Event Count Data

Nathaniel Beck

Department of Politics, NYU, New York, NY 10012, nathaniel.beck@nyu.edu

QII - Week 5

Event Counts

- Suppose y_i counts the number of times that i does something
  - Number of nations joining a sanction
  - Number of times a judge dissents
  - Number of bad checks written by member of congress
  - Number of visits to doctor
- So y takes on values $0, 1, \ldots$ (may in practice be finite, but should be large).
- If only takes on values 0 and 1, use binary models.
- But note argument of Alt, King and Signorino (Political Analysis) that if we have an event count (number of MIDs) that is simply coded
- zero if no MID, one if one or more, than should start with event count (discuss later with censoring)

Poisson

- Simplest model is that y_i is Poisson with mean λ_i. This follows from the assumptions:
  - Events occur over time
  - P(event) is constant over time
  - P(event) is independent of how many prior events occurred
  - P(2 events in small unit of time) $= 0$
- Under these conditions, the number of events that we observe for a unit follows a Poisson distribution characterized by the single parameter density

$$f(y_i) = \frac{e^{-\lambda_i} {\lambda_i}^{y_i}}{y_i!}. \tag{1}$$

- Note that the variance of y_i is also λ, so the model is restrictive (as is any one parameter model).
- As always, need to parametrize λ_i as $f(x_i, \beta)$.
- λ must be positive.
- Usually assume
$$\lambda_i = e^{x_i \beta} \tag{2}$$
- Why? Look at
$$\frac{\partial \lambda_i}{\partial x_j} = \lambda_i \beta_j. \tag{3}$$
  Why is this reasonable? Harder to move from 0 to 1 than from 100 to 101.
- Note that since the expected value of a Poisson is just λ_i, the marginal effect of a change in x_i is just $\lambda_i \beta$.

Likelihood

- Likelihood is then

$$L = \prod_{i=1}^{N} \frac{e^{-\lambda_i} {\lambda_i}^{y_i}}{y_i!} \tag{4}$$

$$= \prod_{i=1}^{N} \frac{e^{-e^{x_i\beta}} (e^{x_i\beta})^{y_i}}{y_i!} \tag{5}$$

- or the log likelihood is

$$\ln L = \sum_{i=1}^{N} [-\lambda_i + y_i x_i \beta - \ln y!] \tag{6}$$

$$= \sum_{i=1}^{N} -e^{x_i\beta} + y_i x_i \beta - \ln y!, \tag{7}$$

- This is a very well behaved likelihood which is easy to maximize.

- The first order condition is very easy to get:

$$\frac{\partial lnL}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} (y_i - \lambda_i)\mathbf{x}_i = \mathbf{0} \tag{8}$$

- Note how similar this is to the logit equation
- Note the constant forces the mean number of events to equal the mean λ_i

Negative Binomial

- The assumption that the variance equals the mean is very strong. Thus the Poisson may be too restrictive
- In practice the variance is usually larger than the assumed by the Poisson. Such a model is called over dispersed.
- Let the mean be as before

$$\lambda_i = e^{x_i \beta} \tag{9}$$

but assume that the variance is

$$\sigma^2 = \lambda_i(1 + \alpha\lambda_i) \tag{10}$$

- One model that can handle this and still allow y to take on only discrete values (DATA ADMISSIBLE) is the NEGATIVE BINOMIAL.
- Thus over-dispersion is

$$\frac{\text{Var}(y_i)}{E(y_i)} = 1 + \alpha E(y_i), \quad \alpha > 0. \tag{11}$$

Other parametrization

- Some people (e.g. King) assume that the over-dispersion (11) is $1 + \alpha$ which leads to a slightly different model. (Would you expect the degree of over dispersion to increase as the mean increases?)
- If you go to a probability text, you will find the usual parametrization of the negative binomial:

$$f(n; r, p) = \begin{pmatrix} n-1 \\ r-1 \end{pmatrix} p^r (1-p)^{n-r} \tag{12}$$

where the mean is $\frac{r}{p}$ and the variance is $\frac{r(1-p)}{p^2}$.
- The 'King' parametrization is

$$f(n; \lambda, \gamma) = \begin{pmatrix} n-1 \\ \lambda(2+\gamma)-1 \end{pmatrix} \left(\frac{1}{2+\gamma}\right)^{\lambda(2+\gamma)} \left(\frac{1+\gamma}{2+\gamma}\right)^{n-\lambda(2+\gamma)} \tag{13}$$

- This difference causes no problems so long as you realize what parametrization you are using!

- Note that this works only for over dispersion.
- The Poisson is not a special case of the negative binomial, which is not defined for $\alpha = 0$.
- Of course you can decide to go with Poisson if *alpha* is small or insignificant.
- Don't have a good way of dealing with under-dispersion, but this appears to be rare. King has a "general event count" distribution which may work.

Censoring

- Sometime the highest event count is N or more (e.g. you might have a few cases with a huge number of events, but only recorded as N or more).
- Then the last entry in the likelihood is not $P(y_i = N)$ but $P(y_i \geq N)$. It is easy to take this into account.
- Thus suppose we go back to the 0-1 coding for MIDs, where 1 is one or more
- Then for a zero, take $P(0)$ from a Poisson or negative binomial, and $P(1) = 1 - P(0)$.
- YABM (not as cool as YACC or YAP)

Truncation at zero

- Sometimes because of our design we only observe $y_i \geq 1$.
- This may be because of choice based sampling.
- That is, instead of doing a survey and asking people how many times they ride the bus, we get on a bus and ask people how many times they ride.
- If we fit a Poisson (or Neg. Bin) we are falsely allowing probability of zero to be positive.
- The simple solution is to change each of the probability entries in the likelihood as follows (with P_{tr} being the truncated (correct) probability and P being the incorrect probability):

$$P_{tr}(y_i = n) = \frac{P(y_i = n)}{1 - P(y_i = 0)}. \tag{14}$$

- Note the assumption we are making is that the process of going from zero to one is the same Poisson process we observe, we just happen not to have chosen to observe zeros. THIS IS A VERY STRONG ASSUMPTION.

Zero Inflated Probability (ZIP)

- Perhaps a better assumption is that the process which govern zero-not zero is different than the process which determines how many times something is done, given it is done.
- Obviously to estimate this model you need info on zeros and non-zeros, though presumably you could get from different samples.
- Think of some binary random variable, z_i which measures if you would ever do something or not. (Some people get counts of zero because they didn't do something but might have, others would never +have even considered doing the thing.)
- If $z_i = 1$ then y_i^* measures how many times you do something. y_i is the observed event count.
- Then we change our probabilities in the likelihood function so (with P_P being the Poisson (or negative binomial), P_B being the probit and P_Z being the ZIP probabilities):

$$P_Z(y_i = 0) = P_B(z_i = 0) + P_B(z_i = 1)P_P(y_i^* = 0) \tag{15}$$
$$P_Z(y_i, y_i > 0) = P(z_i = 1)P_P(y_i) \tag{16}$$

- Note that the probit/logit and Poisson/negative binomial parts may be estimated with the same or different covariates.
- Relationship to hurdle model
- Hurdle - going from 0 to 1, measured by a probit
- Then if we get over the hurdle, event count (with truncation)
- Makes sense - first we ask if someone watches debates at all, then a second and different model for how many