Event Count Models

Evgeny Sedashov, PhD
esedashov@hse.ru

4/02/2026

Introduction

- Another class of variables frequently encountered in social sciences – event counts.
- Number of terror attacks per year.
- Number of protests in a certain country and year.
- Number of crimes of a certain type reported in the specific municipality.

Event Count Models

- Event count variables do not satisfy normality assumption and 0 conditional mean assumption.
- In the typical setting, event count variable has a theoretically defined lower bound of 0, but no theoretically defined upper bound.
- This makes sense – number of events can never be negative, but we cannot know in advance the maximum possible number of protests in a given country and year.
- One needs to search for distributions that correspond to $[0; \infty)$ support.

First Steps

- One easy alternative is to assume that the event count variable follows Poisson distribution.
- The corresponding model is called Poisson regression.
- The main advantage of Poisson regression framework – analytical simplicity.

Exponential Link Function I

- Recall that in any regression setting we would like to model a conditional expectation function:

$$\mathbb{E}[y|\mathbf{x}_i] = f(\mathbf{x}_i)$$

  where $\mathbf{x}_i$ is some vector of values corresponding to independent variables.

- Poisson distribution has only one parameter that completely describes its behavior – the mean $\lambda \in (0; \infty)$.

- We would like to model λ conditioned on X_i.

Exponential Link Function II

- Our traditional equation relating y and $\mathbf{x}_i$:

$$y = \beta_0 + \beta_1 x_{i1} + ... + \beta_n x_{in}$$

- We cannot use this equation directly as a conditional mean because the target quantity, λ_i, is bounded below with 0, while expression above is not.

- To map the linear regression equation into the appropriate scale, we use exponential link function:

$$\mathbb{E}[y|\mathbf{x}_i] = \exp(\beta_0 + \beta_1 x_{i1} + ... + \beta_n x_{in}) = \lambda(\mathbf{x}_i)$$

Likelihood Function I

- Next, recall that the Poisson distribution models the probability of event count $= c$ as

$$\mathbb{P}[C = c] = \frac{\lambda^c e^{-\lambda}}{c!}$$

- Now imagine that instead of drawing counts independently from the same distribution we actually draw counts from different distributions (but also independently), so λ now is not a constant.

- Certain observed parameters-independent variables serve as a signal corresponding to a specific underlying λ.

Likelihood Function II

- Substituting, we obtain

$$\mathbb{P}[C = c] = \frac{\lambda(\mathbf{x}_i)^c e^{-\lambda(\mathbf{x}_i)}}{c!} =$$

$$\frac{\exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_n x_{in})^c e^{-\exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_n x_{in})}}{c!} =$$

$$\frac{\exp(c(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_n x_{in}) - \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_n x_{in}))}{c!}$$

- This looks complicated, but it will be simplified later.

- Since the observations are independent, the total likelihood is

$$\mathcal{L} = \prod_{i=1}^{N} \frac{\lambda(\mathbf{x}_i)^{c_i} e^{-\lambda(\mathbf{x}_i)}}{c_i!}$$

Log-likelihood I

- To simplify the work for the computer, we take the log of this product and convert it into the sum of logs:

$$\log(\mathcal{L}) = \sum_{i=1}^{N} \log(\frac{\lambda(\mathbf{x}_i)^{c_i} e^{-\lambda(\mathbf{x}_i)}}{c_i!}) =$$

$$\sum_{i=1}^{N} [\log(\lambda(\mathbf{x}_i)^{c_i} e^{-\lambda(\mathbf{x}_i)}) - \log(c_i!)]$$

- Notice that $\log(c_i!)$ does not depend on $\mathbf{x}_i$, hence, it won't affect optimization results and can be dropped.

- On the other hand, we can substitute the first part with the numerator from the previous slide, first bullet point.

Log-likelihood II

- Substituting, we obtain:

$$\sum_{i=1}^{N}[c_i(\beta_0+\beta_1 x_{i1}+\ldots+\beta_n x_{in})-\exp(\beta_0+\beta_1 x_{i1}+\ldots+\beta_n x_{in})]$$

  which is the log-likelihood for the Poisson regression model.

- Likelihoods reported by software usually contain $-\log(c_i!)$, but, as we established, this actually does not affect the estimated parameters.

- The computer will find parameters $\hat{\beta}$ for you that maximize this function; various numerical optimization algorithms are available due to the absence of a closed-form solution.

Regression Coefficients

- Similarly to binary response models, regression coefficients in Poisson regression do not have a unit-change interpretation.
- Signs, however, once again have intuitive interpretation, with minus corresponding to a negative effect and plus corresponding to a positive effect.
- Similarly to binary choice models, it is advisable to use graphical tools when interpreting estimates from Poisson regression.

Interpretation

- Interpreting results of Poisson regression is easy.
- One can use endpoint transformation algorithm.
- Or one can use CLARIFY algorithm.

CLARIFY Approach for Poisson Regression I

- Set all variables, except for the variable of interest, to their means or modes.
- Draw the sample (normally 1000 or 10000 observations) from a multivariate normal distribution with mean equal to estimated regression coefficients and variance equal to the variance-covariance matrix – call this sample D.
- Multiply each vector from the matrix of independent variables by the transpose of D.

CLARIFY Approach for Poisson Regression II

- Apply the exponential function to the resulting matrix.
- For each vector of independent variables, compute the mean predicted value, 2.5 and 97.5 percentiles.
- Draw the graph with the variable of interest on the x axis and predicted values on the y axis.
- Predicted values in this context are expected counts.

Nested vs. Non-nested Models I

- MLE framework has an analogue for the F-test that we used in OLS setting to test exclusion restrictions.
- Suppose you have two models, with the Model 1 being the full model and the Model 2 being the nested model, i.e. its independent variables are a subset of independent variables from the full model.
- We would like to know whether the Model 1 provides a significantly better fit to the data than the Model 2.

Nested vs. Non-nested Models II

- Likelihood ratio test provides a simple algorithm for testing such hypotheses.
- The null hypothesis H_0 in this context is the hypothesis that the nested model is true.
- The alternatives hypothesis H_1 is that the full model is true.
- The test statistic is the likelihood ratio:

$$\lambda_{LR} = -2(\log(\mathcal{L}_{nested}) - \log(\mathcal{L}_{full}))$$

where $\log(\mathcal{L}_{nested})$ and $\log(\mathcal{L}_{full})$ are log-likelihoods for nested and full models, respectively.

Nested vs. Non-nested Models III

- Wilks' theorem states that, as the sample size grows, the distribution of likelihood ratio statistic approaches chi-square distribution with degrees of freedom parameter equals to the number of variables excluded from the full model to obtain the nested model.

- Therefore, for the computed likelihood ratio statistic we can obtain the *p*-value as the probability that chi-square random variable exceeds the computed likelihood ratio.

- We then compare this *p*-value to standard critical rejection values 0.05, 0.01, 0.001.

Offset I

- Some settings require adding offset to the event count models.
- What is offset?
- In the protest setting, offset is the population of a country – the greater population means that there are more people who can attend protests.

Offset II

- Offset is not a variable per se and does not have a corresponding parameter – it is included in the model as the independent variable with the coefficient constrained to 1.
- Taking a log of offset gives exposure – it can also be included in the model as the independent variable with the coefficient constrained to 1.
- I recommend using exposure, but offset can be more appropriate under certain circumstances.

Introduction

- Negative Binomial regression is another approach to modelling event count data.
- One can guess that the main difference is rooted in the distributional assumption: instead of assuming the count data follows Poisson distribution, we will now assume that data follows Negative Binomial distribution.
- Why use Negative Binomial distribution at all?

Poisson vs. Negative Binomial I

- Poisson distribution assumes that mean and variance are equal, i.e. mean = variance = λ.
- Not all the data satisfy this assumption.
- For example, if you model dependency of terror attacks on political regime type, you might suspect that democracies are targeted more frequently.
- However, attacks for democracies may be overdispersed, i.e. the variance exceeds the expected mean.
- In situations like this, Negative Binomial model is more useful than Poisson model.

Poisson vs. Negative Binomial II

- To test whether Negative Binomial is more appropriate than Poisson, one would use likelihood ratio test.
- Negative Binomial can be viewed as Poisson regression with an additional estimated overdispersion parameter.
- Hence, standard theory of likelihood ratio testing applies here.

Negative Binomial Interpretation

- Interpretation for Negative Binomial regression follows largely the same rules as for the Poisson regression.
- Only coefficient signs have intuitive interpretation.
- Using graphical approaches based on endpoint transformation or CLARIFY is advisable.

Introduction

- Well-known mass mobilization project dataset codes protest events in different states for years starting from 1990.
- The data was coded by human coders who read newspaper articles from Lexis-Nexis database and recorded a protest event with the date, in addition to characteristics of the event.
- Because the data utilized newspaper articles for protest coding, data generating process involves not only the fact of protest, but also reporting of the protest by a major newspaper.

Data Generating Process

- The example on the previous slides illustrates the common question in event history modelling: "Do zeroes mean what we think they mean?"
- For mass mobilization dataset, 0 can mean absence of protests, but it can also mean lack of the journalists' attention or highly nontransparent information environment in a certain state.
- Models like Poisson or Negative Binomial cannot take this idea into account.
- Zero Inflated Models were introduced to handle issues of this type.

Zero-Inflated Poisson Regression I

- Zero-Inflated Poisson regression assumes that the data generating process is governed by two rules.
- If event count = 0, this observation might have come from the Poisson distribution (no protests), or the event happened, but was not observed.
- If event count > 0, it is assumed that such observations come from the Poisson distribution.

Zero-Inflated Poisson Regression II

- Formally, let p_i be the probability that a protest happened, but was not observed, and $(1 - p_i)$ the probability that protest did not happen.
- Then probability of 0 corresponds to

$$p_i + (1 - p_i)\frac{\lambda(\mathbf{x}_i)^0 e^{-\lambda(\mathbf{x}_i)}}{0!} = p_i + (1 - p_i)e^{-\lambda(\mathbf{x}_i)}$$

- The probability of counts ≥ 1, on the other hand, is

$$(1 - p_i)\frac{\lambda(\mathbf{x}_i)^{c_i} e^{-\lambda(\mathbf{x}_i)}}{c_i!}$$

Likelihood Function I

- The likelihood function is given by

$$\mathcal{L} = \prod_{i=1}^{N} [p_i + (1-p_i)e^{-\lambda(\mathbf{x}_i)}]^{\delta(c_i)}[(1-p_i)\frac{\lambda(\mathbf{x}_i)^{c_i}e^{-\lambda(\mathbf{x}_i)}}{c_i!}]^{1-\delta(c_i)}$$

  where $\delta(c_i) = 1$ if $c_i = 0$.

- Once again, we need to take the log of this function in order to simplify the work for the machine.

Likelihood Function II

- Taking logs, one could obtain:

$$\log(\mathcal{L}) = \sum_{i=1}^{N}[\delta(c_i)\log(p_i+$$
$$(1-p_i)e^{-\lambda(X_i)}) + (1-\delta(c_i))(\log(1-p_i)+$$
$$c_i\log(\lambda(X_i)) - \lambda(X_i) - \log c_i!)]$$

- p_i is usually parameterized using logistic CDF; standard normal CDF is also sometimes employed.
- In essence, zero-inflated models are a mixture of binary response models and event count models.

Other Applications I

- Inflated models sometimes find surprising theoretically-driven applications.
- Miguel et.al. (2011) published an interesting piece on civil war exposure and violence.
- They used data on football players from European professional leagues to test the idea.

Other Applications II

- The main expectation is a straightforward one: exposure to an intense civil conflict makes people more prone to violent behavior – in football this will result in more red and yellow cards.
- There are multiple possible research designs here.
- The simplest design would count the total years of civil war since a specific minimum date (1980 in the paper) in a footballer's country of origin.
- More complicated design would count the total years of civil war in a footballer's country of origin since his birth date.

Other Applications III

- Even more complicated design would check footballer's biographies and match years where a footballer resided in a country with conflict years.
- Finally, if one could obtain detailed residency records for each footballer, one could try checking civil violence intensity in these residency regions.
- Paper found that, indeed, exposure to civil war violence increases yellow and red cards' count for footballers; this holds when accounting for various controls such as footballer's playing position, age, the number of games started, level of inequality and the rule of law in the country of origin.

Other Applications IV

- What if other players can learn about such behavior? In other words, what if somehow, looking at the behavior of players from countries experiencing civil conflict, other players would adopt an avoidance tactic?

- It makes sense: if you know that a certain player is prone to be more aggressive because of his background, you will avoid close-quarter contacts with him and try to keep the distance.

- This is the setup for a zero-inflated model!

Other Applications IV

- Idea can be extrapolated to other sports; for example, one can check whether exposure to crime increases the number of penalty minutes in hockey.
- In the US, Canada and Russia there are detailed crime records at the county level, so one could find the level of crime exposure for a player by averaging over counties where a player resided over his lifetime and then checking whether the number of penalty minutes or total number of penalties is associated with crime exposure.