Bayesian Latent Variable Models for Mixed Outcomes

- Most frequentist and Bayesian work on latent factor modeling of multivariate data has focused on normal linear structures, e.g.,

\[ y_i = \Lambda \xi_i + \epsilon_i \]

- Extensions to allow mixed categorical and continuous outcomes are straightforward using an underlying normal structure

\begin{align*}
  y_{ij} &= g_j(y_{i}^*; \tau_j), \quad \text{for } j = 1, \ldots, p \\
  y_{i}^* &= \Lambda \xi_i + \epsilon_i,
\end{align*}

where \( g_j(\cdot) \) is the identity link for continuous \( y_{ij} \), and a threshold link for categorical outcomes.

- Generalizations to incorporate covariates in the measurement model and on the level of the latent variable are straightforward.
Mixed Variables in the Exponential Family

- Clearly, the underlying normal framework is quite restrictive, and one may instead want to consider models of the form

\[ \eta_i = \Lambda \xi_i \]

where \( \eta_i = (\eta_{i1}, \ldots, \eta_{ip})' \) is a vector of linear predictors specific to the different outcomes.

- Then, outcomes can have different distributions in the exponential family and a separate GLM is defined for each outcome type.

- Because the same latent factors occur in the different GLMs, this structure accommodate correlation.

- Models in this class are sometimes referred to as generalized latent trait models.
Some Articles on Generalized Factor Models

• For a special case, Sammel, Ryan, and Legler (1997, *JRSS-B*) proposed models with a predictor incorporated on the level of the latent variable.

• Moustaki and Knott (2001, *Psychometrika*) described a more general class of models, and proposed simple algorithms of maximum likelihood estimation.

• Dunson (2000, *JRSS-B*) proposed a somewhat different Bayesian approach which accommodated observed and latent variables in the exponential family, while also allowing multi-level data structures.

• Dunson (2003, *JASA*) extended the generalized latent trait model to the longitudinal setting.
Dynamic Latent Trait Models for Multidimensional Longitudinal Data

- Let $y_{ij} = (y_{ij1}, \ldots, y_{ijp})'$ denote a vector of mixed scale outcomes for subject $i$ and time $j$.

- Let $\eta_{ij} = (\eta_{ij1}, \ldots, \eta_{ijp})'$ denote a vector of linear predictors corresponding to $y_{ij}$.

- Define a separate GLM for each outcome at each time point, including a term $\lambda_{jk}' \xi_{ij}$, where $\lambda_{jk}$ are factor loadings for the $j$th time point and $k$th outcome, and $\xi_{ij}$ are time-specific latent traits.
• Define an autoregressive normal model for the time-specific latent traits, with correlation parameters $\phi$.

• Choose prior distributions for regression parameters, factor loadings, residual variances (for normal outcomes), and autoregressive parameters.

• Posterior computation can then proceed via adaptive rejection sampling.
Non-normal latent variables

• In survival analysis and for count outcomes, gamma latent variables are more commonly used due to simplifications in computation and interpretation.

• Thus, in some settings, it might be preferable to consider alternatives to normal factor models.

• Let’s focus on the application to studies of tumorigenesis for motivation.
Skin and Breast Tumorigenesis Studies

- Collect data on time to first detection of tumor (*discrete event time*)
- Increases in tumor burden over time (*repeated counts*)
- Occurrence of internal tumors at death (*multiple binary*)
- Each outcome type can be thought to relate to an underlying count or counts, representing the number of tumors in different categories defined by time, site, and pathology.
Poisson-Gamma Latent Variable Framework
We consider a Poisson underlying variable framework, focusing ini-
tially on the case where outcomes are considered separately.

Count Data:
First suppose that data on subject $i$ consist of a single tumor count $y_i = z_i$. We focus initially on the simple Poisson log linear model:

$$z_i \sim \text{Poisson}(\eta_i), \quad \eta_i = \exp(x_i^T \beta), \quad (1)$$

where $x_i = (x_{i1}, \ldots, x_{iq})'$ is a vector of predictors, and $\beta = (\beta_1, \ldots, \beta_q)'$ are unknown regression coefficients.
**Binary Data:**

- In many cases, the exact number of tumors per animal \( z_i \) is unknown, and data consist instead of a binary indicator variable \( y_i \), which equals one if animal \( i \) has any tumors and zero otherwise.

- Hence, we have \( y_i = I(z_i > 0) \) which implies under expression (1) that

\[
\Pr(y_i = 1 \mid x_i) = 1 - \exp \left\{ - \exp(x_i' \beta) \right\}, \tag{2}
\]

a complementary log-log model. It is straightforward to extend this model to the ordered categorical case, in which the exact count is unknown but categories are available.
**Event Time:**

- Alternatively, the outcome may be the time of tumor onset.

- Let $t_i \in \{1, \ldots, J\}$ denote the minimum of the tumor onset time and the death (censoring) time, let $\delta_{ij} = I(t_i \geq j)$ be an at risk indicator, and let $y_{ij} = 1$ if tumor onset occurs at time $j$ and $y_{ij} = 0$ otherwise.

- If $z_{ij}$ denotes the number of tumors on animal $i$ at time $j$, then $y_{ij} = I(z_{ij} > 0)$ for $j = 1, \ldots, t_i$.

- Assuming that $z_{ij} \sim \text{Poisson}(\eta_{ij})$, where $\eta_{ij} = \exp(x_{ij}'\beta)$ and $x_{ij}$ is a vector of predictors,

$$\Pr(y_{ij} = 1 | \delta_{ij} = 1, x_{ij}) = 1 - \exp\{ - \exp(x_{ij}'\beta) \}, \quad (3)$$

which is a discrete-time version of the proportional hazards model.
• In each of these cases, the observed outcome $y_{ij}$ is linked to an underlying Poisson variable $z_{ij}$.

• Integrating out $z_{ij}$ results in a regression model for $y_{ij}$.

• Although this underlying Poisson structure is motivated by the tumor application, defining binary response and discrete-time hazards models in this manner has computational advantages.

• Relating mixed discrete outcomes $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})'$ to underlying variables $\mathbf{z}_i = (z_{i1}, \ldots, z_{in_i})'$ having a common scale (in this case, Poisson counts) has advantages in defining multivariate distributions for these mixed outcomes.
Additive Gamma Factor Models

• By incorporating shared latent variables in models for the underlying Poisson counts, we define a multivariate distribution for discrete outcome having mixed measurement scales.

• First, we generalize expression (1) to allow over-dispersion by using the standard Poisson-gamma shared frailty model, \( \eta_{ij} = \xi_i \exp(x_{ij}' \beta) \), where \( \xi_i \) is a frailty having a gamma \( \mathcal{G}(\phi^{-1}, \phi^{-1}) \) density with variance \( \phi \).

• The marginal expectation and variance of \( z_{ij} \) are

\[
E(z_{ij} \mid x_{ij}) = \exp(x_{ij}' \beta) \quad \text{and} \quad V(z_{ij} \mid x_{ij}) = \exp(x_{ij}' \beta) + \phi \exp(x_{ij}' \beta)^2,
\]

so that \( \phi \) is a measure of over-dispersion relative to the Poisson distribution.
• We generalize this model to allow for multivariate \( z_i = (z_{i1}, \ldots, z_{ip})' \) by letting

\[
\eta_{ij} = E(z_{ij} | \xi_i, x_{ij}) = (\xi_i' \lambda_j) \exp(x_{ij}' \beta),
\]

(4)

\[
\eta_i = (\eta_{i1}, \ldots, \eta_{ip})' = E(z_i | \xi_i, X_i),
\]

\[
X_i = (x_{i1}, \ldots, x_{ip})' \text{ is a } p \times q \text{ matrix of covariates,}
\]

\[
\beta = (\beta_1, \ldots, \beta_q)' \text{ are unknown regression coefficients,}
\]

\[
\xi_i = (\xi_{i1}, \ldots, \xi_{ir})' \text{ is a vector of independent gamma distributed latent variables}
\]

with \( \xi_{ik} \sim \mathcal{G}(\phi_k^{-1}, \phi_k^{-1}) \), for \( k = 1, \ldots, r \), and \( \lambda_j = (\lambda_{j1}, \ldots, \lambda_{jr})' \) are factor loadings for the \( j \)th outcome, \( j = 1, \ldots, p \).

• By linking the underlying Poisson variables \( z_i \) to the observed outcomes \( y_i \), we can accommodate arbitrary mixtures of discrete outcomes.
• The frailty multiplier, $\xi'_i \lambda_j$, is structured in a general manner to accommodate dependency in the multiple outcomes as well as over-dispersion relative to the Poisson distribution for count outcomes.

• The frailty is defined as a weighted sum of gamma latent variables.

• Under expression (4), the marginal expectation and variance of $z_{ij}$, integrating out the latent gamma variables are

$$E(z_{ij} \mid x_{ij}) = \left( \sum_{k=1}^{r} \lambda_{jk} \right) \exp(x'_{ij} \beta) = \exp(\nu_j + x'_{ij} \beta),$$

$$V(z_{ij} \mid x_{ij}) = E(z_{ij} \mid x_{ij}) + \left( \sum_{k=1}^{r} \phi_k \lambda_{jk}^2 \right) \{ \exp(x'_{ij} \beta) \}^2, \quad (5)$$

where $\nu_j = \log(\sum_k \lambda_{jk})$. 
Derivation of Marginal Variance and Correlation Coefficient

First note that the random variable, $z_{ij}$, can be expressed as a sum of independent Poisson random variables: $z_{ij} = \sum_{k=1}^{r} s_{ijk}$, where

$$s_{ijk} \sim \text{Poisson}(\lambda_{jk} \xi_{ik} \exp(x_i' \beta)),$$

for $k = 1, \ldots, r$.

Therefore, $V(z_{ij} | x_{ij}) = \sum_{k=1}^{r} V(s_{ijk} | x_{ij})$, where the marginal variance of $s_{ijk}$ is

$$V(s_{ijk} | x_{ij}) = E(s_{ijk}^2 | x_{ij}) - E(s_{ijk} | x_{ij})^2$$

Expression (5) follows directly after substituting back into the expression for $V(z_{ij} | x_{ij})$. 

$$V(s_{ijk} | x_{ij}) = E(E(s_{ijk}^2 | \xi_{ik}, x_{ij}) - E(E(s_{ijk} | \xi_{ik}, x_{ij})})^2$$

$$= E[\lambda_{jk} \xi_{ik} \exp(x_i' \beta) + \{\lambda_{jk} \xi_{ik} \exp(x_i' \beta)\}^2] - E\{\lambda_{jk} \xi_{ik} \exp(x_i' \beta)\}^2$$

$$= \lambda_{jk} \exp(x_i' \beta) + \lambda_{jk}^2 \exp(x_i' \beta)(1 + \phi_k) - \lambda_{jk}^2 \exp(x_i' \beta)^2$$

$$= \lambda_{jk} \exp(x_i' \beta) + \phi_k \lambda_{jk}^2 \exp(x_i' \beta)^2.$$
Parameter Interpretation and Properties

- For a count outcome, \( z_{ij} \) is observed directly and (5) provides a closed form for the marginal expectation and variance.

- It is clear that the subject-specific and population-averaged regression models have the same multiplicative form, with only the intercept varying.

- Hence, the regression coefficients, \( \beta_h \), have both conditional and marginal interpretations in terms of logarithms of ratios of expectations.
• The distribution of $z_{ij}$ is over-dispersed relative to the Poisson distribution when $\sum_{k=1}^r \phi_k \lambda_{jk}^2 > 0$.

• Hence, when $y_{ij}$ is a count, the $\lambda_j$ parameters measure not only the correlation between $y_{ij}$ and the other outcomes but also over-dispersion in the marginal distribution of $y_{ij}$ relative to the Poisson distribution.

• To separate these two attributes, one can include a latent variable specific to each count outcome, along with latent variables that load on more than one outcome and hence accommodate correlation.

• For example, if $y_{ij}$ is a count and $\xi_{ik}$ is included to allow over-dispersion, then we would restrict $\lambda_{j'k} = 0$ for all $j' \neq j$. This strategy and issues in model identifiability are discussed in detail in Section 3.1.
For binary outcomes, \( y_{ij} = I(z_{ij} > 0) \), a closed form can be derived for the marginal probability of a response (e.g., one or more tumors), integrating out the underlying Poisson and latent gamma variables:

\[
\Pr(y_{ij} = 1 \mid x_{ij}) = \Pr(z_{ij} > 0 \mid x_{ij}) = 1 - \Pr(s_{ij1} = 0, \ldots, s_{ijr} = 0 \mid x_{ij})
\]

\[
= 1 - \int \prod_{k=1}^{r} \Pr(s_{ijk} = 0 \mid \xi_{ik}, x_{ij}) G(\xi_{ik}; \phi_{k}^{-1}, \phi_{k}^{-1}) d\xi_{ik}
\]

\[
= 1 - \prod_{k=1}^{r} \int_{0}^{\infty} \exp \left( - \lambda_{jk} \xi_{ik} e^{x'_{ij} \beta} \right) \left( \frac{\phi_{k}^{-1}}{\Gamma(\phi_{k}^{-1})} \right) \frac{\phi_{k}^{-1}}{\phi_{k}^{-1}} d\xi_{ik}
\]

\[
= 1 - \prod_{k=1}^{r} \left( \frac{1}{1 + \phi_{k} \lambda_{jk} \exp(x'_{ij} \beta)} \right)^{\phi_{k}^{-1}},
\]

which simplifies to a logistic regression model when \( r = 1 \) and \( \phi_{1} = 1 \).
A useful closed form expression for characterizing the dependency between multiple binary outcomes is
\[
\frac{\Pr(y_{ij} = 0 \mid y_{ij'} = 0, X_i)}{\Pr(y_{ij} = 0 \mid X_i)} = \int \frac{\Pr(y_{ij} = 0 \mid \xi_i, X_i) \Pr(y_{ij'} = 0 \mid \xi_i, X_i) \pi(d\xi_i)}{\Pr(y_{ij} = 0 \mid X_i) \Pr(y_{ij'} = 0 \mid X_i)}
\]
\[
= \prod_{k=1}^{r} \int \exp \left\{-\xi_{ik}(\lambda_{jk}e^{x_{ij}'\beta} + \lambda_{jk'}e^{x_{ij}'\beta})\right\} G(\xi_{ik}; \phi_k^{-1}, \phi_k^{-1}) d\xi_{ik}
\]
\[
= \prod_{k=1}^{r} \left\{ (1 + \phi_k \lambda_{jk}e^{x_{ij}'\beta})(1 + \phi_k \lambda_{jk'}e^{x_{ij}'\beta}) \right\}^{-\phi_k^{-1}}
\]
\[
= \prod_{k=1}^{r} \left( \frac{(1 + \phi_k \lambda_{jk}e^{x_{ij}'\beta})(1 + \phi_k \lambda_{jk'}e^{x_{ij}'\beta})}{1 + \phi_k (\lambda_{jk}e^{x_{ij}'\beta} + \lambda_{jk'}e^{x_{ij}'\beta})} \right)^{\phi_k^{-1}}. \tag{7}
\]
This expression is interpretable as the multiplicative increase in the probability of \( y_{ij} = 0 \) given \( y_{ij'} = 0 \).
Prior Specification

• A Bayesian specification of the model is completed with prior distributions for the factor loadings, $\mathbf{\lambda} = (\lambda_1', \ldots, \lambda_p')$, the frailty variances, $\mathbf{\phi} = (\phi_1, \ldots, \phi_r)'$, and the regression parameters, $\mathbf{\beta}$.

• Allow $\pi(\mathbf{\beta})$ to follow an arbitrary form (e.g., multivariate normal).

• For the frailty precision parameters, $\phi_k^{-1}$, we choose gamma $\mathcal{G}(a_k, b_k)$ priors following common practice (Clayton, 1991).

• For the factor loadings parameters, $\lambda_{jk}$, we also choose gamma priors, $\pi(\lambda_{jk}) = \mathcal{G}(\lambda_{jk}; c_{jk}, d_{jk})$. However, for purposes of identifiability and interpretability, some of the elements of $\mathbf{\lambda}$ must be constrained.
Data Augmentation MCMC Algorithm

We express $z_{ij}$ as the following sum of independent Poisson latent variables:

$$z_{ij} = \sum_{k=1}^{r} s_{ijk}, \quad s_{ijk} \sim \text{Poisson}(\lambda_{jk} \xi_{ik} \exp(x_{ij}' \beta)), k = 1, \ldots, r.$$  \hspace{1cm} (8)

Following standard algebraic routes, it is straightforward to show that the resulting full conditional posterior distributions for $s_{ijk}$, $\xi_{ik}$, and $\lambda_{jk}$ follow simple conjugate forms. After specifying initial values, our algorithm alternates between:

1. Imputing the underlying Poisson variables $s_{ij} = (s_{ij1}, \ldots, s_{ijr})'$ by sampling from their full conditional posterior distribution (for all $i, j$).

2. Updating the latent variables, $\xi_i$, by sampling one at a time from their gamma full conditional distributions.

3. Updating the unknown factor loadings, $\lambda_{jk}$, by sampling one at a time from their gamma full conditional distributions.

4. Updating the frailty variances, $\phi$, and regression parameters, $\beta$. 


Example

• Let $T_i$ denote the number of weeks at which mouse $i$ is examined starting at week 9 and going up to the censoring time.

• Let $\delta_i = 1$ if the mouse developed papillomas by $T_i$ and $\delta_i = 0$ otherwise.

• Let $t_i \in \{1, 2, \ldots, T_i\}$ denote the minimum of $T_i$ and the week of first papilloma occurrence. The data for the three aspects of the tumor response are

  1. binary indicators $(y_{ij} = I(j=t_i)\delta_i$ for $j \leq t_i)$ of papilloma onset;

  2. repeated counts $(y_{i,t_i+1}, \ldots, y_{i,T_i+1})$ of weekly increases in papilloma burden from onset of the first tumor

  3. a binary indicator $(y_{i,T_i+2})$ of the presence of malignant tumors, respectively.