Bayesian Analyses of Multivariate Data

- We have discussed random effects models for repeated observations on a subject.

- These models allow the intercept and regression coefficients in a linear or generalized linear model to vary for the different study subjects in order to accommodate dependency.

- In considering such models, we have focused on cases where the same type of outcome (e.g., blood pressure) is measured repeatedly.

- In such cases, it often makes senses to think of the observations within an individual as independent draws from some distribution, conditional on measured and unmeasured factors for that individual.
Multiple Outcomes of Different Types

- In many cases, the outcome data \( \mathbf{y}_i = (y_{i1}, \ldots, y_{ip})' \) may consist of different but correlated observations for an individual.

- For example, the first outcome may be blood pressure, the second may be body mass index, and the third may be cholesterol.

- Focusing initially on the case where all outcomes are normally distributed, \( \mathbf{y}_i \sim N_p(\mathbf{X}_i \boldsymbol{\beta}, \Sigma) \), where \( \mathbf{X}_i = (x_{i1}, \ldots, x_{ip})' \), \( x_{ij} \) is a vector of predictors for the \( j \)th outcome for \( j = 1, \ldots, p \), \( \boldsymbol{\beta} \) are regression coefficients, and \( \Sigma \) is a \( p \times p \) covariance matrix.

- In this case, it typically does not make sense to assume that the same individual-specific regression coefficients apply to each of the disparate outcomes.

- However, we can model \( \mathbf{y}_i \) as multivariate normal without structuring the covariance matrix or using a hierarchical model.
Standard Bayesian Analyses of Multivariate Normal Outcomes

- Methods for Bayesian inferences in univariate normal linear models generalize directly to the multivariate normal case.

- We can still place conjugate normal priors on the regression coefficients.

- The main difference is in the prior specification for the covariance matrix, since we now have a matrix instead of a single error precision.

- The typical default conjugate choice of prior for $\Sigma$ is the inverse-Wishart density.
Limitations of Wishart priors for $\Sigma^{-1}$

- The Wishart density tends to be quite restrictive in that it prescribes a common degrees of freedom for each of the diagonal elements of $\Sigma$.

- Potentially, one may have abundant prior information about certain entries in $\Sigma$, but limited prior information about other entries.

- In addition, one may want to do inferences on the covariance structure. For example, is the 1st outcome associated with the 3rd outcome, conditionally on the other variables?

- The Wishart prior is not flexible enough to allow for differential prior knowledge about the different elements of $\Sigma$.

- In addition, the Wishart prior does not allow for zero elements.
Modeling of the Covariance Structure

• Because the number of unknowns in $\Sigma$ $(p(p + 1)/2)$ can be large (or huge in gene expression studies), one often wants to reduce the number of free parameters.

• Many possibilities have been proposed in this regard. To list a few:

  – Define a prior distribution that allows off diagonal elements to be zero, while penalizing models with lots of non-zero elements (to favor sparse covariance matrices).
    
    * A challenging problem is to define prior distributions for $\Sigma$ - possibly after decomposition - so that $\Sigma$ is symmetric and positive definite.
    
    * In addition, posterior computation under the necessary restrictions can be very challenging - particularly in large $p$ cases.
    
    * This has been the motivation of much of the recent work on Bayesian graphical models.
– Model the covariance structure using a factor analytic model that incorporates shared latent variables:

\[ y_i = X_i \beta + \Lambda \xi_i + \epsilon_i, \]

where \( X_i \) is a \( p \times q \) matrix of known predictors, \( \beta \) is a \( q \times 1 \) vector of regression coefficients, \( \Lambda \) is a \( p \times r \) factor loadings matrix, \( \xi_i = (\xi_{i1}, \ldots, \xi_{ir})' \) is a \( r \times 1 \) vector of independent standard normal latent variables, and \( \epsilon_i \sim N_p(0, \Sigma) \) is an error residual with diagonal covariance \( \Sigma \).
Factor Analysis and Structural Equations Modeling

- Factor analysis is most useful when there is a priori knowledge about unmeasured variables, which may lead to within-subject dependency.

- For example, a study may be designed specifically to measure some latent trait, such as intelligence, stress level, or overall neurological functioning.

- Outcomes may then consist of different items or tests measuring the latent trait or latent traits.

- In such cases, the structure of the factor loadings matrix can be chosen not only to satisfy minimal identifiability constraints but also to fix the definition of each latent trait.

- Potentially, one can even define a hierarchical model in which certain predictors impact the measured outcomes (“manifest variables”) only through the intermediate latent variables.

- A graphical model can be drawn to illustrate the conditional independence relationships.
Induced Correlation Structure

It is straightforward to derive the correlation coefficient between any two outcomes, $y_{ij}$ and $y_{ij'}$, integrating out the latent factors:

$$
\rho(y_{ij}, y_{ij'}) = \frac{E(y_{ij}y_{ij'}) - E(y_{ij})E(y_{ij'})}{\sqrt{V(y_{ij})V(y_{ij'})}}
$$

$$
= \frac{E\{(\xi_i'\lambda_j + \epsilon_{ij})(\xi_i'\lambda_{j'} + \epsilon_{ij'})\}}{\sigma_j\sigma_{j-1}}
$$

$$
= \frac{\sum_{k=1}^r \lambda_j \lambda_{j'} k E(\xi_{ik}^2)}{\sigma_j\sigma_{j-1}} = \frac{\lambda_j' \lambda_{j'}}{(\sigma_j\sigma_{j-1})}.
$$

Hence, the $j$th and $j'$th outcomes are uncorrelated if and only if $\sum_{k=1}^r \lambda_{jk} \lambda_{j'k} = 0$ (i.e., there aren’t any latent traits that load on both outcomes).

Because the latent traits are assumed in advance to have $N(0,1)$ densities for identifiability and interpretability, the relative values of the factor loadings parameters and residual variances drive the correlation matrix.
Bayesian Inferences in Normal Factor Models

- Assuming the basic factor structure (number of latent traits, locations of structural zeros in the $\Lambda$ matrix) is known \textit{a priori}, we can take advantage of the conditionally normal linear structure to simplify Bayesian inferences.

- In particular, given the latent variables $\xi_i$, the model $y_i = X_i'\alpha + \Lambda \xi_i + \epsilon_i$ can be reexpressed as $y_i = W_i'\theta + \epsilon_i$, where $W_i$ is a $p \times s$ matrix with elements consisting of known predictors $X_i$ and latent variables $\xi_i$, and $\theta$ is a $s \times 1$ parameter vector containing $\alpha$ and the free elements of $\Lambda$. 
• Hence, the prior distribution $\pi(\theta) = N_s(\theta_0, \Sigma_\theta)$ is conditionally-conjugate and the conditional posterior distribution of $\theta$ is multivariate normal. It follows that the regression coefficients and factor loadings can be updated in a single block, making posterior computation much more efficient (always an issue in factor analyses).

• The conditional distributions of the latent variables $\xi_i$ given $\theta$ and $\Sigma$ are also normally distributed, while conjugate inverse-gamma priors can be chosen for the diagonal elements of $\Sigma$. Hence, posterior computation can proceed via a simple Gibbs sampling algorithm.
• It is straightforward to include point masses at zero in the prior for certain factor loadings parameters to allow some uncertainty in the loadings structure.

• In addition, one can potentially choose a prior for the number of factors and use a reversible jump MCMC algorithm for posterior computation, though it is important to note that the data often contain minimal information about the factor structure & many structures may result in very similar likelihoods.

• An important comment is that one should avoid improper priors or highly diffuse but proper priors in factor analyses, in the former case the posterior is often improper and in the latter case mixing will often be horrible.
What about non-Gaussian Outcomes?

• Now suppose that the elements of $y_i$ are binary or belong to some distribution in the exponential family - e.g., Poisson.

• Modifications to allow probit or underlying normal models are straightforward, and are the standard approach used to allow mixed discrete and continuous outcomes (Muthén, 1983, 1984; Dunson, 2000, Dunson, Chen, and Harry, 2003).

• In particular, we let $y_{ij} = g_j(z_{ij}; \tau)$, for $j = 1, \ldots, p$, where $g(\cdot)$ is a threshold link function and $z_i = (z_{i1}, \ldots, z_{ip})'$ is a vector of underlying normal variables (with unit variance when the corresponding $y$ is categorical) following a normal linear factor model.

• Posterior computation then proceeds again using the Albert and Chib (1993) data augmentation trick.
Moving beyond underlying normal models

- Clearly, the underlying normal structure is limited in its generality and there is often interest in considering other model structures.

- As an alternative Moustaki (1996), Sammel, Ryan and Legler (1997), and others have proposed defining a generalized linear model for each of the outcomes separately, incorporating shared latent traits to accommodate dependency.

- In particular, letting \( \eta_i = (\eta_{i1}, \ldots, \eta_{ip})' \) be a vector of linear predictors for the different outcomes, we let \( \eta_i = X_i \alpha + \Lambda \xi_i \).

- Posterior computation can then proceed by Gibbs sampling via adaptive rejection sampling.
Alternative and hierarchical structures

• We have been focusing on a very limited class of models in which known predictors $X_i$ effect the different outcomes directly and the latent variables do not depend on predictors.

• In many cases, $X_i$ may be a (possibly high dimensional) vector of “endogenous variables” measuring latent predictors, $\xi_i$, which we want to relate to the latent traits, $\eta_i$ (using the standard structural equations modeling (SEM) notation).

• To allow for this more general setting we can define a SEM, and the conditional linear normal simplications will again apply when the manifest and latent variables are all normally distributed (the typical assumption).