Example: modeling reaction times of schizophrenics and nonschizophrenics

Bayesian Data Analysis (page 468)\(^1\)

Queensland Center for Schizophrenia Research collection.

\(^1\)This example first appeared in the paper *The analysis of repeated-measures data on schizophrenic reaction times using mixture models*, Berlin, T. and Rubin, D.; Statistics in Medicine, Vol. 14, 747-768 (1995)
Schizophrenia

- Schizophrenia is most likely a group of illnesses affecting the most characteristically human of abilities such as language, planning, emotion and perceptions. It is relatively common, often becomes apparent in early adulthood, and frequently results in prolonged disability.

- Schizophrenia is traditionally defined by symptoms and signs according to one of several sets of diagnostic criteria. Both major classification systems are: The International Classification of Diseases produced by the World Health Organization, and the American Psychiatric Association Fourth Edition of the Diagnostic and Statistical Manual (DSM-IV). Both sets of criteria are very similar and emphasize traditional symptoms such as hallucinations of voices commenting on the person’s actions, delusions, experiences of various forms of interference with the person’s thoughts, incoherent or irrelevant speech, blunting of affect and decline in general level of functioning.

- Although schizophrenia appears to affect males and females in equal numbers, males tend to develop schizophrenia at a younger age than females (modal age of onset for males:females = 20-24:25-29). In general the course of the disorder is worse for males than females. Men have more and longer hospital stays, higher suicide rates, greater substance abuse, poorer living conditions and a worse psychosocial outcome than women.

---

2Extracted from Queensland Center for Schizophrenia Research (http://www.qcsr.uq.edu.au)
Dataset

- Psychologists at Harvard University performed an experiment measuring thirty reaction times for each of seventeen male subjects: eleven non-schizophrenics and six schizophrenics.

- Manual reaction times to visual cues, where subjects watch a screen and move their fingers from one button to another when a signal appears, were measured.

- There are 30 measurements per individual.
Histogram of $Y = \log(\text{response time in milliseconds})$
Non-schizophrenic individuals
Histogram of $Y = \log(\text{response time in milliseconds})$

Schizophrenic individuals
• Response times appear longer for schizophrenics than for non-schizophrenics.

• Response times for most of the schizophrenics appear more variable than those for non-schizophrenics.

• Because of differences in both mean and variance, a simple linear model analysis seems inappropriate. The lack of fit of a standard model can be assessed by considering
  
  – Replicate data consistent with the fitted ANOVA model can be generated by first drawing a value of the common variance from its scaled Inv-$\chi^2$ posterior distributions, then drawing values of the person means from their normal posterior distribution given the drawn value of the common variance. Finally draw 30 observations for each individual from a normal with the appropriate person mean and the common variance.
  
  – Replicate the previous step, say, 1000 times, and each time calculate the largest within-person variance, smallest within-person variance, and average within-person variance among the six schizophrenic individuals.
  
  – Compare those values with the observed data.
• Current psychological theory suggests:

(a) Schizophrenics have an attentional deficiency, implying that on some trials they have difficulty attending to the experimental task, with a corresponding delay in initiating a reaction.

(b) Schizophrenia might be associated with motor reflex retardation implying slower reflexes once a reaction has commenced.

Thus, at least some of the extra-variability arises from an attentional lapse that delays some, but not all, of each schizophrenic’s reaction times.

• Although there are systematic relationships between physical measurements (e.g. reaction times and perceptual responses) and symptom variables (e.g. length of institutionalization and severity of disease), the dataset is not detailed enough or large enough to address such factors in interpreting schizophrenic reaction times.
Proposed models

• Measurements from schizophrenics arise from a mixture of

(a) a component analogous to the distribution of response time for non-schizophrenics.

(b) a mean-shifted component.
Model 1

(a) Response times for non-schizophrenic individuals are treated as arising independently from a normal distribution with a distinct mean $\alpha_j$ for individual $j$ and common variance $\sigma^2_y$.

(b) Each schizophrenic also has a distinct mean $\alpha_j$ for his response times, but more structure is needed to reflect the underlying psychological theory. Attentional deficiency is represented by modeling the distribution of response times for each schizophrenic individual as a two-component mixture: a proportion $(1 - \lambda)$ of the responses from schizophrenic $i$ arise from a component analogous to the distribution of non-schizophrenic response times, with mean $\alpha_i$ and variance $\sigma^2_y$, and a proportion $\lambda$ of the responses from schizophrenic $j$ arise from a shifted component centered at $\alpha_j + \tau$ also with variance $\sigma^2_y$.

(c) The comparison of the typical components of $\alpha = (\alpha_1, \ldots, \alpha_{17})$ for schizophrenics versus non-schizophrenics addresses the magnitude of schizophrenics’ motor reflex retardation. A hierarchical parameter $\beta$ measuring this motor retardation was included.
Model 1 (cont’d)

Define,

- $S_j = I(\text{individual } j \text{ is schizophrenic})$, where $I(\cdot)$ is the indicator function.
- $\lambda$, the probability that an observation will be delayed for a schizophrenic subject to attentional delays;
- $\tau$, the size of the delay when an attentional lapse occurs (on the log scale);
- $\beta$, the average log response time for the non-delayed observations of schizophrenics minus the average log response time for nonschizophrenics;
- $\zeta_{ij} = I(\text{response } i \text{ for schizophrenic } j \text{ is from the shifted component}).$

Further define,

- $\phi = (\lambda, \tau, \mu, \sigma^2_\alpha, \sigma^2_y)$
- $y_{ij}$ the $i$th response of individual $j$.
Model 1 (cont’d)

• The data likelihood is

  – For non-schizophrenic individuals:

    \[ y_{ij}|\alpha_j, \phi \sim N(\alpha_j, \sigma^2_y) \]

  – Among schizophrenic individuals:

    (a) for “non-shifted” reaction times

    \[ y_{ij}|\alpha_j, \phi \sim N(\alpha_j, \sigma^2_y) \]

    (b) for “shifted” reaction times

    \[ y_{ij}|\alpha_j, \phi \sim N(\alpha_j + \tau, \sigma^2_y) \]

• The model can be expressed in terms of the indicators variables

  \[ y_{ij}|\alpha_j, \zeta_{ij}, \phi \sim N(\alpha_j + \tau \zeta_{ij}, \sigma^2_y) \]

  \[ \alpha_j|\zeta, \phi \sim N(\mu + \beta S_j, \sigma^2_{\alpha}) \]

  \[ \zeta_{ij}|\phi \sim \text{Bernoulli}(\lambda S_j) \]
• Hyperprior distribution
  
  – Non-informative uniform joint prior density on $\phi$.

• Some additional considerations
  
  – With the given hyperprior the model is not identified, because the trials unaffected by a positive attentional delay could instead be thought of as being affected by a negative attentional delay. $\tau$ was restricted to be positive to identify the model.
  
  – The variance components $\sigma_y^2$ and $\sigma_\alpha^2$ are of course restricted to be positive as well.
  
  – The mixture component $\lambda$ is actually taken to be uniform on $[0.001, 0.999]$ as values of zero or one would not correspond to mixture distributions.
  
  – The assumption of a common variance $\sigma_y^2$ for shifted and unshifted schizophrenic responses is a possible weakness of this model.
Model 1 (cont’d)

- Crude estimate of parameters

\[
\hat{\alpha}_j = \frac{1}{30} \sum_{i=1}^{30} y_{ij}
\]

\[
\hat{\sigma}_y^2 = \frac{1}{11} \sum_{j=1}^{11} \left[ \frac{1}{29} \sum_{i=1}^{30} (y_{ij} - \bar{y}_j)^2 \right]
\]

\[
\hat{\mu} = \frac{1}{11} \sum_{j=1}^{11} \hat{\alpha}_j
\]

\[
\hat{\beta} = \frac{1}{6} \sum_{j=12}^{17} \hat{\alpha}_j - \frac{1}{11} \sum_{j=1}^{11} \hat{\alpha}_j
\]

\[
\hat{\sigma}_\alpha^2 = \frac{1}{16} \sum_{j=1}^{17} (\alpha_j - \bar{\alpha})^2
\]

It is not necessary to create a preliminary estimates of the indicator variables, \( \zeta_{ij} \), because they are updated as the first step in the ECM and Gibbs sampler computations.
Model 1: ECM algorithm

• **Starting points:** 100 points were drawn at random from a simplified distribution for \( \phi \). Each of those points was used as a starting point for the ECM maximization algorithm to search for modes.

  – The simplified distribution is obtained by adding some randomness to the crude parameter estimates. Specifically, to obtain a sample from the simplified distribution, all parameters are set at the crude point estimates and then divided each parameter by an independent \( \chi_1^2 \) random variable in an attempt to cover the modes of the parameter space with marginal Cauchy distributions (i.e., to ensure that the 100 draws were sufficiently spread out).
Model 1: ECM algorithm (cont’d)

- **E step:** we determine the expected joint log posterior density, averaging $\zeta$ over its posterior distribution given the last guessed value of $\theta^{old}$.

The expected “complete-data” log posterior density is:

$$E_{old}(\log p(\zeta, \theta | y)) = \text{const.} + \sum_{j=1}^{17} \log(N(\alpha_j | \mu + \beta S_j, \sigma^2_\alpha)) +$$

$$\sum_{j=1}^{17} \sum_{i=1}^{30} \left[ \log(N(y_{ij} | \alpha_j, \sigma^2_y))(1 - E_{old}(\zeta_{ij})) + \log(N(y_{ij} | \alpha_j + \tau, \sigma^2_y))E_{old}(\zeta_{ij}) \right] +$$

$$\sum_{j=12}^{17} \sum_{i=1}^{30} \left[ \log(1 - \lambda)(1 - E_{old}(\zeta_{ij})) + \log(\lambda)E_{old}(\zeta_{ij}) \right]$$

We must compute $E_{old}(\zeta_{ij})$ for each observation $(i, j)$. Given $\theta^{old}$ and $y$, the indicators $\zeta_{ij}$ are independent, with conditional posterior densities,

$$P(\zeta_{ij} = 0 | \theta^{old}, y) = 1 - z_{ij}$$

$$P(\zeta_{ij} = 1 | \theta^{old}, y) = z_{ij}$$
Model 1: ECM algorithm

• E step (cont’d):

Thus $E_{\text{old}}(\zeta_{ij}) = z_{ij}$, where

$$z_{ij} = \frac{\lambda_{\text{old}}N(y_{ij}|\alpha_{ij}^{\text{old}}+.5\tau^{\text{old}},\sigma_{y}^{2\text{old}})}{(1-\lambda_{\text{old}})N(y_{ij}|\alpha_{ij}^{\text{old}},\sigma_{y}^{2\text{old}})+\lambda_{\text{old}}N(y_{ij}|\alpha_{ij}^{\text{old}}+.5\tau^{\text{old}},\sigma_{y}^{2\text{old}})}$$

$$= \left[ 1 + \frac{(1-\lambda_{\text{old}})}{\lambda_{\text{old}}} \exp(\sigma_{y}^{-2\text{old}}((\alpha_{ij}^{\text{old}} - y_{ij})\tau^{\text{old}} + .5(\tau^{\text{old}})^2)) \right]^{-1}$$

For each $(i, j)$, the above expression is a function of $(y, \theta)$ and can be computed based on the data $y$ and the current guess, $\theta^{\text{old}}$.

• M step:

(a) Update $\lambda$

$$\lambda_{\text{new}} = \frac{1}{6 \times 30} \sum_{j=12}^{17} \sum_{i=1}^{30} z_{ij}$$

(b) Update $\alpha_{j}$. For each $j$,

$$\alpha_{j}^{\text{new}} = \frac{\sigma_{y}^{2}(\mu + \beta S_{j}) + \sigma_{\alpha}^{2} \sum_{i=1}^{30} (y_{ij} - z_{ij}\tau)}{\sigma_{y}^{2} + 30\sigma_{\alpha}^{2}}$$
Model 1: ECM algorithm

- M step: (cont’d)

  (c) Update $\tau$

  \[
  \tau_{\text{new}} = \frac{\sum_{j=12}^{17} \sum_{i=1}^{30} z_{ij} (y_{ij} - \alpha_j)}{\sum_{j=12}^{17} \sum_{i=1}^{30} z_{ij}}
  \]

  (d) Update $\sigma_y^2$

  \[
  \sigma_{y_{\text{new}}}^2 = \frac{1}{17 \times 30} \sum_{j=12}^{17} \sum_{i=1}^{30} (y_{ij} - \alpha_j - z_{ij} \tau)^2
  \]

  (e) Update $\mu$

  \[
  \mu_{\text{new}} = \frac{1}{11} \sum_{j=1}^{11} \alpha_j
  \]

  (f) Update $\beta$

  \[
  \beta_{\text{new}} = \frac{1}{6} \sum_{j=12}^{17} \alpha_j - \frac{1}{11} \sum_{j=1}^{11} \alpha_j
  \]

  (g) Update $\sigma_\alpha^2$

  \[
  \sigma_{\alpha_{\text{new}}}^2 = \frac{1}{17} \sum_{j=1}^{17} (\alpha_j - \mu - \beta S_j)^2
  \]
Model 1: Creating an approximate distribution

- After 100 iterations of ECM from each of 100 starting points, three local maxima of \((\alpha, \phi)\) were found: a major mode and two minor modes.

The minor modes are substantively uninteresting, corresponding to near-degenerate models with the mixture parameter \(\lambda\) near 0, and had little support in the data. Thus, they were ignored, and the target distribution could be considered unimodal for practical purposes. Had we included those two minor modes, any draws from them would have had essentially zero importance weights and would almost certainly have not appeared in the importance-weighted samples.

- Once the mode has been found, a multivariate \(t_4\) approximation for \(\theta\) was constructed. The multivariate \(t_4\) distribution was centered at the mode with scale determined by the second derivative matrix at the mode, which was computed by numerical differentiation.
Model 1: Creating an approximate distribution (cont’d)

- 2,000 independent random samples of $\theta$ were drawn from the multivariate $t$ approximation. An histogram of the relative values of the 1,000 largest log importance weights shows little variation – an indication of the adequacy of the overdispersed approximation as a basis for taking draws to be resampled to create the starting distribution.

Logarithms of the largest importance ratios from the multivariate $t$ approximation
Model 1: Why all this trouble?

Why do we have to draw the starting points of the Gibbs sampler by importance-weighted resampling.

- It is possible for the Gibbs sampler to exhibit slow convergence.

To illustrate this point, 10 sequences of 200 steps were drawn using Gibbs sampling. However, the starting points were drawn directly from the initial approximate distribution (i.e. the one whose marginals are Cauchy distributions). Inference for some scalar estimads were based on the last halves of the sequences.

<table>
<thead>
<tr>
<th></th>
<th>Posterior interval</th>
<th>Potential scale reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>2.5%</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.27</td>
<td>-0.17</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.84</td>
<td>0.70</td>
</tr>
<tr>
<td>-2log(density)</td>
<td>757.18</td>
<td>681.63</td>
</tr>
</tbody>
</table>
Model 1: Why all this trouble?  
(cont’d)

- The high potential scale reductions clearly show that the simulations are far from convergence. To understand better what is happening, the last halves of the 10 sequences of log posterior densities were plotted.

*Log posterior densities for ten simulated sequences*
Model 1: Why all this trouble? (cont’d)

• The single sequence that stands alone started and remains in the neighborhood of one of the minor modes found earlier by maximization. Since the minor mode is of no scientific interest and has a negligible support in the data, we discard this sequence, as almost certainly would have occurred with importance resampling.

• Thus, the use of an overdispersed starting distribution, which, if well chosen, can lead to conservative yet relatively efficient inferences is all-important.
Model 1: Gibbs sampler

- Let $\theta = (\alpha, \phi)$. Then, the marginal posterior distribution of the model parameters $\theta$ is a product of mixture forms:

$$p(\theta|y) \propto \prod_{j=1}^{17} N(\alpha_j|\mu + \beta S_j, \sigma^2_\alpha) \times \prod_{j=1}^{17} \prod_{i=1}^{30} \left[ (1 - \lambda)N(y_{ij}|\alpha_j, \sigma^2_y) + \lambda S_j N(y_{ij}|\alpha_j + \tau, \sigma^2_y) \right]$$

- The Gibbs sampler is easy to apply for our model because the full conditional posterior distributions – $p(\phi|\alpha, \zeta, y)$, $p(\alpha|\phi, \zeta, y)$, and $p(\zeta|\phi, \alpha, y)$ – have standard forms and can be easily sampled from.

- Recall that a set of ten starting points was drawn by importance resampling from the $t_4$ approximation centered at the major mode. This distribution is intended to approximate our ideal starting conditions: for each scalar stimand of interest, the mean is close to the target mean and the variance is greater than the target variance.
**Model 1: One cycle of the Gibbs sampler**

**Step 1** - Update $\zeta_{ij}$ for $(i, j) \in \{1, \ldots, 30\} \times \{12, \ldots, 17\}$

$$\zeta_{ij} | \theta, y \sim \text{Bernoulli}(z_{ij})$$

The indicators $\zeta_{ij}$ are fixed at 0 for the non-schizophrenic sub-
jects ($j < 12$).

**Step 2** - Update $\alpha_j$ for $j \in \{1, \ldots, 17\}$

$$\alpha_j | \phi, \zeta, y \sim \text{N} \left( \frac{\sigma_y^2 (\mu + \beta S_j) + \sigma_\alpha^2 \sum_{i=1}^{30} (y_{ij} - \zeta_{ij} \tau)}{\sigma_y^2 + 30 \sigma_\alpha^2}, \frac{1}{\sigma_\alpha^2} + \frac{30}{\sigma_y^2} \right)$$

**Step 3** - Update $\lambda$

$$\lambda | \tau, \mu, \sigma_\alpha^2, \sigma_y^2, \alpha, \zeta \sim \text{Beta}(h + 1, 180 - h + 1)$$

where $h = \sum_{j=12}^{17} \sum_{i=1}^{30} \zeta_{ij}$

**Step 4** - Update $\sigma_y^2$

$$\sigma_y^2 | \alpha, \lambda, \zeta \sim \text{Inv-} \chi^2 \left( 508, \frac{1}{508} \sum_{j=1}^{17} \sum_{i=1}^{30} (y_{ij} - \alpha_j - \zeta_{ij} \tau)^2 \right)$$
Model 1: One cycle of the Gibbs sampler (cont’d)

Step 5 - Update $\sigma^2_{\alpha}$

$$\sigma^2_{\alpha}|\alpha, \lambda, \zeta \sim \text{Inv-}\chi^2 \left(15, \frac{1}{15} \sum_{j=1}^{17} (\alpha_j - \mu - \beta S_j)^2\right)$$

Step 6 - Update $\tau$

$$\tau|\alpha, \lambda, \zeta, \sigma^2_{\alpha}, \sigma^2_y \sim N \left(\frac{\sum_{j=12}^{17} \sum_{i=1}^{30} \zeta_{ij} (y_{ij} - \alpha_j)}{\sum_{j=12}^{17} \sum_{i=1}^{30} \zeta_{ij}}, \frac{\sigma^2_y}{\sum_{j=12}^{17} \sum_{i=1}^{30} \zeta_{ij}}\right)$$

Step 7 - Update $\mu$

$$\mu|\alpha, \lambda, \zeta, \sigma^2_{\alpha}, \sigma^2_y \sim N \left(\frac{1}{17} \sum_{j=1}^{17} (\alpha_j - \beta S_j), \frac{\sigma^2_{\alpha}}{17}\right)$$

Step 8 - Update $\beta$

$$\beta|\alpha, \lambda, \zeta, \sigma^2_{\alpha}, \sigma^2_y \sim N \left(\frac{1}{6} \sum_{j=12}^{17} (\alpha_j - \mu), \frac{\sigma^2_{\alpha}}{6}\right)$$

Step 9 - Return to Step 1 and begin a new iteration
Model 1: Gibbs sampler (cont’d)

• Possible difficulties at a degenerate point

If all $\zeta_{ij}$’s are zero, then the mean and variance of the conditional distribution of $\beta$ are undefined, because $\tau$ has an improper prior distribution, and, conditional on $\sum_{ij} \zeta_{ij} = 0$, there are no delayed reactions and thus no information about $\tau$. Strictly speaking, this means that our posterior distribution is improper. For the data at hand, however, this degenerate point has extremely low posterior probability and is not reached by any of our simulations. If the data were such that $\sum_{ij} \zeta_{ij} = 0$ were a realistic possibility, it would be necessary to assign an informative prior distribution for $\tau$.

• Gibbs sampler implementation

Ten sequences of 200 iterations each were simulated. The first half of each sequence was discarded. Hence, we obtain posterior intervals for all quantities of interest from the quantiles of the 1000 simulations from the second halves of the sequences.
Model 1: Gibbs sampling results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>mean</th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
<th>Potential scale reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>5.73</td>
<td>5.66</td>
<td>5.71</td>
<td>5.73</td>
<td>5.76</td>
<td>5.80</td>
<td>1.00 1.00</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>5.89</td>
<td>5.82</td>
<td>5.86</td>
<td>5.89</td>
<td>5.91</td>
<td>5.95</td>
<td>1.00 1.00</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>5.71</td>
<td>5.64</td>
<td>5.69</td>
<td>5.71</td>
<td>5.73</td>
<td>5.78</td>
<td>1.00 1.01</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>5.71</td>
<td>5.64</td>
<td>5.68</td>
<td>5.71</td>
<td>5.73</td>
<td>5.77</td>
<td>1.00 1.02</td>
</tr>
<tr>
<td>$\alpha_5$</td>
<td>5.58</td>
<td>5.51</td>
<td>5.56</td>
<td>5.58</td>
<td>5.60</td>
<td>5.65</td>
<td>1.00 1.01</td>
</tr>
<tr>
<td>$\alpha_6$</td>
<td>5.80</td>
<td>5.73</td>
<td>5.77</td>
<td>5.80</td>
<td>5.82</td>
<td>5.86</td>
<td>1.00 1.00</td>
</tr>
<tr>
<td>$\alpha_7$</td>
<td>5.86</td>
<td>5.79</td>
<td>5.83</td>
<td>5.86</td>
<td>5.88</td>
<td>5.92</td>
<td>1.00 1.00</td>
</tr>
<tr>
<td>$\alpha_8$</td>
<td>5.59</td>
<td>5.52</td>
<td>5.56</td>
<td>5.59</td>
<td>5.61</td>
<td>5.65</td>
<td>1.00 1.00</td>
</tr>
<tr>
<td>$\alpha_9$</td>
<td>5.55</td>
<td>5.48</td>
<td>5.53</td>
<td>5.55</td>
<td>5.57</td>
<td>5.62</td>
<td>1.00 1.00</td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>5.77</td>
<td>5.71</td>
<td>5.75</td>
<td>5.77</td>
<td>5.80</td>
<td>5.84</td>
<td>1.00 1.01</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>5.72</td>
<td>5.65</td>
<td>5.69</td>
<td>5.72</td>
<td>5.74</td>
<td>5.78</td>
<td>1.00 1.01</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>5.73</td>
<td>5.66</td>
<td>5.71</td>
<td>5.73</td>
<td>5.75</td>
<td>5.80</td>
<td>1.00 1.00</td>
</tr>
<tr>
<td>$\alpha_{13}$</td>
<td>6.03</td>
<td>5.97</td>
<td>6.01</td>
<td>6.03</td>
<td>6.05</td>
<td>6.10</td>
<td>1.00 1.00</td>
</tr>
<tr>
<td>$\alpha_{14}$</td>
<td>6.01</td>
<td>5.93</td>
<td>5.98</td>
<td>6.01</td>
<td>6.04</td>
<td>6.09</td>
<td>1.00 1.01</td>
</tr>
<tr>
<td>$\alpha_{15}$</td>
<td>6.19</td>
<td>6.08</td>
<td>6.15</td>
<td>6.19</td>
<td>6.22</td>
<td>6.29</td>
<td>1.03 1.07</td>
</tr>
<tr>
<td>$\alpha_{16}$</td>
<td>6.19</td>
<td>6.11</td>
<td>6.16</td>
<td>6.19</td>
<td>6.22</td>
<td>6.26</td>
<td>1.01 1.03</td>
</tr>
<tr>
<td>$\alpha_{17}$</td>
<td>6.07</td>
<td>6.00</td>
<td>6.04</td>
<td>6.07</td>
<td>6.09</td>
<td>6.14</td>
<td>1.01 1.02</td>
</tr>
<tr>
<td>$\sigma_\alpha$</td>
<td>0.14</td>
<td>0.09</td>
<td>0.12</td>
<td>0.14</td>
<td>0.16</td>
<td>0.21</td>
<td>1.00 1.00</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.32</td>
<td>0.17</td>
<td>0.27</td>
<td>0.32</td>
<td>0.37</td>
<td>0.48</td>
<td>1.01 1.02</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.12</td>
<td>0.07</td>
<td>0.10</td>
<td>0.12</td>
<td>0.14</td>
<td>0.18</td>
<td>1.02 1.04</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.85</td>
<td>0.74</td>
<td>0.81</td>
<td>0.85</td>
<td>0.88</td>
<td>0.96</td>
<td>1.02 1.05</td>
</tr>
<tr>
<td>$\sigma_y$</td>
<td>0.19</td>
<td>0.18</td>
<td>0.18</td>
<td>0.19</td>
<td>0.19</td>
<td>0.20</td>
<td>1.01 1.02</td>
</tr>
<tr>
<td>$\sigma_\alpha/\sigma_y$</td>
<td>0.74</td>
<td>0.50</td>
<td>0.64</td>
<td>0.73</td>
<td>0.85</td>
<td>1.11</td>
<td>1.00 1.00</td>
</tr>
</tbody>
</table>

-2log(density)  747.33  727.81  739.92  746.88  753.84  768.35  1.01  1.01
Model 1: Gibbs sampler

Inference from the iterative simulations

- Estimated potential scale reductions The estimated potential scale reductions shown in the previous table are close to 1, they suggest that further simulation will not markedly improve our estimates of the scalar stimands shown.

More precise estimation of the means and variances of the target distribution, as would be achieved by further simulation, would not narrow the estimated posterior intervals much – nearly all the width of the intervals is due to the posterior variances themselves, not uncertainty due to simulation variability.
Model 1: Gibbs sampler
Inference from the iterative simulations

• Results for this model suggest that approximately 7-18 per cent of each schizophrenic’s responses suffer from an attentional lapse ($\lambda$), with a $110\% = e^{.74} - 1$ to $161\% = e^{.96} - 1$ increase in mean response time due to this lapse ($\tau$). The interpretation here is based on the fact that $\tau$ measures increases on the scale of $100 \times \log$(response time). Note that the 95% interval for $\tau$ also excludes the null value of zero, which is on the boundary of the parameter space; the real reason for rejecting the model with $\lambda = 0$ is not this interval, but rather the evidence that a simple linear model does not come close to explaining the observed variability in the data. This group of schizophrenics also shows evidence of motor retardation resulting in $(e^{.17} - 1) = 19\%$ to $(e^{.48} - 1) = 62\%$ slower mean reaction times ($\bar{\alpha}_{schiz} - \bar{\alpha}_{non-schiz}$). These results suggest a rather dramatic attentional delay when it occurs, though it appears to be infrequent, combined with a more modest delay due to motor retardation.
Model 1: Checking the model

Inference from the iterative simulations (cont’d)

• Before accepting the results from model 1 as scientifically meaningful, we should be confident that the model is not contradicted by the data.

• Bayesian model monitoring using posterior predictive checks, can be used to see whether important observed features of the data could plausibly have arisen under the assumed model.

• Operationally, we examine whether the posterior predictive distribution could lead to future replicate data that look like our data set.

• This framework views data analysis as a process that involves model development, model fitting, and model monitoring, returning to the model development stage when inadequacies in model fit are found.
Model 1: Checking the model
Inference from the iterative simulations (cont’d)

• A general strategy for the model-monitoring step is to identify a statistic that is scientifically relevant and that would not automatically be well fit by the assumed model, then to generate a number of replicate data sets from the posterior distribution of parameters from the fitted model, and finally to ascertain whether the value of the statistic in the observed data is extreme relative to the distribution of the value of the statistic calculated from the generated data sets.

• The histograms of schizophrenics’ reaction times indicate that there is substantial variation in the within-person response time variance. To investigate whether the model can explain this feature of the data, we compute $s_j$, the standard deviation of the 30 log reaction times $y_{ij}$, for each schizophrenic individual ($j = 12, \ldots, 17$). Then, two test quantities were defined: $T_{\text{min}}$ and $T_{\text{max}}$, the smallest and largest of the six values $s_j$.  

31
Model 1: Checking the model

Inference from the iterative simulations (cont’d)

• We simulate predictive datasets from the normal–mixture model for each of the 1000 draws of the parameters from the posterior distribution. For each of those 1000 simulated datasets, $y^{rep}$, we compute the two test quantities, $T_{min}(y^{rep})$ and $T_{max}(y^{rep})$

• We drew a scatterplot of the 1000 simulated values of the test quantities, with the observed values (indicated by ×).

• With regard to these test quantities, the observed data $y$ are atypical of the posterior predictive distribution – $T_{min}$ is too low and $T_{max}$ is too high with estimated $p$-values of 0.000 and 1.000 (to three decimal places).

• A reformulation of the model is needed to fit the data more accurately.
Expanding the model

The following models are possible extensions of model 1.

• Model 2

A first extension of model 1, model 2 allows the variance of shifted responses to be different from the variance of non-shifted responses for schizophrenics.

\[
y_{ij} | \alpha_j, \zeta_{ij}, \phi \sim N(\alpha_j + \tau \zeta_{ij}, (1 - \zeta_{ij})\sigma_{y1}^2 + \zeta_{ij}\sigma_{y2}^2)
\]

\[
\alpha_j | \zeta, \phi \sim N(\mu + \beta S_j, \sigma_\alpha^2)
\]

\[
\zeta_{ij} | \phi \sim \text{Bernoulli}(\lambda S_j)
\]
Expanding the model (cont’d)

• Model 3

Model 3 postulates that there are two types of schizophrenic individuals: those who are not susceptible to attentional deficiency, and those who have a proportion $\lambda$ of shifted response times due to attentional deficiency. Consequently, a missing indicator $W_j$ for individual $j$ which equals 1 if some responses can be shifted and 0 if they are never shifted is introduced. The separate variances assumption of model 2 is dropped in favor of the common variance of model 1 to see if the indicator $W_j$’s alone can account for the spread in variances among schizophrenics.

\[
y_{ij}|\alpha_j, \zeta_{ij}, \phi \sim N(\alpha_j + \tau \zeta_{ij}, \sigma_y^2)\\
\alpha_j|\zeta, \phi \sim N(\mu + \beta S_j, \sigma_\alpha^2)\\
\zeta_{ij}|\phi \sim \text{Bernoulli}(\lambda S_j W_j)\\
W_j|S, \theta \sim \text{Bernoulli}(\omega S_j)
\]
Expanding the model (cont’d)

- Model 4

Model 4 includes both the possibility that not all schizophrenic individuals exhibit attentional deficiency as well as the possibility that response times shifted due to attentional deficiency have different variances from non-shifted response times.

\[ y_{ij} | \alpha_j, \zeta_{ij}, \phi \sim N(\alpha_j + \tau \zeta_{ij}, (1 - \zeta_{ij}) \sigma_{y1}^2 + \zeta_{ij} \sigma_{y2}^2) \]

\[ \alpha_j | \zeta, \phi \sim N(\mu + \beta S_j, \sigma^2_{\alpha}) \]

\[ \zeta_{ij} | \phi \sim \text{Bernoulli}(\lambda S_j W_j) \]

\[ W_j | S, \theta \sim \text{Bernoulli}(\omega S_j) \]
Expanding the model (cont’d)

• **Possible alternatives to model 4**

Possible extensions of model 4 include

– letting $\tau$ vary across individual schizophrenics and estimating a mean and a variance for the individual values; e.g. assuming

$$\tau_j \sim N(\mu_\tau, \sigma^2_\tau)$$

– letting $\lambda$ vary across individual schizophrenics; e.g. assuming either

$$\lambda_j \sim \text{Beta}(\lambda_1, \lambda_2)$$

or

$$\text{logit}(\lambda_j) \sim N(\mu_\lambda, \sigma^2_\lambda)$$

• Any of these alternatives would naturally be accompanied by treating person means for non-schizophrenics and the means of the non-shifted component of schizophrenic reaction times as random effects, i.e. $\alpha_j \sim N(\mu_\alpha, \sigma^2_\alpha)$.