \) = \( \bar{g} \phi + M_{\bar{g}} + \gamma_{\bar{g}} \)

where \( \bar{g} \) = \( \bar{g} \phi \) and \( (1 + e^{\phi} \bar{g}) \bar{g} - (1 + e^{\phi} \bar{g}) \bar{g} - \sigma = (\phi) \bar{g} \)

and \( (\phi) \bar{g} \) for \( 0 < (\phi) \bar{g} < (\phi) \bar{g} \) and satisfies \( \sigma \) and \( \sigma \) to respect to \( \bar{g} \) with \( \bar{g} \). (12)

\[
\begin{align*}
\frac{u/\bar{g} \phi (\phi - 1) + q\bar{g}}{1} \bar{g} + \sigma = (\phi) \bar{g} = \bar{g} \\
\frac{\bar{g} \phi + M_{\bar{g}} + \gamma_{\bar{g}}}{1} \bar{g} + \sigma + uN = (\phi) \bar{g} = \bar{g}
\end{align*}
\]

extreme points \( (\phi) \bar{g} \) where

To maximize under the restriction \( \bar{g} \) use Lagrange's method to obtain the possible

\[
\begin{align*}
\frac{u/\bar{g} \phi (\phi - 1) + q\bar{g}}{1} \bar{g} + \sigma = (\phi) \bar{g} = \bar{g} \\
\frac{\bar{g} \phi + M_{\bar{g}} + \gamma_{\bar{g}}}{1} \bar{g} + \sigma + uN - \sigma = (\phi) \bar{g} = \bar{g}
\end{align*}
\]

Luttrell can be written as

\[
\begin{align*}
\phi_{\text{max}} \text{H}_{\text{2}}(\text{O}, 1928) \frac{u}{\bar{g}} \frac{\bar{g} + \sigma}{1} = \phi \text{ Luttrell}
\end{align*}
\]

Then consider the profile log posterior

\[
\begin{align*}
\phi_{\text{max}} \text{H}_{\text{2}}(\text{O}, 1928) \frac{u}{\bar{g}} \frac{\bar{g} + \sigma}{1} = \phi \text{ Luttrell}
\end{align*}
\]
Thompson used a similar polynomial to produce examples of bimodal posteriors.

Having established this, we need the real solutions of $(\phi)\bar{b}$ for all $\phi$.

Lemma 6.1

For the cubic polynomial $g$: $(\phi)\bar{b}$ and $\nabla \frac{\partial}{\partial \phi} = \nabla \frac{\partial}{\partial \phi}$. (6)

Recall the definition of $\phi$, $\bar{b}$, and $g$.

\[
\begin{align*}
\text{where} \quad 0 = (\phi)\bar{b} \quad \text{and} \quad 0 > (\phi)\bar{b} \quad \text{if} \quad 0 = (\phi)\bar{b} \quad \text{and} \quad 0 > (\phi)\bar{b}.
\end{align*}
\]

For the real solutions of $g$, we have two different real solutions:

- $0 = (\phi)\bar{b}$ then $0 = (\phi)\bar{b}$ and $0 > (\phi)\bar{b}$ because $0 > (\phi)\bar{b} > 0$. Hence, only one real solution.
- $0 > (\phi)\bar{b}$ then $0 = (\phi)\bar{b} \bar{b}$ and $0 < (\phi)\bar{b}$.

Thus:

- $0 < \phi g \left( \frac{\partial}{\partial \phi} + \phi \bar{b} \bar{b} \right) \bar{b} \bar{b} + f(\phi)\bar{b} \bar{b} = (1)\bar{b}$ and $0 < \phi g \left( \frac{\partial}{\partial \phi} + \phi \bar{b} \bar{b} \right) \bar{b} \bar{b} + f(\phi)\bar{b} \bar{b}$

To prove the Lemma: note that $0 = (\phi)\bar{b}$ has at least one real solution in $(\phi, 1)$.

If $0 < \phi g \left( \frac{\partial}{\partial \phi} + \phi \bar{b} \bar{b} \right) \bar{b} \bar{b} + f(\phi)\bar{b} \bar{b}$ then $0 = (\phi)\bar{b}$ has only one real solution and

- $0 < \phi g \left( \frac{\partial}{\partial \phi} + \phi \bar{b} \bar{b} \right) \bar{b} \bar{b} + f(\phi)\bar{b} \bar{b}$.

All the real solutions are in $(\phi, 1)$.

\[
\begin{align*}
\text{Otherwise,} \quad 0 = (\phi)\bar{b} \quad \text{has only one real solution:} \quad 0 = (\phi)\bar{b} \quad \text{has two real solutions:} \quad 0 = (\phi)\bar{b} \quad \text{has three real solutions:} \quad 0 = (\phi)\bar{b}.
\end{align*}
\]
1. When \( 0 < b \neq 0 \), 32) has three different real solutions.

From points 2. and 3. all the real solutions are in (0, 0).

From the theory of cubic polynomials

\[
0 = b + c + \varepsilon f
\]

To prove Theorem 2.1, we only need to find the maximising \( \phi \) as

\[
(\varepsilon f(1) > 0)
\]

Trove Theorem 2.1, we only need to find the maximising \( \phi \) as

\[
(\varepsilon f(1) > 0)
\]
Thus, we only need to prove that the solutions here can be written as in (10) and (11) for

2.1. Thus, we only need to prove that the solutions here can be written as in (10) and (11) and respectively in Theorem 2.1, and items 3 and 4 above correspond to case 3 in Theorem 2. Since (22) is just a transformation of (22), (22) and (22) are the same.

\[
\frac{\varepsilon}{x} \Delta + \left( \frac{\varepsilon}{b} \right) \Delta - \frac{\varepsilon}{b} - \left( \frac{\varepsilon}{b} \right) \Delta + \frac{\varepsilon}{b} = u_i
\]

1. When \(0 < \varepsilon\), has only one real solution:

\[\varepsilon = \varepsilon_i = \varepsilon_i = u_i\]

2. When \(0 = b = d\) and \(0 = 0\), has three real solutions:

\[0 = \varepsilon \left( \frac{\varepsilon}{b} \right) \Delta - \left( \frac{\varepsilon}{b} \right) \Delta = \varepsilon_i\]

3. When \(0 \neq \varepsilon \left( \frac{\varepsilon}{b} \right)\), has three real solutions:

\[0 \neq \varepsilon \left( \frac{\varepsilon}{b} \right)\]

\[\frac{\varepsilon}{x} \Delta - \frac{\varepsilon}{b} - \left( \frac{\varepsilon}{b} \right) \Delta + \frac{\varepsilon}{b} = u_i\]

where

\[
\frac{\varepsilon}{x} \Delta + \left( \frac{\varepsilon}{b} \right) \Delta - \frac{\varepsilon}{b} - \left( \frac{\varepsilon}{b} \right) \Delta + \frac{\varepsilon}{b} = u_i
\]
corresponds to the only mode while
\[
\frac{f + \phi}{f - \phi} + \frac{2}{f - \phi} = \frac{f + \phi}{f - \phi} + \frac{2}{f - \phi}
\]

**Case 2.** When \( \nu \) is positive, we find that the form in (10)

\[
(\xi, \phi) > (\xi) > (1)^{\phi} \nu > (\phi) \nu \in \mathbb{R}.
\]

Since \( \phi > 0 \), consider \( \nu > 0 \). The complete solution is positive, and

\[
\frac{e^{\nu}}{\nu} \quad \text{is also positive, and}
\]

\[
\left[ \phi \sin \nu - \phi \cos \nu \right] \frac{e^{\nu}}{\nu} = \frac{e^{\nu}}{\nu} \quad \text{and}
\]

\[
\left[ \phi \sin \nu + \phi \cos \nu \right] \frac{e^{\nu}}{\nu} = \frac{e^{\nu}}{\nu}
\]

Similarly,

\[
\phi \cos \nu \frac{e^{\nu}}{\nu} = \left[ \phi \sin \nu - \phi \cos \nu \right] \frac{e^{\nu}}{\nu} + \left[ \phi \sin \nu + \phi \cos \nu \right] \frac{e^{\nu}}{\nu}
\]

\[
\phi \cos \nu \frac{e^{\nu}}{\nu} + \phi \cos \nu \frac{e^{\nu}}{\nu} = \nu
\]

Thus, \( \nu \geq \frac{\xi + 1}{\phi} \) for all \( \phi > 0 \) where

\[
\frac{e^{\nu}}{\nu} - \frac{\xi}{\phi} = \left( \frac{\xi}{\phi} \right) \frac{e^{\nu}}{\nu} - \frac{\xi}{\phi}
\]
<table>
<thead>
<tr>
<th>B/951/0</th>
<th>B/954/3</th>
<th>B/952/7</th>
<th>B/953/1</th>
</tr>
</thead>
<tbody>
<tr>
<td>J/868/1</td>
<td>J/865/2</td>
<td>J/863/5</td>
<td>J/861/1</td>
</tr>
<tr>
<td>B/959/1</td>
<td>B/956/2</td>
<td>B/953/1</td>
<td>B/950/1</td>
</tr>
<tr>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
</tr>
</tbody>
</table>

**Table 3:** Significant features of the datasets. "DHP" is the log height of the least mode in minutes.