Hierarchical Models

Yuki Shiraito

POLSCI 798 Advanced Topics in Quantitative Methodology
University of Michigan

Winter 2019

Overdispersion

- Hierarchical models
  - Simple models as subcomponents of a larger model
  - Useful to account for correlation/dependence/group structure
  - Examples of problems to address
    1. Overdispersion in count data
    2. Dependence in TSCS/panel data

- Overdispersion common in count data
  - Count data: # of violent attacks, # of ..., etc.
  - $X_i \overset{\text{i.i.d.}}{\sim} \text{Bern}(p)$
    - $\widehat{p}_{\text{MLE}} = \operatorname{argmax}_p \sum_{i=1}^{n} \{X_i \log p + (1 - X_i) \log(1 - p)\} = \overline{X}_n \xrightarrow{\text{p}} p$
    - $\widehat{\mathbb{V}(X)}_{\text{MLE}} = \overline{X}_n(1 - \overline{X}_n) \xrightarrow{\text{p}} p(1 - p) = \mathbb{V}(X)$
  - $X_i \overset{\text{i.i.d.}}{\sim} \text{Pois}(\lambda)$
    - $\widehat{\lambda}_{\text{MLE}} = \operatorname{argmax}_\lambda \sum_{i=1}^{n} \{X_i \log \lambda - \lambda\} = \overline{X}_n \xrightarrow{\text{p}} \lambda$
    - $\widehat{\mathbb{V}(X)}_{\text{MLE}} = \overline{X}_n \xrightarrow{\text{p}} \lambda = \mathbb{V}(X)$
  - $\widehat{\sigma^2}_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \xrightarrow{\text{p}} \mathbb{V}(X)$ under no parametric assumptions
  - *Overdispersion* if $\widehat{\sigma^2}_n \gg \widehat{\mathbb{V}(X)}_{\text{MLE}}$

Poisson Regression Mixture Models

- Poisson regression model
$$Y_i \mid \mathrm{X}_i, \beta \overset{\text{indep.}}{\sim} \text{Pois}\left(e^{\mathrm{X}_i^\top \beta}\right) \Leftrightarrow p\left(Y_i \mid \mathrm{X}_i, \beta\right) = \frac{1}{Y_i!}\left(e^{\mathrm{X}_i^\top \beta}\right)^{Y_i} e^{-e^{\mathrm{X}_i^\top \beta}}$$
with some prior $p(\beta)$
  - $\mathbb{E}\left[Y_i \mid \mathrm{X}_i, \beta\right] = e^{\mathrm{X}_i^\top \beta},\ \mathbb{V}\left(Y_i \mid \mathrm{X}_i, \beta\right) = e^{\mathrm{X}_i^\top \beta}$
- Possible source of overdispersion: Unit heterogeneity
- Poisson mixture model
$$Y_i \mid \mathrm{X}_i, \beta, Z_i \overset{\text{indep.}}{\sim} \text{Pois}\left(Z_i e^{\mathrm{X}_i^\top \beta}\right)$$
$$Z_i \overset{\text{i.i.d.}}{\sim} \text{Gamma}(a, a)$$
with some prior $p(a, \beta)$
  - "Random intercept": $Z_i e^{\mathrm{X}_i^\top \beta} = e^{\mathrm{X}_i^\top \beta + \varepsilon_i}$ where $Z_i = e^{\varepsilon_i}$
  - Z_i is latent, and the Gamma is the mixing distribution
  - Hierarchical (multilevel) structure
  - Likelihood:
$$p(Y_i \mid \mathrm{X}_i, \beta, a) = \int_0^\infty p(Y_i \mid \mathrm{X}_i, \beta, Z_i = z) p(z \mid a) dz$$

The Gamma Distribution

- Density of the Gamma distribution (shape-rate):
$$p(z) = \frac{\eta^\zeta}{\Gamma(\zeta)} z^{\eta-1} e^{-\eta z} \quad \text{for } 0 < z < \infty$$
where $\zeta > 0$ (shape) and $\eta > 0$ (rate, a.k.a. inverse scale)
- Alternative shape-scale parameterization:
$$p(z) = \frac{1}{\Gamma(k)\theta^k} z^{k-1} e^{-\frac{z}{\theta}} \quad \text{for } 0 < z < \infty$$
where $k > 0$ (shape) and $\theta > 0$ (scale)
- Normalizing constant:
$$\int_0^\infty p(z)dz = \frac{\eta^\zeta}{\Gamma(\zeta)} \int_0^\infty z^{\eta-1} e^{-\eta z} dz = 1 \Leftrightarrow \int_0^\infty z^{\eta-1} e^{-\eta z} dz = \frac{\Gamma(\zeta)}{\eta^\zeta}$$
- Gamma function:
$$\Gamma(\eta) \equiv \int_0^\infty x^{\eta-1} \exp(-x) dx$$
- $\Gamma(\eta + 1) = \eta\Gamma(\eta)$ for $\eta > 0$ and $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$

Negative-Binomial as an Overdispersed Poisson

- Likelihood of the Poisson mixture model:

$$\begin{aligned} p(Y_i \mid \mathrm{X}_i, \beta, a) &= \int_0^\infty \frac{1}{Y_i!}\left(z e^{\mathrm{X}_i^\top \beta}\right)^{Y_i} e^{-z e^{\mathrm{X}_i^\top \beta}} \frac{a^a}{\Gamma(a)} z^{a-1} e^{-az} dz \\ &= \frac{\left(e^{\mathrm{X}_i^\top \beta}\right)^{Y_i} a^a \Gamma(Y_i + a)}{Y_i! \Gamma(a) \left(e^{\mathrm{X}_i^\top \beta} + a\right)^{Y_i + a}} \\ &= \frac{(Y_i + a - 1)!}{Y_i!(a-1)!} \left(\frac{e^{\mathrm{X}_i^\top \beta}}{e^{\mathrm{X}_i^\top \beta} + a}\right)^{Y_i} \left(\frac{a}{e^{\mathrm{X}_i^\top \beta} + a}\right)^a \end{aligned}$$

  - Negative binomial regression model
  - Dispersion parameter a (some packages estimate $1/a$)
  - Mean $= e^{\mathrm{X}_i^\top \beta}$ and Variance $= e^{\mathrm{X}_i^\top \beta} + \frac{\left(e^{\mathrm{X}_i^\top \beta}\right)^2}{a}$ (overdispersed!)
- MCMC algorithm for Bayesian inference
  - Alternating draws of Z_i, a, and β
  - Not conditionally conjugate $\rightsquigarrow$ Metropolis-Hastings (M-H) draws

Zero-Inflated Poisson Models

- Another source of overdispersion: Zero inflation
- Zero-inflated Poisson regression model:

$$\begin{cases} Y_i \mid Z_i = 0 \overset{\text{i.i.d.}}{\sim} \mathbb{P}\left(Y_i = 0\right) = 1, \\ Y_i \mid \mathrm{X}_i, \beta, Z_i = 1 \overset{\text{indep.}}{\sim} \text{Pois}\left(e^{\mathrm{X}_i^\top \beta}\right) \end{cases}$$

$$Z_i \overset{\text{i.i.d.}}{\sim} \text{Bern}\left(\tau\right)$$

  - Example of finite mixture models: DGP depends on indicator Z_i
  - Overdispersion (underdispersion) if $\tau + \tau(1-\tau)e^{\mathrm{X}_i^\top \beta} > (<) 1$
  - Conditionally conjugate prior on τ: Beta distribution
- MCMC algorithm:
  1. Conditional draw of Z_i:

$$\begin{cases} \mathbb{P}\left(Z_i^{(s)} = 1 \,\middle|\, Y_i \geq 1\right) = 1 \\ \mathbb{P}\left(Z_i^{(s)} = 1 \,\middle|\, Y_i = 0, \beta^{(s-1)}, \tau^{(s-1)}\right) = \dfrac{\tau^{(s-1)} e^{-\exp\left(\mathrm{x}_i^\top \beta^{(s-1)}\right)}}{\left(1-\tau^{(s-1)}\right) + \tau^{(s-1)} e^{-\exp\left(\mathrm{x}_i^\top \beta^{(s-1)}\right)}} \end{cases}$$

  2. Gibbs draw of τ: Beta-Bernoulli with $Z_i^{(s)}, i = 1, \ldots, n$ as data
  3. M-H draw of β: Poisson regression only using i s.t. $Z_i^{(s)} = 1$

Accounting for Data Structure

- Data are often structured
  1. Cluster randomized experiments. E.g. Randomized across villages
  2. Repeated measures for the same units. E.g. fMRI
  3. Nested groups. E.g. Students $\subset$ classes $\subset$ schools
  4. Panel/TSCS data. E.g. Country-year, State-year, etc...

- Cluster randomized experiments
  - Possible interference between units in the same group
  - No interference across groups
  - Potential outcomes given a group-level treatment: $Y_i(T_j = t)$
  - Linear model: $Y_i(t) = \mu_0 + \tau t + \varepsilon_i$
  - $\mathrm{Cov}(\varepsilon_i, \varepsilon_{i'}) = \rho_\varepsilon$ if i and i' are in the same group
  - Cluster robust standard errors:

$$\mathbb{V}(\widehat{(\hat{\mu}_0, \hat{\tau})} \mid T) = \left(\sum_{j=1}^{m} \mathbf{X}_j^\top \mathbf{X}_j\right)^{-1} \left(\sum_{j=1}^{m} \mathbf{X}_j^\top \hat{\varepsilon}_j \hat{\varepsilon}_j^\top \mathbf{X}_j\right) \left(\sum_{j=1}^{m} \mathbf{X}_j^\top \mathbf{X}_j\right)^{-1}$$

  - $\hat{\mu}_0, \hat{\tau}$: OLS estimators
  - $\mathbf{X}_j = [1 \, T_j]$: Design matrix for cluster j
  - $\hat{\varepsilon}_j$: Residuals for cluster j

Least Squares Estimators for Grouped Data

- Unobserved effects model for grouped data
$$Y_i = X_i^\top \beta + \alpha_{j[i]} + \varepsilon_i$$
    - $j[i]$: Group j to which unit i belongs
    - $\alpha_{j[i]}$: Intercept specific to group $j[i]$
- Random effects assumption: For all i and j
$$\mathbb{E}[\varepsilon_i \mid \{X_{i'} : i' \in j\}, \alpha_j] = 0 \text{ and } \mathbb{E}[\alpha_j \mid \{X_{i'} : i' \in j\}] = 0$$
- Feasible generalized least squares (FGLS) estimator:
$$\hat{\beta}_{\text{RE}} = \left(\sum_{j=1}^{m} \mathbf{X}_j^\top \widehat{\Omega}^{-1} \mathbf{X}_j \right)^{-1} \left(\sum_{j=1}^{m} \mathbf{X}_j^\top \widehat{\Omega}^{-1} Y_j \right)$$
- Fixed effects assumption: For all i and j
$$\mathbb{E}[\varepsilon_i \mid \{X_i : i \in j\}, \alpha_j] = 0$$
- Within estimator:
$$\hat{\beta}_{\text{FE}} = \left(\sum_{j=1}^{m} \ddot{\mathbf{X}}_j^\top \ddot{\mathbf{X}}_j \right)^{-1} \left(\sum_{j=1}^{m} \ddot{\mathbf{X}}_j^\top \ddot{Y}_j \right), \quad \ddot{\mathbf{X}}_j \equiv \mathbf{X}_j - \overline{\mathbf{X}}_j$$

Problems with Common Practice

- None of the above directly carries to MLE
- Incidental parameter problem
  - MLE is inconsistent if # of parameters grows
  - Canonical case: Panel data with $i = 1, \ldots, N$ and $t = 0, 1$

$$p\left(Y_{it} = 1 \mid a_i, \beta\right) = \frac{e^{a_i + t\beta}}{1 + e^{a_i + t\beta}} \quad \Rightarrow \quad \hat{\beta}_{\text{MLE}} \overset{\text{p}}{\to} 2\beta \ \text{ as } \ N \to \infty$$

  - Problematic under any model: Logit, probit, Poisson, etc.
- Robust standard error with model misspecification
  - MLE assuming independence: $\hat{\theta}_{\text{MLE}} = \operatorname{argmax} \prod_i \mathcal{L}(Y_i \mid X_i, \theta)$
  - Clustering $\Leftrightarrow$ dependence: Wrong likelihood $\rightsquigarrow$ wrong eatimates!

$$\hat{\theta}_{\text{MLE}} \overset{\text{p}}{\to} \theta_0 \equiv \operatorname*{argmin}_{\theta \in \Theta} \underbrace{\int \log \frac{p(Y_i)}{\mathcal{L}(Y_i \mid \theta)} p(Y_i)\, dY_i}_{\textit{Kullback-Leibler divergence}}$$

  - Asymptotic distribution *around wrong point estimates*:

$$\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta_0) \overset{\text{d}}{\to} \mathcal{N}\left(0,\ \mathbb{E}(-\mathbf{H}_i(\theta_0))^{-1} \mathbb{E}(s_i(\theta_0) s_i(\theta_0)^\top) \mathbb{E}(-\mathbf{H}_i(\theta_0))^{-1}\right)$$

  - "Correct" standard errors for "wrong" estimates

Shrinkage in Multilevel Models

- Multilevel regression models
  - Data structure incorporated into model
  - Shrinkage toward upper levels
- Simple Gaussian hierarchical model
  - Groups: $j = 1, \ldots, J$
  - Individuals: $i = 1, \ldots, n_j$ in group j
  - Model:

$$\begin{array}{ll} \text{Individual level} & Y_{ij} \overset{\text{indep.}}{\sim} \mathcal{N}\left(v_j, \sigma^2\right) \\ \text{Group level} & v_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu, \omega^2\right) \end{array}$$

  - Prior on a hyperparameter: $p(\mu) \propto 1$
  - Variance parameters assumed to be known
- Posterior distribution of $v_j, j = 1, \ldots, J$, given μ:

$$v_j \,\middle|\, \mathbf{Y}, \mu \overset{\text{indep.}}{\sim} \mathcal{N}\left(\frac{\frac{1}{\sigma^2/n_j}\overline{Y}_{\cdot j} + \frac{1}{\omega^2}\mu}{\frac{1}{\sigma^2/n_j} + \frac{1}{\omega^2}}, \frac{1}{\frac{1}{\sigma^2/n_j} + \frac{1}{\omega^2}} \right)$$

- Group-level model as prior on individual-level parameters
- *Shrinkage to the grand mean*

Complete, Partial, and No Pooling

- Posterior inference on hyperparameter
  - Marginal posterior density of μ:

$$p(\mu \mid \mathrm{Y}) \propto \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\mu) \prod_{j=1}^{J} p(v_j \mid \mu) \prod_{i=1}^{n_j} p(Y_{ij} \mid v_j)\, dv_1 \ldots dv_J$$

  - Reduced form: $Y_{ij} = \mu + \zeta_j + \varepsilon_{ij}, \quad \zeta_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \omega^2), \; \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$
  - Marginal posterior distribution of μ:

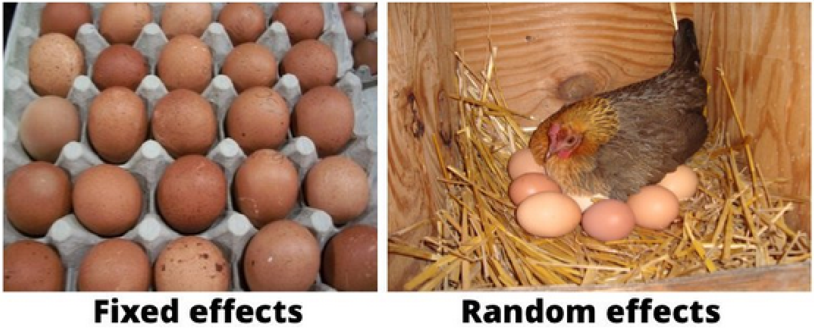
$$\mu \mid \mathrm{Y} \sim \mathcal{N}\left(\frac{\sum_{j=1}^{J} \frac{1}{\sigma^2/n_j + \omega^2} \overline{Y}_{\cdot j}}{\sum_{j=1}^{J} \frac{1}{\sigma^2/n_j + \omega^2}}, \; \frac{1}{\sum_{j=1}^{J} \frac{1}{\sigma^2/n_j + \omega^2}} \right)$$

- Partial pooling ("random effects"): $\omega \in (0, \infty)$
  - Heterogeneous but related groups
  - Other groups partially used through μ to estimate v_j
- No pooling ("fixed effects"): $\omega \to \infty$
  - Independent groups: $p(v_1, \ldots, v_J) = \prod_{j=1}^{J} p(v_j \mid \mu) \propto 1$
  - No information in other groups used to estimate v_j: $\mathbb{E}[v_j \mid \mathrm{Y}] = \overline{Y}_{\cdot j}$
- Complete pooling: $\omega \to 0$
  - Homogeneous groups: $v_1 = \cdots = v_J = \mu$
  - All observations equally used to estimate μ: $\mathbb{E}[\mu \mid \mathrm{Y}] = \overline{Y}_{\cdot\cdot}$

Will Lowe
@conjugateprior

Following

When modeling grouped data, don't forget the chicken.



9:00 AM - 12 May 2016

Multilevel Linear Models

- Linear mixed-effects models
  - General form:
    $$Y_i = \mathrm{X}_i^\top \beta + \mathrm{Z}_i^\top \lambda_i + \varepsilon_i,$$
  - Z_i: Typically a subset of X_i
  - Fixed effects β: Parameters shared by all observations
  - Random effects λ_i: Partially pooled parameters
- Varying-intercepts models
  - Correlated random effects (CRE)
    $$Y_i = \mathrm{X}_i^\top \beta + a_{j(i)} + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$
    $$a_j = \overline{\mathrm{X}}_j^\top \gamma + \zeta_j \quad \zeta_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \omega^2\right)$$
    - MLE of β in the CRE model = Within estimator $\hat{\beta}_{FE}$
  - Non-nested random effects
    $$Y_i = \mathrm{X}_i^\top \beta + a_{j(i)} + \delta_{t(i)} + \varepsilon_i, \quad \varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$
    $$a_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(a, \omega^2\right) \quad \delta_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\delta, \psi^2\right)$$
    - Identified up to a constant: $a_{j(i)} + \delta_{t(i)} = (a_{j(i)} - c) + (\delta_{t(i)} + c)$
- Varying-coefficients models

Gibbs Sampler for Multilevel Linear Models

- MCMC and multilevel models go together very well
  - Upper level models function as priors on lower level parameters
  - Lower level parameters function ad data for upper level models
  - ⇝ Full conditionals reduce to posteriors of component models
- Example: Varying-intercepts model

$$Y_i = X_i^\top \beta + a_{j(i)} + \varepsilon_i, \quad \varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$
$$a_j = Z_j^\top \delta + \zeta_j, \quad \zeta_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \omega^2\right)$$

- Joint posterior:

$$p\left(\beta, \delta, \sigma, \omega, \{a_j\}_{j=1}^{J} \,\middle|\, \mathbf{X}, \mathbf{Z}, Y\right)$$
$$\propto \underbrace{p\left(\beta, \delta, \sigma, \omega\right)}_{\text{prior}} \underbrace{\prod_{j=1}^{J} p\left(a_j \,\middle|\, \mathbf{Z}, \delta, \omega\right)}_{\text{group}} \underbrace{\prod_{i=1}^{N} p\left(Y_i \,\middle|\, \mathbf{X}, \beta, \sigma, a_{j(i)}\right)}_{\text{individual}}$$
$$\propto p\left(\beta, \delta, \sigma, \omega\right) \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}\omega} e^{-\frac{\left(a_j - Z_j^\top \delta\right)^2}{2\omega^2}} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(Y_i - X_i^\top \beta - a_j(i)\right)^2}{2\sigma^2}}$$

- Sample β and σ: Given δ, ω, and $a_j, j = 1, \ldots, J$:

$$p\left(\beta, \sigma \,\middle|\, \delta, \omega, \{a_j\}_{j=1}^{J}, \mathbf{X}, \mathbf{Z}, \mathrm{Y}\right)$$
$$\propto p(\beta, \sigma \mid \delta, \omega) \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{\left\{\left(Y_i - a_{j(i)}\right) - X_i^\top \beta\right\}^2}{2\sigma^2}}$$

  - Reduced form: $Y_i - a_{j(i)} = X_i^\top \beta + \varepsilon_i$
  - Gibbs step 1: Bayesian regression of $Y_i - a_{j(i)}^{(s)}$ on X_i
- Sample $a_j, j = 1, \ldots, J$: Given δ, ω, β, and σ

$$p\left(\{a_j\}_{j=1}^{J} \,\middle|\, \beta, \sigma, \delta, \omega, \mathbf{X}, \mathbf{Z}, \mathrm{Y}\right)$$
$$\propto \prod_{j=1}^{J} e^{-\frac{\left(a_j - Z_j^\top \delta\right)^2}{2\omega^2}} \prod_{i=1}^{N} e^{-\frac{\left\{\left(Y_i - X_i^\top \beta\right) - a_{j(i)}\right\}^2}{2\sigma^2}}$$

  - Reduced form: $Y_i - X_i^\top \beta = a_{j(i)} + \varepsilon_i$
  - Gibbs step 2:
    - Bayesian regression of $Y_i - X_i^\top \beta^{(s)}$ on the intercept for each group
    - Error variance is known: ${\sigma^{(s)}}^2$
    - Prior on a_j: $\mathcal{N}\left(Z_j^\top \delta^{(s)}, {\omega^{(s)}}^2\right)$
  - Group-level model as prior on a_j

- Sample δ and ω: Given β, σ, and $a_j, j = 1, \ldots, J$:
$$p\left(\delta, \omega \,\middle|\, \beta, \sigma, \{a_j\}_{j=1}^J, \mathbf{X}, \mathbf{Z}, Y\right)$$
$$\propto p(\delta, \omega \mid \beta, \sigma) \prod_{j=1}^J \frac{1}{\sqrt{2\pi}\omega} e^{-\frac{\left(a_j - Z_j^\top \delta\right)^2}{2\omega^2}}$$
  - Gibbs step 3: Bayesian regression of $a_j^{(s)}$ on Z_j
  - a_j is "data" when sampling hyperparameters δ and ω
- Alternative approach
  - Sampling β and δ in one step
    - Reduced form: $Y_i = X_i^\top \beta + Z_{j(i)}^\top \delta + \zeta_{j(i)} + \varepsilon_i$
    - Bayesian regression of Y_i on $[X_i^\top \, Z_{j(i)}^\top]^\top$
    - Error variance: Block diagonal matrix where the block for group j is
$$\Omega_j = \begin{pmatrix} \omega^2 + \sigma^2 & \omega^2 & \cdots & \omega^2 \\ \omega^2 & \omega^2 + \sigma^2 & \vdots & \omega^2 \\ \vdots & \cdots & \ddots & \vdots \\ \omega^2 & \cdots & \cdots & \omega^2 + \sigma^2 \end{pmatrix}$$
    - Y_i's are correlated within groups through ζ_j
  - $\{a_j\}_{j=1}^J, \omega, \sigma$: Identical to the simple Gaussian hierarchical model

Multilevel Probit Regression

- Multilevel model nested in the latent variable representation:

$$Y_i = \begin{cases} 0 & (U_i \leq 0) \\ 1 & (U_i > 0) \end{cases}$$

$$U_i = \mathrm{X}_i^\top \beta + a_{j(i)} + \varepsilon_i, \; \varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$$

$$a_j = Z_j^\top \delta + \zeta_j, \qquad \zeta_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \omega^2)$$

- Gibbs sampler: Simply adding one step for U_i!
  - $Y_i = 0$: $U_i^{(s)}$ drawn from $\mathcal{TN}_{(-\infty,0]}\left(\mathrm{X}_i^\top \beta^{(s-1)} + a_{j(i)}^{(s-1)}, 1\right)$
  - $Y_i = 1$: $U_i^{(s)}$ drawn from $\mathcal{TN}_{(0,\infty)}\left(\mathrm{X}_i^\top \beta^{(s-1)} + a_{j(i)}^{(s-1)}, 1\right)$
- Or, `rstanarm` provides a generic implementation
- Prior
  - Proper, conditionally conjugate prior:
    - Gaussian for coefficients
    - Scaled inverse-χ^2 for ω^2
  - $\mathbb{V}(\varepsilon_i) = 1$ for identification
  - β, δ, and ω not too large; otherwise, complete separation

Generalized Linear Models (GLM)

- Generalized linear models (McCullagh and Nelder 1989)
  1. Response variable: $Y_i \overset{\text{i.i.d.}}{\sim} p\left(Y_i \mid \eta_i, \varphi\right)$
  2. Linear predictor: $\eta_i = \mathrm{X}_i^\top \beta$
  3. Link function: $\mathbb{E}[Y_i] = g^{-1}(\eta_i)$
  4. Dispersion parameter (optional): φ
- Examples:
  1. Gaussian
     1. Gaussian response: $Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\eta_i, \varphi\right)$
     2. Identity link: $\mathbb{E}[Y_i] = \mathrm{X}_i^\top \beta$
     3. Dispersion: $\varphi = \sigma^2$
  2. Poisson
     1. Poisson response: $Y_i \overset{\text{i.i.d.}}{\sim} \text{Pois}\left(g^{-1}\left(\eta_i\right)\right)$
     2. Log link: $\mathbb{E}[Y_i] = e^{\eta_i}$
  3. Binomial
     1. Binomial response with known n_i: $Y_i \overset{\text{i.i.d.}}{\sim} \text{Binom}(g^{-1}\left(\eta_i\right), n_i)$
     2. Logit or probit link: $\mathbb{E}[Y_i/n_i] = e^{\eta_i}/(1 + e^{\eta_i})$ or $\mathbb{E}[Y_i/n_i] = \Phi(\eta_i)$
  4. Other distributions: t, Gamma, etc.

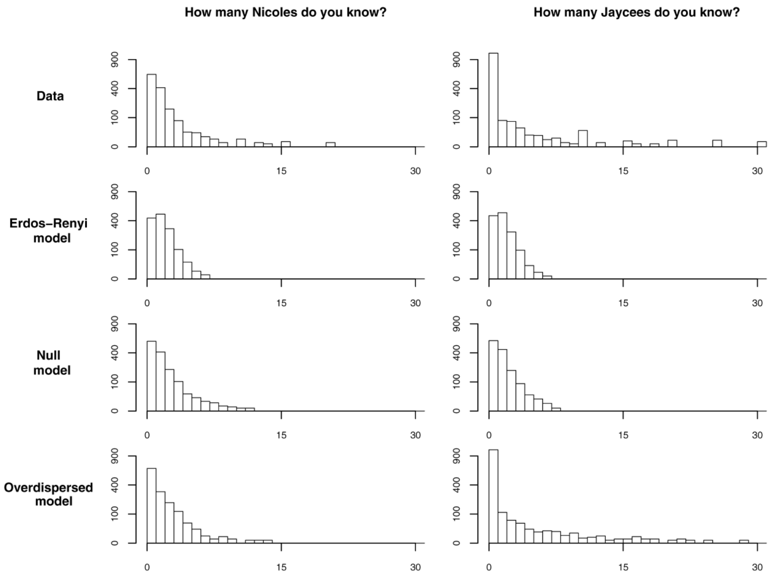
Generalized Linear Mixed Models (GLMM)

- Generalized Linear Mixed Models
  - Linear mixed-effects model for η_i: $\eta_i = X_i^\top \beta + Z_i^\top \lambda_i$
  - Overdispersion due to heterogeneity across groups
  - Usually not conditionally conjugate $\rightsquigarrow$ MH or `Stan` (`rstanarm`)
- Example: Multilevel Poisson regression for social network
  - Survey: Individuals $i = 1, \ldots, N$ and items $k = 1, \ldots, K$
  - Item k: *"How many [members of social group k] do you know?"*
  - Estimate the size of i's social network (# of acquintances)
  - Estimate the size of hard-to-count subpopulations (e.g., homeless)
  - Erdos-Renyi: Equal size of social network

$$Y_i \overset{\text{i.i.d.}}{\sim} \text{Pois}\left(e^{\alpha+\beta_k}\right)$$

  - Overdispersed model:
    - Response: $Y_{ik} \overset{\text{i.i.d.}}{\sim} \text{Pois}\left(e^{\alpha_i+\beta_k+\gamma_{ik}}\right)$
    - Popularity of i: $\alpha_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu_\alpha, \sigma_\alpha^2\right)$
    - Size of group k: $\beta_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu_\beta, \sigma_\beta^2\right)$
    - Overdispersion: $e^{\gamma_{ik}} \overset{\text{i.i.d.}}{\sim} \text{Gamma}\left(1/(\omega_k - 1), 1/(\omega_k - 1)\right)$
    - Uninformative uniform prior on $\mu_\alpha, \mu_\beta, \sigma_\alpha, \sigma_\beta$; $p(1/\omega_k) \propto 1$

How Many X's Do You Know?



Summary

- Overdispersion in count data
  - Modeling unit heterogeneity as continuous mixture
  - Modeling zero inflation as finite mixture
- Group structure in data
  - Problematic approach: Incidental parameters and model misspecification
  - Multilevel modeling: Partial pooling $\rightsquigarrow$ shrinkage to mean
  - MCMC steps for component models

- Readings for review
  1. Overdispersion and pooling:
     - **BDA3** Ch. 5
  2. Multilevel regression models:
     - **BDA3** Ch. 15–6
     - Gelman and Hill (2007) Ch. 11–6
  3. Multilevel model for social network analysis
     - Zheng et. al. (2006) "How Many People Do You Know in Prison?"