PoliSci 4782 Political Analysis II

Count Outcome Models

Seamus Wagner

The Ohio State University

Event Count Applications

- Event count: number of events in a time period for some units
- A count variable implies time and unit dimensions
  - Sometimes a rate is better than a count for modeling
- Y must be 0 or some positive integers, without an upper limit
- A few examples:
  - The annual number of international conflict incidents
  - The annual number of presidential appointments to the Supreme Court
  - The annual number of coups d'etat in African states

Road Map for Modeling Count Outcomes

- Use **rate** ($= \frac{count}{size\_of\_unit}$) instead of count as the outcome variable and run a linear model, if counts are highly sensitive to the size of unit
  - county population $\rightsquigarrow$ COVID cases
- **Linear models** on counts also work, if counts are sufficiently large (probability distribution approximates the normality)
  - people don't have objections to running linear regression on population
- Use the following models:
  - **Poisson models**, if Poisson dispersion can be assumed
  - **quasi-Poisson** or **negative binomial**, if data is over-dispersed
  - *zero-inflated models/hurdle models*, if there are excess zero counts

Poisson Model

1. The **outcome variable**:

$$Y \sim Poisson(\lambda)$$

2. The **systematic component**:

$$\sum_{j=1}^{k} X_j \beta_j = \boldsymbol{X\beta}$$

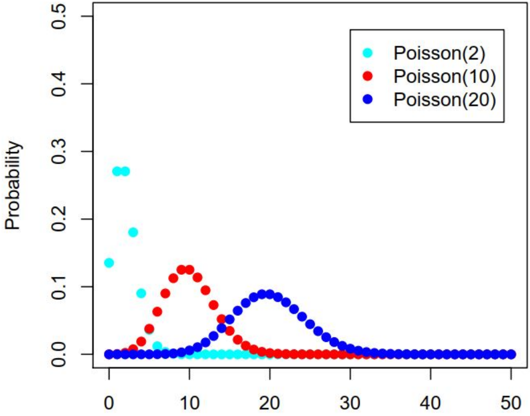
3. The **link function**

$$\ln(\lambda) = \sum_{j=1}^{k} X_j \beta_j$$

4. The **mean function** (inverse to the link function)

$$\lambda = \exp(\sum_{j=1}^{k} X_j \beta_j)$$

Poisson Probability Density



$$Pr(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \qquad y = 0, 1, 2 \ldots \qquad E(Y) = \text{Var}(Y) = \lambda$$

Poisson GLM

- We parameterize λ in the Poisson probability density function using $\lambda = \exp(\beta \boldsymbol{X})$ to make our generalized linear model
- When λ (the mean of Y) increases, Poisson distribution approximates the normal distribution (why we don't need Poisson to model population)
- Pay attention to the Poisson dispersion assumption $E(Y|\boldsymbol{X}) = Var(Y|\boldsymbol{X}) = \lambda$. Otherwise, Poisson models do not work:
  - If we assume Poisson dispersion but data are over-dispersed $[E(Y|\boldsymbol{X}) < Var(Y|\boldsymbol{X})]$, estimated standard errors are too small
  - If we assume Poisson dispersion, but data are under-dispersed $[E(Y|\boldsymbol{X}) > Var(Y|\boldsymbol{X})]$, estimated standard errors are too large (rarely happens)

Poisson Model Interpretation

The mean function for Poisson models is
$\lambda = \exp(\boldsymbol{\beta X}) = \exp(\beta_0) \times \exp(\beta_1 x_1) \cdots \times \exp(\beta_k x_k)$

- Similar to the odds interpretation for logit, variable X_i affects Y through $\exp(\beta_i)$ in a multiplicative way
- $\exp(\beta_i)$ indicates the expected multiplicative difference in Y given one unit change in X_i, holding all other X constant

An Example in Gelman and Hill 2006

- i indexes street intersections and y_i is the number of collisions at intersection i in a given year

- Consider three parameters:
  - a constant term
  - a measure of the average speed of traffic near the intersection (X_1)
  - an indicator for whether the intersection has a traffic signal (X_2)

- Suppose we estimated the model and found $\alpha = 2.8$, $\beta_1 = 0.012$, and $\beta_2 = -0.20$

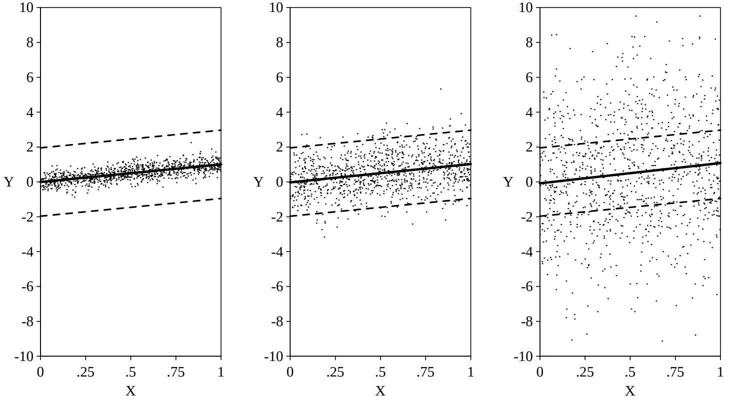
- We could write this in the following way

$$y_i \sim Poisson(\exp(2.8 + 0.012X_{i1} - 0.20X_{i2}))$$

Interpreting Poisson Results

- Coefficients β need to be exponentiated and treated as multiplicative effects.
- The coefficient of X_{i1} is the expected difference in y (on the logarithmic scale) for each additional unit of X_{i1}.
- Thus, the expected multiplicative increase is $e^{0.012} = 1.012$ or a 1.2% positive difference in the value of Y for a unit increase in X.
- Likewise, for X_{i2}, $e^{-0.20} = 0.82$ yielding a reduction of 18% (Consider what happens when we multiple y by 0.82).

Under-, Poisson-, and Over-Dispersion



(a) (b) (c)

A more rigorous test is provided by Cameron and Trivedi (1990) and can be easily performed in R

Dealing with Over-Dispersion

- **Quasi-Poisson**:
  - The idea is simple: introducing an extra parameter ϕ ($\frac{\sum_i (y-\lambda)^2/\lambda}{n-k}$) to Poisson such that we don't have to be restricted to $E(Y|X) = Var(Y|X) = \lambda$

- **Negative-binomial**:
  - $Y \sim NegBin(r, p)$
  - the original idea of *NegBin* is about the number of successes in a sequence of independent and identically distributed Bernoulli trials (governed by the probability parameter p) before a specified (non-random) number of failures (denoted r) occurs

We interpret coefficients from the two models in the same way as we do in Poisson.

Coming Up

- Lab tutorial on Poisson, quasi-Poisson, and negative binomial (also dispersion test)
- Lab tutorial on log likelihood ratio test
- Other discrete outcome models (ordered and unordered categorical variables) next week