A Bayesian Approach for Joint Modeling of Cluster Size and Subunit-Specific Outcomes (Dunson, Chen, and Harry, 2003)

Valerie Poynor

March 11, 2010
Scenario

In studies involving clustered data (ie. longitudinal studies and developmental toxicity experiments), it is often the case that the size of the cluster is associated with outcomes measured on the individual subunits within the cluster.

When variables are correlated it is desirable to model them jointly. However, most methods assume that the cluster size is either independent of the individual subunit outcomes OR the model is conditioned on the cluster size. Neither approach captures the underlying association between the two variables.
**Example**: Rodent Teratology Studies

A dam is randomly placed into a treatment group that will be exposed to a particular dose level of a developmental toxicant at some stage (generally during organogenesis) over a period of time. In this situation, we consider the litters to be the clusters and the fetuses to be the subunits. Various outcomes such as individual fetal weight are measured on the subunits.

For an untreated dam (under good conditions) we would expect to see:

\[
\begin{align*}
\underbrace{\text{fetal weight}} & \quad \overset{(-)}{\leftrightarrow} \quad \underbrace{\text{litter size}} \\
\downarrow (\uparrow) & \quad \quad \quad \quad \quad \quad \uparrow (\downarrow)
\end{align*}
\]

A Bayesian Approach for Joint Modeling of Cluster Size and Subunit-Specific Outcomes (Dunson, Chen, and Harry, 2003)
Now, consider a dam that has been exposed to a known developmental toxin. Then we might see the following trend:

<table>
<thead>
<tr>
<th>fetal weight</th>
<th>dosage level</th>
</tr>
</thead>
<tbody>
<tr>
<td>↓(↑)</td>
<td>↑(↓)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>litter size</th>
<th>dosage level</th>
</tr>
</thead>
<tbody>
<tr>
<td>↓(↑)</td>
<td>↑(↓)</td>
</tr>
</tbody>
</table>

Under the independent framework, this would lead us to believe that as dose goes up we would see both litter size AND fetal weight go down, but this is obviously not the entire story. The correlation between litter size and fetal weight may make it seem like the dose level has less effect than in reality or vice versa. Similarly, the conditioning on litter size could also lead to false conclusions since litter size is correlated with both dose level and fetal weight in a competing manner.
Solution: Model fetal-specific outcomes and litter size JOINTLY!!!
Basic Notation and Set-Up

- **Indices:**
  - \( i = 1, \ldots, n \) indicates the \( i \)th cluster
  - \( j = 1, \ldots, s_i \) indicates the \( j \)th subunit in the \( i \)th cluster \((s_i \text{ is the number of subunits in the } i \text{th cluster})\)
  - \( k = 1, \ldots, p \) indicates the outcome variable which wlog is continuous for \( k = 1, \ldots, p_c \) and binary for \( k = p_c + 1, \ldots, p \)

- **\( p \)-dimensional outcome vector:** \( y_{ij} = (y_{ij1}, \ldots, y_{ijp})^T \)

- As in Muthen (1984), the observed outcome \( y_{ijk} \) is linked to an underlying normal variable \( y^*_{ijk} \) through \( g_k(\cdot) \)

\[
y_{ijk} = g_k(y^*_{ijk}) \quad \text{where}
\]

- \( g_k(y^*_{ijk}) = I(y^*_{ijk} > 0) \) for binary outcome
- \( g_k(y^*_{ijk}) \) is a known 1-1 function (ie. identity or log) for continuous outcome
The multilevel regression model for \( y_{ijk}^* = (y_{ij1}^*, \ldots, y_{ijp}^*)^T \) is given by

\[
y_{ij}^* = \mu + \alpha x_i + \Lambda W_{1i} \xi_i + \Gamma W_{2i} \eta_{ij} + \epsilon_{ij}
\]

- \( \mu = (\mu_1, \ldots, \mu_p)^T \) are outcome-specific intercept parameters,
- \( \alpha = [\alpha_1, \ldots, \alpha_p]^T \) is a pxq matrix of regression coefficients,
- \( x_i \) is a qx1 vector of cluster-level covariates,
- \( \Lambda = [\lambda_1, \ldots, \lambda]^T, \Gamma = [\gamma_1, \ldots, \gamma]^T \) are pxr factor loadings matrices,
- \( W_{hi} = \text{diag}(w_{hi1}, \ldots, w_{hir}) \) for \( h = 1, 2 \) are known diagonal weighting matrices,
- \( \xi_i = (\xi_{i1}, \ldots, \xi_{ir})^T, \eta_{ij} = (\eta_{ij1}, \ldots, \eta_{ijr})^T \) are vectors of iid \( N(0,1) \) cluster and subunit-level latent variables, respectively,
- \( \epsilon_{ij} = (\epsilon_{ij1}, \ldots, \epsilon_{ijp})^T \sim N_p(0, \Sigma) \) are error residuals with \( \Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2) \)
- where the elements of \( \xi_i, \eta_{ij}, \) and \( \epsilon_{ij} \) are independent
As for the cluster size $s_i$ we use (for $j = 1, \ldots, T - 1$)

$$Pr(s_i = j | s_i \geq j, x_i, \xi_i) = F(\delta_j - x_i \beta - \lambda_{p+1}^T W_{1i} \xi_i)$$

$$Pr(s_i = j | x_i, \xi_i) = F(\delta_j - x_i \beta - \lambda_{p+1}^T W_{1i} \xi_i)$$

$$\times \prod_{h=1}^{j-1} (1 - F(\delta_j - x_i \beta - \lambda_{p+1}^T W_{1i} \xi_i))$$

- $F(.)$ is a known 1-1 function mapping $\mathbb{R} \rightarrow [0, 1]$
- $\beta$ are regression coefficients
- $\delta = (\delta_1, \ldots, \delta_{T-1})^T \in \mathbb{R}^{T-1}$ are intercept parameters that characterize the baseline cluster size distribution given $x_i = 0$ (support $\{1, \ldots, T\}$).
- the subscript $j$ is added to $x_i$ to allow consideration for models in which the shape of the cluster distribution varies for subjects in different groups
- $\lambda_{p+1}^T W_{1i} \xi_i$ flexibly accommodates dependency between $s_i$ and $y_i = (y_{i1}, \ldots, y_{isi})$ all the while reducing dimensionality
Let $\theta = (\theta_1^T, \theta_2^T)^T$

where $\theta_1 = (\mu^T, \alpha_1^T, ..., \alpha_p^T, \lambda_1^T, ..., \lambda_p^T, \gamma_1^T, ..., \gamma_p^T)^T$ are the parameters in expression (1)

and $\theta_2 = (\delta^T, \beta^T, \lambda_{p+1}^T)^T$ are the parameters in expression (2)

then place priors:

$$\theta \sim N(\theta_0, \Psi_0)$$

$$\sigma_k^{-2} \sim \text{gamma}(c_{0k}, d_{0k})$$

where the hyperparameters, $\theta_0, \Psi_0, c_{0k}$ and $d_{0k}$ are specified by the investigator. To avoid identifiability problems, $\sigma_k = 1$, if the kth outcome is binary (for $k = p_c, ..., p$)
Considering $g_k(.)$ to be the identity link important variances and covariances are given by

\[
\begin{align*}
\text{var}(y_{ij}) &= \Lambda W_{1i} W_{1i} \Lambda^T + \Gamma W_{2i} W_{21} \Gamma^T + \Sigma \\
\text{cov}(y_{ij}, y_{i'j}) &= 0 \\
\text{cov}(y_{ij}, y_{ij'}) &= \Lambda W_{1i} W_{1i} \Lambda^T \\
\text{cov}(y_{ijk}, y_{ijk'}) &= \lambda_k^T W_{1i} W_{1i} \lambda_{k'} + \gamma_k^T W_{2i} W_{2i} \gamma_{k'} \\
\text{cov}(y_{ijk}, s_i) &= \int \lambda_k^T W_{1i} \xi_i E(s_i|x_i, \xi_i) d\Phi(\xi_{i1})...d\Phi(\xi_{ir})
\end{align*}
\]

where $E(s_i|x_i, \xi_i) = \sum_{j=1}^T jPr(s_i = j|x_i, \xi_i)$
The correlation between the cluster size and the kth outcome for the jth subunit in the ith cluster can be calculated by use of (4) and (5):

$$
\rho(y_{ijk}, s_i) = \frac{\int \lambda_k W_{1i} \xi_i E(s_i|x_i, \xi_i) d\Phi(\xi_{i1})...d\Phi(\xi_{ir})}{\sqrt{(\lambda_k W_{1i} W_{1i}^T \lambda_k + \gamma_k W_{2i} W_{2i}^T \gamma_k + \sigma_k) \int \text{var}(s_i|x_i, \xi_i) d\Phi(\xi_{i1})...d\Phi(\xi_{ir})}}
$$

where $\text{var}(s_i|x_i, \xi_i) = \sum_{j=1}^{T} j^2 Pr(s_i|x_i, \xi_i) - \left\{ E(s_i|x_i, \xi_i) \right\}^2$.

The correlation between the kth subunit-level outcome and is 0 only if $\lambda_{kl} \lambda_{p+1,l} = 0$. 
Biasness of ignoring cluster size

By calculating the expectation of $y_{ijk}$ given the size of the cluster, we can obtain the amount of bias created by modeling cluster size and subunit-specific outcomes independently:

$$E(y_{ijk}|x_i = x, s_i = j) = \int E(y_{ijk}|x, \xi_i) \pi(\xi_i|x, s_i = j) d\xi_i$$

$$= \mu_k + x^T \alpha_k$$

EV ignoring $s_i$

$$+ \frac{\int \lambda_k^T W_1 \xi_i Pr(s_i = j|x, \xi_i) d\Phi(\xi_{i1})...d\Phi(\xi_{ir})}{\int Pr(s_i = j|x, \xi_i) d\Phi(\xi_{i1})...d\Phi(\xi_{ir})}$$

bias

A Bayesian Approach for Joint Modeling of Cluster Size and Subunit-Specific Outcomes (Dunson, Chen, and Harry, 2003)
Constraints are needed on the variance components \((\Lambda, \Gamma, \Sigma)\)

Case when \(W_{hi} = I_{r \times r}\) and \(\lambda_{p+1} = 0\)

\(\rightarrow (1)\) is a standard multilevel factor analytic model

\(\rightarrow \sigma_k = 1\) for binary outcomes

\(\rightarrow \Lambda\Lambda^T\) and \(\Gamma\Gamma^T\) must be uniquely defined by \(\Lambda\) and \(\Gamma\), respectively, by placing restrictions on the later (e.g. for \(p \geq 3\) set the upper diagonals to 0 and choose \(r \leq p/2\))

In the case where the predictors are elements in the weighting matrix, the standard restrictions apply to the submatrices \(\Lambda_l\) and \(\Gamma_l\) where \(l = 1, \ldots, m\) (\(m \leq q\) the number of predictor elements in the weighting matrix.)
Developmental Toxicology Experiment by National Toxicology Program, (Price et al., 1985)

During organogenesis, pregnant dams were exposed to one of four different dose levels of ethylene glycol. Fetal weight (continuous variable) and percentage of malformations (indicator is a binary variable) in each litter were measured.

<table>
<thead>
<tr>
<th>Dose (mg/kg)</th>
<th>No. dams</th>
<th>Mean litter size</th>
<th>Mean fetal weight (g)</th>
<th>Malformation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>11.9</td>
<td>0.97</td>
<td>0.3</td>
</tr>
<tr>
<td>0.75</td>
<td>24</td>
<td>11.5</td>
<td>0.88</td>
<td>9.4</td>
</tr>
<tr>
<td>1.50</td>
<td>22</td>
<td>10.4</td>
<td>0.76</td>
<td>38.9</td>
</tr>
<tr>
<td>3.00</td>
<td>23</td>
<td>9.8</td>
<td>0.70</td>
<td>57.1</td>
</tr>
</tbody>
</table>
Data

A Bayesian Approach for Joint Modeling of Cluster Size and Subunit-Specific Outcomes (Dunson, Chen, and Harry, 2003)
Fetal weight:

\[ y_{ijk} = \mu_1 + \alpha_1 x_i + \lambda_1 \xi_i + \gamma_1 \eta_{ij} + \epsilon_{ij1} \]

where \( \lambda_1, \gamma_1 > 0 \) for identifiability, \( \epsilon_{ij1} \sim N(0, \sigma^2) \).

Malformation: let \( y_{ij2} = 1 \) if the jth fetus is malformed and 0 o.w. Let \( y_{ij2}^* \) represent the underlying normal variable such that \( y_{ij2} = I(y_{ij2}^* > 0) \),

\[ y_{ij2}^* = \mu_2 + \alpha_2 x_i + \lambda_2 \xi_i + \gamma_2 \eta_{ij} + \epsilon_{ij2} \]

where \( \epsilon_{ij1} \sim N(0, 1) \) for identifiability. For both model parts, \( \xi_i \) and \( \eta_{ij} \) are independently distributed standard normal random latent variables as described earlier.
Litter Size: Mixed effects extension of the continuation ratio probit model.

\[
Pr(s_i = j|x_i, \xi_i) = \Phi(\delta_j - \beta x_i - \lambda_3 \xi_i) \\
\times \prod_{h=1}^{j-1} \{1 - \Phi(\delta_h - \beta x_i - \lambda_3 \xi_i)\}
\]

for \(j = 1, \ldots, T - 1\). The vector \(\delta = (\delta_1, \ldots, \delta_{15})'\) characterizes the baseline (control group) distribution of the litter size. From historical data and expert experience it is assumed \(s_i \in \{1, \ldots, 16\}\).
A Bayesian Approach for Joint Modeling of Cluster Size and Subunit-Specific Outcomes  
(Dunson, Chen, and Harry, 2003)

Informative prior parameters for $\mu_1, \mu_2$, and $\delta$ were selected using historical data means and 10 times historical data estimated variances. All other parameters above were given mean 0 and variance 4 (considered large for the anticipated sizes of regression coefficients). The prior for the precision parameter was chosen to be $\sigma^{-2} \sim \text{gamma}(0.01, 0.01)$, which has mean 1 and variance 100.
Model and Analysis

A Bayesian Approach for Joint Modeling of Cluster Size and Subunit-Specific Outcomes (Dunson, Chen, and Harry, 2003)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Mean</th>
<th>SD</th>
<th>95% HPD credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weight model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$\mu_1$</td>
<td>0.955</td>
<td>0.014</td>
<td>(0.927, 0.984)</td>
</tr>
<tr>
<td>Dose</td>
<td>$\alpha_1$</td>
<td>-0.088</td>
<td>0.009</td>
<td>(-0.105, -0.071)</td>
</tr>
<tr>
<td>Litter-level latent variable</td>
<td>$\lambda_1$</td>
<td>0.09</td>
<td>0.008</td>
<td>(0.075, 0.104)</td>
</tr>
<tr>
<td>Fetus-level latent variable</td>
<td>$\gamma_1$</td>
<td>0.017</td>
<td>0.01</td>
<td>(0.001, 0.039)</td>
</tr>
<tr>
<td><strong>Malformation model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$\mu_2$</td>
<td>-3.284</td>
<td>0.96</td>
<td>(-5.194, -1.922)</td>
</tr>
<tr>
<td>Dose</td>
<td>$\alpha_2$</td>
<td>1.289</td>
<td>0.384</td>
<td>(0.716, 2.039)</td>
</tr>
<tr>
<td>Litter-level latent variable</td>
<td>$\lambda_2$</td>
<td>-0.88</td>
<td>0.28</td>
<td>(-1.425, -0.444)</td>
</tr>
<tr>
<td>Fetus-level latent variable</td>
<td>$\gamma_2$</td>
<td>-1.125</td>
<td>0.637</td>
<td>(-2.348, -0.083)</td>
</tr>
<tr>
<td><strong>Litter size model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose</td>
<td>$\beta$</td>
<td>-0.131</td>
<td>0.059</td>
<td>(-0.253, -0.02)</td>
</tr>
<tr>
<td>Litter-level latent variable</td>
<td>$\lambda_3$</td>
<td>-0.161</td>
<td>0.08</td>
<td>(-0.322, -0.009)</td>
</tr>
</tbody>
</table>
A Bayesian Approach for Joint Modeling of Cluster Size and Subunit-Specific Outcomes  (Dunson, Chen, and Harry, 2003)
MLE version and extension to outcome specific cluster level latent variable (Gueorguieva 2007)
  → Sample Size
  → Decent prior knowledge

Assumption of latent variable is normally distributed
  → Semiparametric methods (Dunson 2007)
A Bayesian Approach for Joint Modeling of Cluster Size and Subunit-Specific Outcomes (Dunson, Chen, and Harry, 2003)