Zero variance estimates

Hardly anything is known about estimates on the boundary of the parameter space.

When a zero variance estimate maximizes the RL, I want to know:
▶ are the data consistent with “large” values of that variance,
▶ i.e., does the RL have a flat left tail in that variance?

This section
▶ examines the BOWREM in some detail;
▶ gives short comments about a nested ANOVA model; and
▶ finishes with some thoughts about tools.
Balanced one-way RE model (BOWREM)

\[ y_{ij} = \beta_0 + u_i + \epsilon_{ij}, \]  
\[ N \text{ groups, } m \text{ per group, } u_i \sim N(0, \sigma_s^2), \epsilon_{ij} \sim N(0, \sigma_e^2). \]

Define \( n = Nm, S_E = \sum_{ij} (y_{ij} - \bar{y}_i)^2, S_M = \sum_i (\bar{y}_i - \bar{y}.)^2 \)

The log RL is:

\[
-\frac{n - N}{2} \log(\sigma_e^2) - \frac{1}{2} \frac{S_E}{\sigma_e^2} - \frac{N - 1}{2} \log(\sigma_s^2 m + \sigma_e^2) - \frac{m}{2} \frac{S_M}{(\sigma_s^2 m + \sigma_e^2)},
\]

It is maximized by:

if \( \frac{S_M}{N-1} \geq \frac{S_E}{m(n-N)} \)

\[ \hat{\sigma}_e^2 = \frac{S_E}{n-N}, \quad \hat{\sigma}_s^2 = \frac{S_M}{N-1} - \frac{\hat{\sigma}_e^2}{m} \]

if \( \frac{S_M}{N-1} < \frac{S_E}{m(n-N)} \)

\[ \hat{\sigma}_e^2 = \frac{(S_E + mS_M)}{(n-1)}, \quad \hat{\sigma}_s^2 = 0. \]
When is this RL flat near $\hat{\sigma}^2_s = 0$?

Consider the log RL’s derivative wrt $\sigma^2_s$, evaluated at $\sigma^2_s = 0$ and $\hat{\sigma}^2_e$:

$$\frac{\partial \log \text{ RL}}{\partial \sigma^2_s} = \left( \frac{\hat{\sigma}^2_e}{m} \right)^{-1} \frac{N - 1}{2} \left( \frac{S_M}{N - 1} \left( \frac{\hat{\sigma}^2_e}{m} \right)^{-1} - 1 \right).$$

If $S_M/(N - 1) < \hat{\sigma}^2_e/m$, this derivative < 0 and $\hat{\sigma}^2_s = 0$.

The derivative is small in absolute value if either

- $\hat{\sigma}^2_e/m$ is large or
- $\frac{S_M}{N - 1} \left( \frac{\hat{\sigma}^2_e}{m} \right)^{-1}$ is close to 1.

These two conditions have different implications.
\[
\frac{\partial \log \text{ RL}}{\partial \sigma_s^2} = \left( \frac{\hat{\sigma}_e^2}{m} \right)^{-1} \frac{N - 1}{2} \left( \frac{S_M}{N - 1} \left( \frac{\hat{\sigma}_e^2}{m} \right)^{-1} - 1 \right).
\]

If \(\hat{\sigma}_e^2 / m\) is large, the design and data provide low resolution for \(\sigma_s^2\).

- A wide interval of positive \(\sigma_s^2\) have RL near the max value.
- The RL provides this information; it is routinely ignored.

To increase the design’s resolution and get \(\hat{\sigma}_s^2 > 0\):
- Increase \(m\), holding constant \(N\), \(S_M\), and \(\hat{\sigma}_e^2\).
- Simply increasing \(N\) doesn’t work:
  - Holding constant \(S_M / (N - 1)\) and \(\hat{\sigma}_e^2 / m\), this leaves the key condition \(S_M / (N - 1) < \hat{\sigma}_e^2 / m\) unchanged.
  - The derivative does become larger in magnitude (steeper dropoff).
\[
\frac{\partial \log \text{RL}}{\partial \sigma_s^2} = \left( \frac{\hat{\sigma}_e^2}{m} \right)^{-1} \frac{N - 1}{2} \left( \frac{S_M}{N - 1} \left( \frac{\hat{\sigma}_e^2}{m} \right)^{-1} - 1 \right).
\]

If \(\hat{\sigma}_e^2/m - S_M/(N - 1) < 0\) but close to 0, \(\partial \log \text{RL}/\partial \sigma_s^2 < 0\) and small because the peak is close to \(\sigma_s^2 = 0\).

The RL may but does not necessarily decline slowly from \(\sigma_s^2 = 0\)

- When \(S_M/(N - 1) > 0.5\hat{\sigma}_e^2/m\), \(\partial^2 \log \text{RL}/\partial (\sigma_s^2)^2 < 0\)
  - and the restricted likelihood can drop off quickly.

Otherwise, it drops off slowly.
A more complicated ANOVA (Epidermal nerve density)

The real example:

The investigators want to compare biopsy (old) vs. blister (new).

19 subjects; at calf and foot, 2 blisters.

They’re interested in the between-blister variation.

Here are the max-RL estimates of variance components (CIs from SAS)

<table>
<thead>
<tr>
<th>Variance component</th>
<th>Variance Estimate</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>18,031</td>
<td>8,473 61,169</td>
</tr>
<tr>
<td>Subject-by-Location</td>
<td>9,561</td>
<td>4,684 29,197</td>
</tr>
<tr>
<td>Blister within Subject/Location</td>
<td>0</td>
<td>0 0</td>
</tr>
<tr>
<td>Residual</td>
<td>6,696</td>
<td>5,181 8,992</td>
</tr>
</tbody>
</table>
Consider a simplified version with 20 subjects and balance

Location is a fixed effect

Four variance components:

- $\sigma^2_{s1}$ variation between subjects
- $\sigma^2_{s2}$ variation between subjects in the difference between location
- $\sigma^2_{s3}$ variation between blisters within subject and location
- $\sigma^2_e (j = 4)$ error (variation between images within a blister)

$$
\log RL(\sigma^2_{s1}, \sigma^2_{s2}, \sigma^2_{s3}, \sigma^2_e | y) = -0.5 \sum_{j=1}^{4} DF_j \left[ \log \theta_j + \hat{\theta}_j^u / \theta_j \right],
$$

$DF_j$ are the usual ANOVA DF; $\hat{\theta}_j^u$ are the usual mean squares.

Degrad the design’s resolution by increasing the error mean square, $\hat{\theta}_4^u$. What happens?
Mean squares are 15, 7, and 3 for sub, sub×loc, blister(sub×loc)

<table>
<thead>
<tr>
<th>$\hat{\theta}_4^u$</th>
<th>$\hat{\sigma}_{s1}^2$</th>
<th>$\hat{\sigma}_{s2}^2$</th>
<th>$\hat{\sigma}_{s3}^2$</th>
<th>$\hat{\sigma}_e^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>1.00</td>
<td>0.50</td>
<td>2.00</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td>3.00</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>0.83</td>
<td>0.00</td>
<td>3.67</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>8</td>
<td>1.00</td>
<td>0.17</td>
<td>0.00</td>
<td>6.33</td>
</tr>
<tr>
<td>9</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>7.00</td>
</tr>
<tr>
<td>10</td>
<td>0.93</td>
<td>0.00</td>
<td>0.00</td>
<td>7.58</td>
</tr>
<tr>
<td>11</td>
<td>0.86</td>
<td>0.00</td>
<td>0.00</td>
<td>8.15</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>21</td>
<td>0.14</td>
<td>0.00</td>
<td>0.00</td>
<td>13.91</td>
</tr>
<tr>
<td>22</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>14.48</td>
</tr>
<tr>
<td>23</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>15.05</td>
</tr>
</tbody>
</table>

The analogous thing happens if we fix error MS and increase $\hat{\theta}_3^u$. This is unexplained (as far as I know).
Some thoughts about tools

My question: When a zero variance estimate maximizes the RL, are the data consistent with “large” values of that variance?

The obvious form for this information is a simple one-sided CI.

- We might use derivatives; but how to calibrate “small”?
- A simple CI, if one exists, avoids this calibration problem.
- This would be useful to Bayesians because it’s simple and fast.
- And it’s OK if the CI’s coverage tends to be low.

The obvious candidate would use the profile log RL.

- Upper end: \( \sigma_s^2 \) profile log RL is reduced by \( c \) from its max.
- Some software already computes the profile RL.
- Problem: Which \( c \) to use? Solution: Big simulation experiment.