Multiple local maxima in the RL and posterior

The literature here is very thin, though $\exists$ (Henn & Hodges 2014 *ISR*).

Clearly, multiple maxima occur more readily

- in the likelihood than in the RL. (Won't discuss further)
- in the marginal posterior than in the RL.

The rest of this section will

- briefly survey the little that's known about the RL, and
- mostly focus on the posterior (which is weird enough).

Multiple local maxima in the RL

Henn & Hodges 2014 *ISR* found 1 (one) report of an RL with multiple local maxima (Welham & Thompson 2009 *CSDA*).

H & H 2014 *ISR* manufactured another example, as I'll describe below.

Recently, Philip Reiss found another one (as yet unpublished).

I'll use the re-expressed RL to show how to manufacture examples.

For two-variance models, we've written the RL as

$$\log RL(\sigma_s^2, \sigma_e^2|\mathbf{y}) = B - \frac{n - s_X - s_Z}{2}\log(\sigma_e^2) - \frac{1}{2\sigma_e^2}\mathbf{y}'\mathbf{\Gamma}_c\mathbf{\Gamma}_c'\mathbf{y} \quad (1)$$

$$-\frac{1}{2}\sum_{j=1}^{s_Z}\left[\log(\sigma_s^2 a_j + \sigma_e^2) + \frac{\hat{v}_j^2}{\sigma_s^2 a_j + \sigma_e^2}\right] \quad (2)$$

for $s_X = \text{rank}(\mathbf{X}), s_Z = \text{rank}(\mathbf{X}|\mathbf{Z}) - \text{rank}(\mathbf{Z})$.

The RL has the form of a likelihood from a gamma-errors GLM with identity link (2) times a prior for σ_e^2 (1).

$\Rightarrow$ the RL can have a second "mode" if this "prior" and "likelihood" provide very different information about σ_e^2.

H & H 2014 *ISR* manufactured an example this way.

$\Rightarrow$ (2) can have two "mode"s $\Rightarrow$ *perhaps* $\exists$ RLs with 3 "modes".

Multiple local maxima in the posterior

This section tries to frighten you with strange but true results about bimodality in posterior distributions for two simple problems.

Point: Multimodal posteriors probably happen far more often than is generally known or acknowledged, so it would be helpful to develop a collection of examples as a basis for further research.

Balanced one-way random effect model

Model: $y_{ij} = \mu + u_i + \epsilon_{ij}$, $i = 1, \ldots, N$ groups, $j = 1, \ldots, m$ obs/group
$n = Nm$, u_i iid $N(0, \sigma_s^2)$, ϵ_{ij} iid $N(0, \sigma_e^2)$.

The RL has a single maximum $\Rightarrow$ multimodal posteriors arise from conflict between the RL and prior.

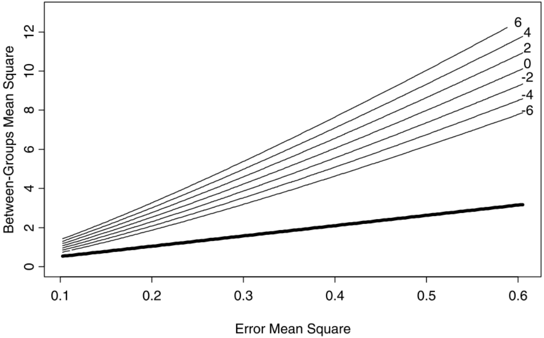
Liu & Hodges (JRSSB 2003) gave conditions under which these posteriors are unimodal or bimodal:

- joint posterior of $(\mu, u_1, \ldots, u_N, \sigma_s^2, \sigma_e^2)$
- marginal posterior of $(\mu, \sigma_s^2, \sigma_e^2)$
- marginal posterior of (σ_s^2, σ_e^2).

$\exists \leq 2$ modes; these 3 posteriors need not have the same modality.

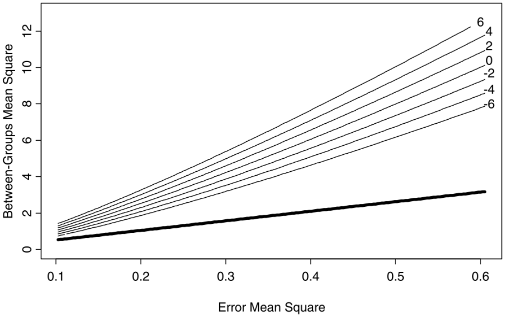
For each posterior, the stationary points are solutions to a cubic equation and utterly without intuition.

Typical results: joint posterior, $N = 10$ groups, $m = 10$ obs/group, $\sigma_s^2, \sigma_e^2 \sim$ IG(0.001,0.001).



A dataset is a point: error MS(horiz) and between-groups MS (vert)

Datasets below and above the thick line give unimodal and bimodal posteriors respectively.



Small between-groups MS $\Rightarrow$ one mode (“much shrinkage”).

When the between-groups MS reaches the thick line, a distant second mode arises (“little shrinkage”), and mass moves to it.

As the between-groups MS increases, mass shifts to “little shrinkage”.

Contour lines: height “little shrinkage” mode minus height “much shrinkage” mode (log units).

How often does this happen in real datasets?

Considering 67 balanced one-way datasets from JH's statistical practice, with inverse gamma (0.001,0.001) priors for the variances:

- 47 datasets (70%) gave bimodal posteriors for $(\mu, \sigma_s^2, \sigma_e^2)$;
- 4 datasets (6%) had modes differing in height by $\leq$ 4 natural-logs.

Answer: It happens all the time BUT the second mode is usually — though not always — so small you'd never notice it.

Facts, from inconvenient to truly strange

No IG prior on the variances guarantees a unimodal posterior.

Increasing N (fixing other things) does not necessarily fix bimodality.

The region of datasets giving 2 modes isn't necessarily convex.
$N = 15$, $m = 20$, within-group SS fixed at 5.2, $\sigma_s^2, \sigma_e^2 \sim$ IG(0.001,0.001):

- unimodal when the between-groups sum of squares is 0.9,
- bimodal when it's 1.1,
- unimodal again when it's 1.5,
- bimodal again when it reaches 300.

This oddity appears to be rare.

What should we to think about all this?

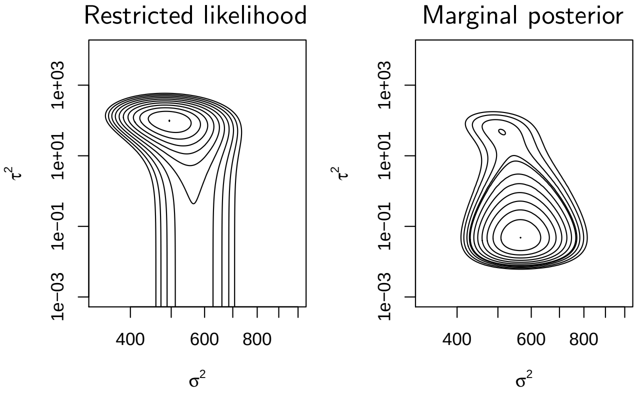
de Finetti: If this model and prior capture Your beliefs, then a posterior with a mixture of extreme beliefs is the coherent (i.e., correct) result.

de Finetti: This mixture of extremes merely shows hidden strength in a specification made almost automatic by habit.

The message, perhaps, is that we understand this specification less well than we would like to think — and it's the simplest MLM.

A two-level model: The HMO data revisited

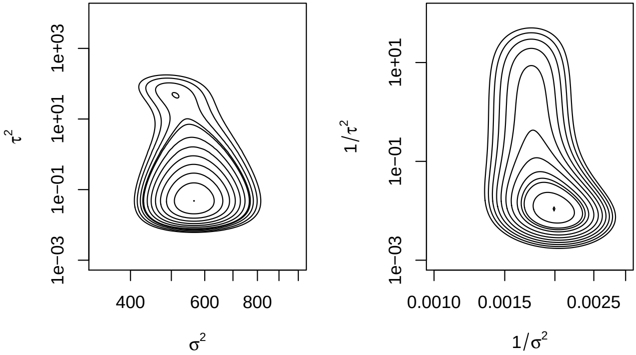
Wakefield: Bimodal posterior for between-state variance (τ^2) and within-state variance (σ^2) $\Rightarrow$ "there are two competing explanations for the observed variability in the data". Well, maybe not ...



I'd say: The RL has fairly weak info about τ^2 and the prior punctures it.

Re-parameterize to precisions and the modality changes!

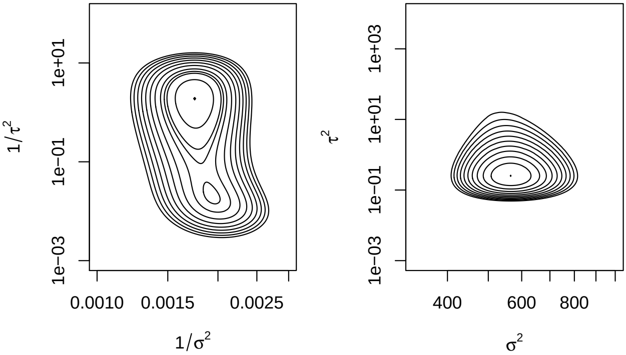
Simply changing variables changes a bimodal to a unimodal posterior.



This happens because the mode for τ^2 created by the prior is spread out over a wide range of $1/\tau^2$ and becomes a "shoulder".

Here's an example without a shoulder

Just change the prior on the error precision $1/\sigma^2$ to a gamma with parameters $\alpha = 3$ and $\beta = 1$



Lisa Henn stumbled on the first example but produced the second easily

$\Rightarrow$ either she was really lucky or it's easy to manufacture examples.