Intro to GLM – Day 5: Count Data

Federico Vegetti
Central European University

ECPR Summer School in Methods and Techniques
12 August 2016

Counts as indicators

- In political science, some phenomena are observed by counting events
  - The number of parliamentary questions asked by an MP in a legislature
  - The number of wars in which a country has been involved in a certain period of time
  - The number of coups in a certain region
- These counts are often proxies of broader, usually latent phenomena that we want to analyze
- For instance, we can have an idea of how “pacific” a country is by counting by the number of conflict events in which it participates

Count data

- What are the characteristics of count data?
  - Counts can not be smaller than zero
  - Counts often present many zeros
  - Distribution of counts are usually right-skewed
- For these reasons (mostly the first two), OLS methods are not appropriate
- As with other cases we encountered these days, we need to transform the distribution of the error and the functional form to depart from the usual normality and linearity assumptions

Review of GLM

The three steps of GLM

1. Specify the joint distribution of $Y|X$. Which type of random process has generated our data?
2. Specify the function to transform the expectation of Y into the linear predictor
3. Specify the equation of the linear predictor. How does it relate to our Xs?

Poisson distribution

- A distribution that is commonly assumed for counts is the Poisson distribution

$$P(Y_i|\lambda) = \frac{exp(-\lambda)\lambda^{y_i}}{y_i!} \quad \text{for } \lambda > 0$$

- Where:

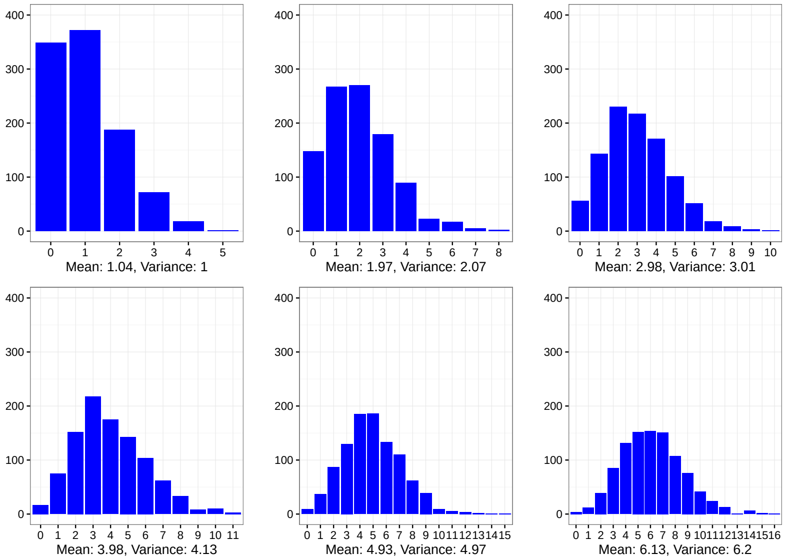
$$\lambda = E(y) = Var(y)$$

(yes, λ is both the conditional mean of y and its variance)

- The probability function of the population is the product over the individual observations:

$$P(y_i|\lambda) = \prod_{i=1}^{N} \frac{exp(-\lambda)\lambda^{y_i}}{y_i!}$$

How λ affects the distribution



Link function

- Because λ is a (mean) count, it can take only positive values
- We need a function to drop this constraint
- The most straightforward way is to take the *log*:

$$X\beta = log(\lambda)$$

- So the response function is:

$$\lambda = E(Y) = exp(X\beta)$$

- The likelihood function of Poisson models consists in plugging $exp(X\beta)$ into the probability function
- The software will do the usual work maximizing the log-likelihood function

Quantities of interest

- As we saw, the transformation performed by the link function is a simple log transformation
- This means that our coefficients are logs of the expected (increment in) counts, given 1 point increase of X
- Therefore, the calculation of a meaningful quantity is straightforward: to obtain λ, we only need to exponentiate the predicted values on the linear predictor
- A quantity of interest to present after Poisson models is a draw from a Poisson distribution given the value of λ conditional on interesting values of X

Incident rate ratios

- While we can obtain λ for desired values of X easily, the coefficients are not so easily interpreted
- Let's suppose we have a variable X_1 which has a negative coefficient, like *-1*
- If we exponentiate *-1*, we get a positive value: `exp(-1) = 0.37`
- What does that mean?
- While the raw coefficients can be interpreted as changes in log counts with a unit increase of X, their exponent can be interpreted as "*incident rate ratios*"
  - Incident rate refers here to the number of events occurring within a certain interval
  - The ratio works as with odds ratios
  - In our case, an IRR of `0.37` means that for every unit increase in X, the incidence rate of Y becomes *0.37* times as big as it was before – AKA it is reduced by 63%

Assumptions of Poisson models

Several assumptions, the last two more important

- Every period starts with zero events
- More than one event can not occur at the same time
- The probability of observing an event during a certain interval is the same across intervals
- The probability of observing an event during a certain interval does not depend on whether we observed an event in any other previous interval

Overdispersion

- Turns out that the most problematic assumption of Poisson models is that $\lambda = E(y_i) = Var(y_i)$
- In many practical applications, the variance of our data is much larger than the mean
- If that is the case, the Poisson distribution is not an appropriate description of the process that generated our data
- What consequences?
  - If we underestimate the dispersion of the data, our model will produce standard errors that are smaller than they should be

The Negative Binomial model

- A solution for overdispersion is to use "negative binomial" regression
- The logic is simple:
  - We fit a Poisson model, but we treat λ as an unobservable random variable that follows a Gamma distribution with mean λ and scale parameter θ
- What is a "scale parameter"?
  - A scale parameter is the parameter that governs the dispersion of a random variable
  - E.g. in a normal distribution the scale parameter is the standard deviation
- So compared to the Poisson model, in the Negative Binomial model we estimate one more parameter

Negative Binomial function

- The observed count follows now a negative binomial distribution:

$$P(y_i|\lambda,\theta) = \frac{\Gamma(y_i+\theta)}{y!\Gamma(\theta)} \times \frac{\lambda_i^{y_i}\theta^\theta}{(\lambda_i+\theta)^{\lambda_i+\theta}}$$

- While the link function is the same as in the Poisson model:

$$E(Y) = \lambda = exp(X\beta)$$

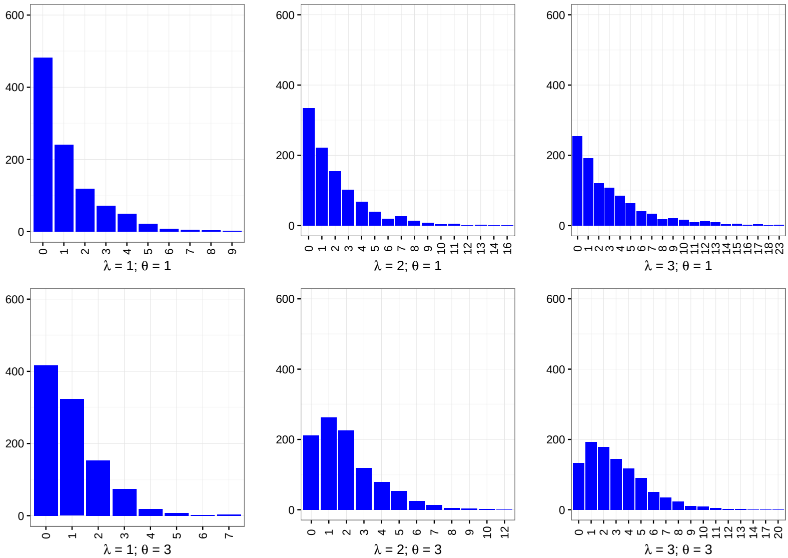
- But now the variance of Y is different:

$$Var(Y) = \lambda + \frac{\lambda^2}{\theta}$$

On the dispersion parameter

- By modeling the variance in this way, we state that the conditional variance of Y increases more rapidly than its expected value
- However, given the nature of the count process, there is still a connection between the two
- When the value of θ is very high, then λ^2/θ tends to 0
- When this is the case, we go back to the situation where $Var(Y) = \lambda$, and the negative binomial model approximates the Poisson model
- Since θ is estimated, we can see to what extent we gain in terms of "correctness" when we use a negative binomial instead of a Poisson model

How λ and θ affect the distribution



Interpretation

- The link function in negative binomial models is still the log of the linear predictor
- Therefore, coefficients are interpreted in the same way as in Poisson models
- In fact, negative binomial models produce coefficients that are very similar to the ones produced by Poisson models
- What changes are the standard errors
- The value of the dispersion parameters θ estimated by the model informs us about the overdispersion of the data

Too many zeros?

- Sometimes a count variable has a large amount of zeros
- This is of course a case of overdispersion
- However, when zeros are really *a lot* compared to the other values, a negative binomial model might fail to fit the data properly
- If we think of what process generated such data, we may take a different perspective
- Perhaps, the process generating the zeros is not the same as the process generating the count

A two-step data generating process

- Suppose we want to study whether different working conditions lead to different patterns of alcohol consumption or abuse
- We take a sample of people and we count how many days they have been drinking alcoholic beverages within a month time
- Our sample will inevitable include people who do not drink alcohol *at all*, and this will result in a lot of cases with count *0*
- Those "zeros" will be of 2 types:
  - **Structural**: individuals that are not at risk for the surveyed behavior
  - **Random**: zeros that happen due to sample variability (e.g. being on a diet)
- In other words, we have two groups in the sample
- They may differ systematically in terms of individual characteristics

Zero-inflated models

- The goal of zero-inflated models is to model two processes simultaneously:
  - The probability to be in the "dry" group
  - The count of drinking days within a month
- Zero-inflated regression models are a *mixture* of two different models:
  - A binary model predicting the probability to be in the "zero" group (usually a logit or probit)
  - A count model for those out of the "zero" group (usually a Poisson or negative binomial)
- The two models can include different predictors

Example: zero-inflated Poisson model

- For the zero-inflated Poisson model, the probability distribution of the outcome is made by the two following functions:

$$P(y_i = 0|X_i, Z_i) = \pi + (1 - \pi)exp(-\lambda)$$

$$P(y_i|X_i, Z_i) = (1 - \pi)\frac{exp(-\lambda)\lambda^{y_i}}{y_i!} \text{ for } y_i > 0$$

- Where :
  - $\pi = F(\beta X)$ and $\lambda = exp(\gamma Z)$
  - F can be either the logistic or normal CDF
  - Predictors X can the same as predictors Z or not
- Parameters are interpreted in the same way as in the respective binary and count models
- In the zero-inflated negative binomial model, the second equation is the one changing

Considerations

- Not all cases of overdispersion justify using a zero-inflated model
- The zero-inflated model should be theoretically motivated
  - Is the process that we want to explain really a mixture?
  - Is the step between *0* and *0*$<$ so different from e.g. *1* and *2*?
- Sometimes a “simple” negative binomial model can take care of overdispersion, without arguing for zero-inflation