Hierarchical Poisson model

- Count data are often modeled using a Poisson model.

- If $y \sim \text{Poisson}(\mu)$ then $E(y) = \text{var}(y) = \mu$.

- When counts are assumed exchangeable given $\mu$ and the rates $\mu$ can also be assumed to be exchangeable, a Gamma population model for the rates is often chosen.

- The hierarchical model is then

\[
\begin{align*}
y_i & \sim \text{Poisson}(\mu_i) \\
\mu_i & \sim \text{Gamma}(\alpha, \beta).
\end{align*}
\]

- Priors for the hyperparameters are often taken to be Gamma (or exponential):

\[
\begin{align*}
\alpha & \sim \text{Gamma}(a, b) \\
\beta & \sim \text{Gamma}(c, d),
\end{align*}
\]

with $(a, b, c, d)$ known.
Hierarchical Poisson model (cont’d)

- The joint posterior distribution is

\[
p(\mu, \alpha, \beta | y) \propto \prod_i \mu_i^{y_i} \exp\{-\mu_i\} \mu_i^{\alpha-1} \exp\{-\mu_i\beta\} \alpha^{a-1} \exp\{-\alpha b\} \beta^{c-1} \exp\{-\beta d\}
\]

- To carry out Gibbs sampling we need to find the full conditional distributions.

- Conditional for \( \mu_i \) is

\[
p(\mu_i | \text{all}) \propto \mu_i^{y_i+\alpha-1} \exp\{-\mu_i(\beta + 1)\},
\]

which is proportional to a Gamma with parameters \((y_i + \alpha, \beta + 1)\).

- The full conditional for \( \alpha \) is

\[
p(\alpha | \text{all}) \propto \prod_i \mu_i^{\alpha-1} \alpha^{a-1} \exp\{\alpha b\}.
\]

- The conditional for \( \alpha \) does not have a standard form.
Hierarchical Poisson model (cont’d)

• For $\beta$:

$$p(\beta | \text{all}) \propto \Pi_i \exp\{-\beta \mu_i\} \beta^{c-1} \exp\{-\beta d\} \propto \beta^{c-1} \exp\{-\beta (\sum_i \mu_i + d)\},$$

which is proportional to a Gamma with parameters $(c, \sum_i \mu_i + d)$.

• Computation:

  – Given $\alpha, \beta$, draw each $\mu_i$ from the corresponding Gamma conditional.
  
  – Draw $\alpha$ using a Metropolis step or rejection sampling or inverse cdf method.
  
  – Draw $\beta$ from the Gamma conditional.

• See Italian marriages example.
Poisson regression

- When rates are not exchangeable, we need to incorporate covariates into the model. Often we are interested in the association between one or more covariate and the outcome.

- It is possible (but not easy) to incorporate covariates into the Poisson-Gamma model.

- Christiansen and Morris (1997, JASA) propose the following model:

- Sampling distribution, where \( e_i \) is a known exposure:
  \[
  y_i | \lambda_i \sim \text{Poisson}(\lambda_ie_i).
  \]
  Under model, \( E(y_i/e_i) = \lambda_i \).

- Population distribution for the rates:
  \[
  \lambda_i | \alpha \sim \text{Gamma}(\zeta, \zeta/\mu_i),
  \]
  with \( \log(\mu_i) = x'_i\beta \), and \( \alpha = (\beta_0, \beta_1, \ldots, \beta_{k-1}, \zeta) \).

- \( \zeta \) is thought of as an unobserved prior count.
Hierarchical Poisson model (cont’d)

• Under population model,

\[ E(\lambda_i) = \frac{\zeta}{\zeta/\mu_i} = \mu_i \]
\[ CV^2(\lambda_i) = \frac{\mu_i^2}{\zeta \mu_i^2} \cdot \frac{1}{\zeta} = \frac{1}{\zeta} \]

• For \( k = 0 \), \( \mu_i \) is known. For \( k = 1 \), \( \mu_i \) are exchangeable. For \( k \geq 2 \), \( \mu_i \) are (unconditionally) nonexchangeable.

• In all cases, standardized rates \( \lambda_i/\mu_i \) are Gamma(\( \zeta, \zeta \)), are exchangeable, and have expectation 1.

• The covariates can include random effects.
Hierarchical Poisson model (cont’d)

• To complete specification of model, we need priors on $\alpha$.

• Christensen and Morris (1997) suggest:
  – $\beta$ and $\zeta$ independent a priori.
  – Non-informative prior on $\beta$’s associated to ‘fixed’ effects.
  – For $\zeta$ a proper prior of the form:

$$p(\zeta|y_0) \propto \frac{y_0}{(\zeta + y_0)^2},$$

where $y_0$ is the prior guess for the median of $\zeta$.

• Small values of $y_0$ (for example, $y_0 < \hat{\zeta}$ and $\hat{\zeta}$ the MLE of $\zeta$) provide less information.
Poisson regression

• When the rates cannot be assumed to be exchangeable, it is common to choose a generalized linear model of the form:

\[
p(y|\beta) \propto \prod_i \exp\{-\lambda_i\} \lambda_i^{y_i}
= \prod_i \exp\{-\exp(\eta_i)\}[\exp(\eta_i)]^{y_i},
\]

for \( \eta_i = x_i'\beta \) and \( \log(\lambda_i) = \eta_i \).

• The vector of covariates can include one or more random effects to accommodate additional dispersion (see epilepsy example).

• The second-level distribution for the \( \beta \)'s will typically be flat (if covariate is a ‘fixed’ effect) or normal

\[
\beta_j \sim \text{Normal}(\beta_{j0}, \sigma_{\beta_j}^2)
\]

if \( j \)th covariate is a random effect. The variance \( \sigma_{\beta_j}^2 \) represents the between ‘batch’ variability.
Epilepsy example

- From Breslow and Clayton, 1993, JASA.

- Fifty nine epileptic patients in a clinical trial were randomized to a new drug: $T = 1$ is the drug and $T = 0$ is the placebo.

- Covariates included:
  - Baseline data: number of seizures during eight weeks preceding trial
  - Age in years.

- Outcomes: number of seizures during the two weeks preceding each of four clinical visits.

- Data suggest that number of seizures was significantly lower prior to fourth visit, so an indicator was used for V4 versus the others.

- Two random effects in the model:
  - A patient-level effect to introduce between patient variability.
– A patients by visit effect to introduce between visit within patient dispersion.
Epilepsy study – Program and results

model {
  for(j in 1 : N) {
    for(k in 1 : T) {
      + alpha.Trt * (Trt[j] - Trt.bar)
      + alpha.BT  * (BT[j] - BT.bar)
      + alpha.Age * (log.Age[j] - log.Age.bar)
      + alpha.V4  * (V4[k] - V4.bar)
      + b1[j] + b[j, k]
      y[j, k] ~ dpois(mu[j, k])
      b[j, k] ~ dnorm(0.0, tau.b);       # subject*visit random effects
    }
    b1[j] ~ dnorm(0.0, tau.b1)        # subject random effects
    log.Base4[j] <- log(Base[j] / 4)
    log.Age[j] <- log(Age[j])
  }
  # covariate means:
  log.Age.bar <- mean(log.Age[])  
  Trt.bar  <- mean(Trt[])  
  BT.bar <- mean(BT[])  
  log.Base4.bar <- mean(log.Base4[])  
  V4.bar <- mean(V4[])  

  # priors:
  a0 ~ dnorm(0.0,1.0E-4)  
  alpha.Base ~ dnorm(0.0,1.0E-4)  
  alpha.Trt ~ dnorm(0.0,1.0E-4);  
  alpha.BT ~ dnorm(0.0,1.0E-4)  
  alpha.Age ~ dnorm(0.0,1.0E-4)  
  alpha.V4 ~ dnorm(0.0,1.0E-4)  
  tau.b1 ~ dgamma(1.0E-3,1.0E-3); sigma.b1 <- 1.0 / sqrt(tau.b1)  
  tau.b ~ dgamma(1.0E-3,1.0E-3); sigma.b <- 1.0/ sqrt(tau.b)  

  # re-calculate intercept on original scale:
}

# re-calculate intercept on original scale:
}
Individual random effects

Difference in expected seizure counts between fourth and first measurements
Diff in seizure counts between V4 and V1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std</th>
<th>2.5th</th>
<th>Median</th>
<th>97.5th</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha.Age</td>
<td>0.4677</td>
<td>0.3557</td>
<td>-0.2407</td>
<td>0.4744</td>
<td>1.172</td>
</tr>
<tr>
<td>alpha.Base</td>
<td>0.8815</td>
<td>0.1459</td>
<td>0.5908</td>
<td>0.8849</td>
<td>1.165</td>
</tr>
<tr>
<td>alpha.Trt</td>
<td>-0.9587</td>
<td>0.4557</td>
<td>-1.794</td>
<td>-0.9637</td>
<td>-0.0676</td>
</tr>
<tr>
<td>alpha.V4</td>
<td>-0.1013</td>
<td>0.08818</td>
<td>-0.273</td>
<td>-0.09978</td>
<td>0.07268</td>
</tr>
</tbody>
</table>