Extentions to Models for Count Data

Author: Nicholas Reich, transcribed by Herb Susmann edited by Hachem Saddiki

Course: Categorical Data Analysis (BIOSTATS 743)

Extensions to Models for Count Data

There are several ways to extend models for count data in order to capture properties like overdispersion.

- Poisson model with adjustment for overdispersion (see previous notes)
- Poisson-Gamma Model
- Generalized Linear Mixed Models (GLMMs)

Poisson-Gamma Model

A Poisson-Gamma model is one way to account for overdispersion in models of count data. The model has two parts:

- First, we assume that the outcome variable follows a Poisson distribution: $Y|\lambda \sim \text{Poisson}(\lambda)$.
- Second, we assume that the rate parameter for that Poisson distribution itself follows a Gamma distribution: $\lambda \sim \text{Gamma}(k, \mu)$.
  - Under this parameterization, $\text{E}[\lambda] = \mu$, $\text{Var}[\lambda] = \mu^2/k$.
  - We can alternatively parameterize in terms of a dispersion parameter $\gamma = 1/k$.

Poisson-Gamma Model

- Under these two assumptions, the marginal distribution of Y follows a negative binomial distribution: $Y \sim \mathrm{NegativeBinomial}(\mathrm{k}, \mu)$
- See this blog post for a proof that the Poisson-Gamma model is a negative binomial distribution.
- The mean and variance of the Poisson-Gamma model is not equal (as opposed to a Poisson model), which allows it to account for overdispersion.

Poisson-Gamma Model

- The expected value of Y is given by:

$$\begin{aligned}
\mathrm{E}[Y] &= \mathrm{E}[\mathrm{E}[Y|\lambda]] \\
&= \mathrm{E}[\lambda] \\
&= \mu
\end{aligned}$$

- And the variance:

$$\begin{aligned}
\mathrm{Var}[Y] &= \mathrm{E}[\mathrm{Var}[Y|\lambda]] + \mathrm{Var}[\mathrm{E}(Y|\lambda)] \\
&= \mathrm{E}[\lambda] + \mathrm{Var}[\lambda] \\
&= \mu + \mu^2/k, \text{ or equivalently} \\
&= \mu + \gamma\mu^2
\end{aligned}$$

- Note that as $\gamma \to 0$ ($k \to \infty$), the distribution of Y approaches a Poisson distribution.

Generalized Linear Mixed Models

Another approach is to use a Generalized Linear Mixed Model (GLMM).

- First, assume that the outcome Y_i follows a Poisson distribution.
- Assume the link-transformed expected value of the outcome is a linear function of the covariates and random effects:

$$\log(\mathrm{E}[Y_i|\mu_i]) = X_{ij}^T\beta + \mu_i$$

- Finally, assume that the random effects u_i follow a distribution:

$$u_i \sim N(0, \sigma^2)$$

- This example uses a log link and assumes the u_i are normally distributed.

Generalized Linear Mixed Models

- Note that you need to use a link that transforms the linear predictor to a non-negative value. For example, the identity link leads to structural problems because a negative linear predictor implies a negative expected count, which is impossible.

Other choices are possible for the distribution of u_i:

- Assuming $u_i \sim \mathrm{Gamma}(1, \gamma)$ implies a negative binomially distributed outcome Y.
- Another possible choice is assume u_i follow a log-normal distribution.
- Each choice implies a different structure for the random intercepts.