Quantitative Methods in Political Science: Count Models

UNIVERSITY OF MANNHEIM
School of Social Sciences

Thomas Gschwend | Oliver Rittmann | Viktoriia Semenova | David M. Grundmanns

Week 12 - 24 November 2021

The Course

Roadmap

- Understand and model stochastic processes
- Understand statistical inference
- Implement it mathematically and learn how to estimate it
  - OLS
  - Maximum Likelihood
- Implement it using software
  - R
  - Basic programming skills

Overview: Week 12

Count Models

Example Today: One-Sided Violence Against Civilians

Count Models as Generalized Linear Models

- Poisson Models
- Negative Binomial

Count Models

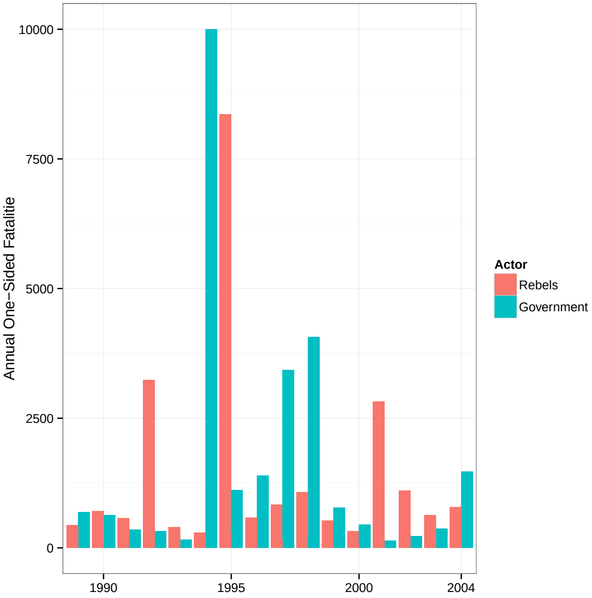
Count Models

- Oftentimes our dependent variables are counts of discrete events
  - number of bills passed in legislature per month
  - number of parliamentary questions asked per MP per year
  - number of military conflicts per year
  - number of Coups d'Etat in black African states
  - number of news stories about a politician per day
  - number of presidential vetoes per presidential term
- No (theoretical) upper limit on the number of observed events

Example Today: One-Sided Violence Against Civilians

Example Today: One-Sided Violence Against Civilians

- Kristine Eck & Lisa Hultan present a data-set on direct and deliberate killings of civilians (one sided violence) in interstate armed conflicts, during 1989-2004.
- Here, we are especially interested in the question if one-sided violence committed by governments or rebel groups



Annual One−Sided Fatalitie
10000
7500
5000
2500
0
1990
1995
2000
2004
Actor
Rebels
Government

How to model counts using Poisson ?

- Counts take on discrete values $(0, 1, \dots)$ and are bounded between 0 and $+\infty$
- We typically cannot observe the underlying data generating process (how events occur)
- We only observe the number of events at the end of the “observation period”.
    - For example, we can count or estimate the number of one-sided violence in a year, but conceiving a list of killed/not-killed is difficult
- $Pr(\text{event at time } t \mid \text{all events up to time } t-1)$ is constant for all t, i.e., the probability of an event occurring at a certain time is constant and independent of all previous events.

Count Models as Generalized Linear Models

The structure of Generalized Linear Models

Like in the logit/probit case, we can use the *generalized linear model* setup

A Generalized Linear Model (GLM) consist three components

- *Stochastic Component* (1), specifying the conditional distribution of the dependent variable Y_i
- *Systematic Component*, consisting of a linear function of predictors (2), e.g.,

$$\eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} \tag{1}$$

- and (3) a *Link Function* $h(\cdot)$ which transforms the expectation of the dependent variable, $\mu_i = E(Y_i)$, to the linear predictor

$$h(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} \tag{2}$$

Generalized Linear Models

- Models we discussed so far:

Model	Distribution	Link	$h(\mu_i) = \eta_i$	Range Y_i
Linear	Gaussian	Identity	(μ_i)	$(-\infty, \infty)$
Logit	Bernoulli	Logit	$log \frac{\mu_i}{1-\mu_i}$	$[0, 1]$
Probit	Bernoulli	Probit	$\Phi^{-1}(\mu_i)$	$[0, 1]$

Poisson Model

- Y_i is drawn from a Poisson distribution with (only a single) parameter λ_i :

$$Y_i \sim Poisson(\lambda_i)$$

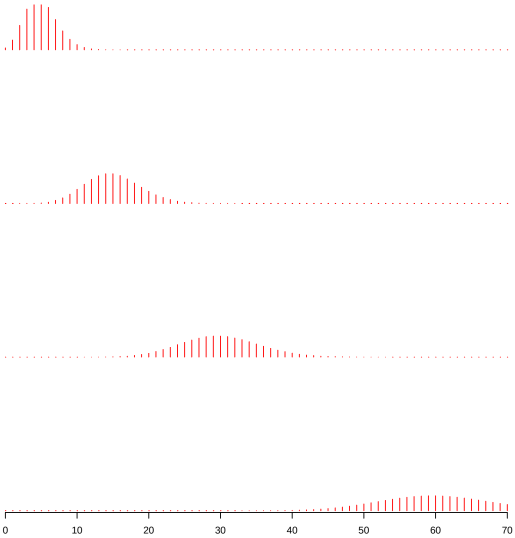
- The Poisson PDF for a single observation:

$$f_{Poisson}(y_i|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} & \text{for } \lambda > 0 \text{ and } y_i = 0, 1, \ldots \\ 0 & otherwise \end{cases}$$

- The probability density of all the data (i.e., N observations, given that Y_i and Y_j are independent conditional on X for all $i \neq j$ and identically distributed) is the product of all N individual observations:

$$Pr(Y|\lambda) = \prod_{i=1}^{N} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

Poisson with $\lambda = [5, 15, 30, 60]$



Poisson Model

- It can be shown that for a Poisson distribution λ_i is the mean and variance parameter, i.e., $E(Y_i) = Var(Y_i) = \lambda_i$

- We will use an exponential response function, because $\lambda_i > 0$

$$E(Y_i) = \lambda_i = e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik}} = e^{X_i \beta}$$

- Remember we need to assume that the probability of an event occurring at a certain time is constant and independent of all previous events.

Poisson Model

- Deriving the likelihood function:

$$L(\lambda|Y) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$$

$$\ln L(\lambda|Y) = \sum_{i=1}^{n}(y_i ln(\lambda_i) - \lambda_i - ln(y_i!))$$

$$\ln L(\lambda|Y) = \sum_{i=1}^{n}(y_i ln(\lambda_i) - \lambda_i)$$

$$\ln L(\beta|Y) = \sum_{i=1}^{n}(y_i(X_i\beta) - e^{X_i\beta})$$

- As before, we maximize the log-likelihood to get $\hat{\beta}$.

Poisson Model

- Quantities of interest:
  - Expected values (counts) given specific values of X

$$E(Y_i|X_i) = \lambda_i = e^{X_i\beta}$$

  - First-differences (entertain interesting counterfactual)
  - Predicted values (counts)
- As usual, to simulate from a Poisson model we do the following:
  1. Draw $\tilde{\beta}$'s repeatedly from the multivariate normal $N(\hat{\beta}, \hat{V}(\hat{\beta}))$, i.e. the simulated sampling distribution of the β's to account for *estimation* uncertainty
  2. Define scenario of interest by setting the IV's to particular values (X_c)
  3. Compute $\tilde{\lambda}_c = e^{X_c\tilde{\beta}}$ and take the mean over all $\tilde{\lambda}_c$'s to get the expected counts
  4. Draw Y_c from $f_{Poisson}(Y|\tilde{\lambda}_c)$ to account for *fundamental* uncertainty if you wanna get predicted counts, given $\tilde{\lambda}_c$

Results for Example: Poisson Model

	Dependent variable:		
	Count of One Sided Violence		
	(1)	(2)	(3)
		Government	Rebel
Civil War	0.626***	0.222***	0.901***
	(0.008)	(0.013)	(0.011)
Autocracy	0.473***	0.439***	0.550***
	(0.011)	(0.016)	(0.016)
Democracy	−0.066***	−1.403***	0.402***
	(0.014)	(0.032)	(0.017)
Government	−0.073***		
	(0.008)		
One sided Violence t-1	0.001***	0.001***	0.001***
	(0.00000)	(0.00001)	(0.00001)
Constant	2.878***	3.572***	2.306***
	(0.015)	(0.022)	(0.021)
Observations	1,178	435	743
Log Likelihood	−158,122.600	−65,303.790	−89,707.900
Akaike Inf. Crit.	316,257.300	130,617.600	179,425.800
Note:			*p<0.1; **p<0.05; ***p<0.01
			Excluding observation Rwanda 1994

Negative Binomial Model

- A limitation of the Poisson distribution is the equality of the conditional mean and variance of y .

$$Var(Y|X) = E(Y|X) = \lambda$$

- Often our data are overdispersed, i.e. the variance is actually larger than the mean.

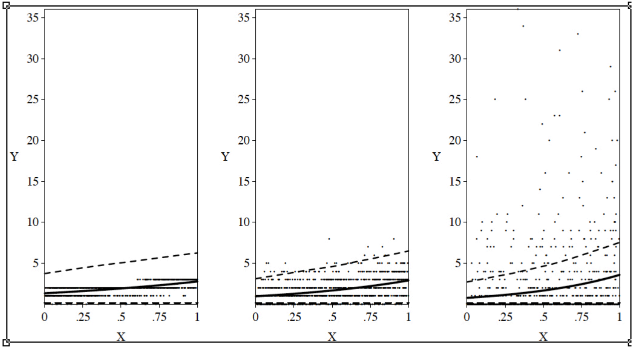
$$Var(Y|X) > E(Y|X)$$

- The consequence of overdispersion is that the standard errors are biased towards zero (i.e., they are too small)
- We could use robust standard errors, but then we only get the mean right (very limited!)
- Therefore, we model overdispersion, i.e. we parameterize it. This yields the the so-called negative binomial model.

What happens without an extra dispersion parameter?

Suppose we run a poisson when counts are …

…underdispersed, $Var(Y_i|X_i) = E(Y_i|X_i)$, or overdispersed.



$E(Y_i|X_i)$ and 95% CI

Negative Binomial Model

- The basic idea is to adopt a Poisson model for the count Y_i, but to suppose that the expected count λ^* is itself an unobservable random variable that is Gamma distributed with mean λ and scale (or overdispersion) parameter θ.

- Then the observed count Y_i follows a *negative binomial distribution:*

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

- Mean: $E(Y_i) = \lambda_i = e^{X_i \beta}$

Negative Binomial Model

- The negative binomial model includes a parameter that can take care of the overdispersed data structure:

$$Var(Y_i) = \lambda_i + \frac{1}{\theta}\lambda_i^2, \text{ or}$$

$$Var(Y_i) = \lambda_i + \alpha\lambda_i^2$$

- The negative binomial model has one more parameter than the Poisson
- If $\alpha = 0$ (or $\theta \to \infty$), the negative binomial model is identical to the Poisson model
- This means we can test whether the negative binomial or the Poisson model is more appropriate for our data.

Negative Binomial Model

- Let $\alpha = \frac{1}{\theta}$ represent the inverse of the scale parameter.
- Our null hypothesis $H_0 : \alpha = 0$ ("Poisson model")
- Our alternative hypothesis $H_a : \alpha > 0$.
- Compute likelihood-ratio test (LRT) from chi-square distribution with one degree of freedom.
  - In order to obtain corresponding p-value note that right-tailed p-value must be halved since the negative binomial over-dispersion parameter is restricted to be positive (one-sided test)

Results for Example: Negative Binomial Model

	Dependent variable:		
	Count of One Sided Violence		
	(1)	(2)	(3)
		Government	Rebel
Civil War	1.042***	0.903	1.012**
	(0.340)	(0.629)	(0.406)
Autocracy	0.561	0.822	0.347
	(0.369)	(0.703)	(0.431)
Democracy	0.394	−0.322	0.580
	(0.410)	(0.817)	(0.468)
Government	−0.002		
	(0.296)		
One sided Violence t-1	0.009***	0.008***	0.010***
	(0.001)	(0.001)	(0.001)
Constant	1.237**	1.449	1.282**
	(0.526)	(0.988)	(0.609)
Observations	1,178	435	743
Log Likelihood	−2,407.617	−718.369	−1,679.689
α	0.042*** (0.003)	0.029*** (0.004)	0.053*** (0.004)
Akaike Inf. Crit.	4,827.235	1,446.738	3,369.378
Note:			*p<0.1; **p<0.05; ***p<0.01
			Excluding observation Rwanda 1994

Use simulated quantity of interest instead