Event count models

Silje Synnøve Lyder Hermansen

2023-03-23

The dependent variable

Count data

Count data is common in political science

Count data

Count data is common in political science

- Discrete: consists only in integers (0, 1, 2, . . . no digits)

Count data

Count data is common in political science

- Discrete: consists only in integers (0, 1, 2, . . . no digits)
- Bounded at zero, often long tail upwards.

Count data

Count data is common in political science

- Discrete: consists only in integers (0, 1, 2, ... no digits)
- Bounded at zero, often long tail upwards.

Count models: What are they good for?

When do we use count models?

The data generating process allows us to

When do we use count models?

The data generating process allows us to

- observe and count a number of events and

When do we use count models?

The data generating process allows us to

- observe and count a number of events and
- define a time frame or geographical space for the occurence(s)

When do we use count models?

The data generating process allows us to

- observe and count a number of events and
- define a time frame or geographical space for the occurence(s)

⇒ *e.g. goals in a soccer game, number of meetings between decision makers, violent events, legislative proposals, etc.*

Why not a binomial logistic regression?

These are indeed binary outcomes

Why not a binomial logistic regression?

These are indeed binary outcomes but we don't have information on the event level

Why not a binomial logistic regression?

These are indeed binary outcomes but we don't have information on the event level

$\Rightarrow$ *Variables are on the exposure level; related to when (where) the events took place.*

Why not OLS?

The variable could be approximated to a continuous measure but

Why not OLS?

The variable could be approximated to a continuous measure but

- it is bounded at zero, so predictions would be wrong

Why not OLS?

The variable could be approximated to a continuous measure but

- it is bounded at zero, so predictions would be wrong $\rightarrow$ *same problems as logit*

Why not OLS?

The variable could be approximated to a continuous measure but

- it is bounded at zero, so predictions would be wrong $\rightarrow$ *same problems as logit*
- it is skewed. Some people add a constant and logtransform: $log(y + 0.1)$

Why not OLS?

The variable could be approximated to a continuous measure but

- it is bounded at zero, so predictions would be wrong $\rightarrow$ *same problems as logit*
- it is skewed. Some people add a constant and logtransform: $log(y + 0.1) \rightarrow$ *but heteroskedasticity and non normal errors remain*

Why not OLS?

The variable could be approximated to a continuous measure but

- it is bounded at zero, so predictions would be wrong $\rightarrow$ *same problems as logit*
- it is skewed. Some people add a constant and logtransform: $log(y + 0.1) \rightarrow$ *but heteroskedasticity and non normal errors remain*

$\Rightarrow$ *We replace the normal distribution with another probability distribution*

The generalized linear model strategy

There are many count models

The generalized linear model strategy

There are many count models

- **Poisson model:** (aka the A-plan) the base-line

The generalized linear model strategy

There are many count models

- **Poisson model:** (aka the A-plan) the base-line
- **Other models:** (aka the B-plans) to address problems with the poisson

The generalized linear model strategy

There are many count models

- **Poisson model:** (aka the A-plan) the base-line
- **Other models:** (aka the B-plans) to address problems with the poisson
  - quasi-poisson
  - negative binomial
  - zero-inflated model
  - hurdle model

The Poisson model

Poisson process

The poisson distribution maps probabilities of events within a window to outcomes

Poisson process

The poisson distribution maps probabilities of events within a window to outcomes

- **Recoding** uses two pieces of information
  - **Exposure ($t, t + h$):** A window of opportunity between two bounaries (geographical or spacial)

Poisson process

The poisson distribution maps probabilities of events within a window to outcomes

- **Recoding** uses two pieces of information
  - **Exposure ($t, t + h$):** A window of opportunity between two bounaries (geographical or spacial)
  - **Probability of event (λ):** Simply the logtransformed mean of events within that window

Poisson process

The poisson distribution maps probabilities of events within a window to outcomes

- **Recoding** uses two pieces of information
  - **Exposure ($t, t + h$):** A window of opportunity between two bounaries (geographical or spacial)
  - **Probability of event (λ):** Simply the logtransformed mean of events within that window
    - Probability of event $= h\lambda$

Poisson process

The poisson distribution maps probabilities of events within a window to outcomes

- **Recoding** uses two pieces of information
  - **Exposure ($t, t + h$):** A window of opportunity between two bounaries (geographical or spacial)
  - **Probability of event (λ):** Simply the logtransformed mean of events within that window
    - Probability of event = $h\lambda$
    - Probability of no event = 1 - $h\lambda$

Poisson process

The poisson distribution maps probabilities of events within a window to outcomes

- **Recoding** uses two pieces of information
  - **Exposure ($t, t + h$):** A window of opportunity between two bounaries (geographical or spacial)
  - **Probability of event (λ):** Simply the logtransformed mean of events within that window
    - Probability of event $= h\lambda$
    - Probability of no event $=$ 1 - $h\lambda$
- **Poisson distribution** maps the latent variable to our observed values by calculating the probability of events. $y \sim Pois(\lambda)$

$\Rightarrow$ *Check out the [podcast] (http://leardigressions.com/episodes/2020/3/1/better-know-a-distribution-the-poisson-distribution)*

Formula

The equation the model estimates:

$$E(y_i) \equiv h\lambda_i = h \times exp(\alpha + \beta \times x_i) \tag{1}$$

Estimation of the exposure

What to do with the exposure parameter?

$$E(y_i) \equiv h\lambda_i = h \times exp(\alpha + \beta \times x_i) \quad (2)$$

Two strategies :

- **Offset:** Move it into the equation but constrain parameter: $exp(\alpha + \beta \times x_i + 1 \times log(h_i))$

Estimation of the exposure

What to do with the exposure parameter?

$$E(y_i) \equiv h\lambda_i = h \times exp(\alpha + \beta \times x_i) \tag{2}$$

Two strategies :

- **Offset:** Move it into the equation but constrain parameter: $exp(\alpha + \beta \times x_i + 1 \times log(h_i)) \rightarrow$ *we don't see it in the BUTON*

Estimation of the exposure

What to do with the exposure parameter?

$$E(y_i) \equiv h\lambda_i = h \times exp(\alpha + \beta \times x_i) \tag{2}$$

Two strategies :

- **Offset:** Move it into the equation but constrain parameter: $exp(\alpha + \beta \times x_i + 1 \times log(h_i)) \rightarrow$ *we don't see it in the BUTON*
- **Estimate a parameter:** $exp(\alpha + \beta_1 \times x_i + \beta_2 \times log(h_i))$

Estimation of the exposure

What to do with the exposure parameter?

$$E(y_i) \equiv h\lambda_i = h \times exp(\alpha + \beta \times x_i) \quad (2)$$

Two strategies :

- **Offset:** Move it into the equation but constrain parameter: $exp(\alpha + \beta \times x_i + 1 \times log(h_i)) \rightarrow$ *we don't see it in the BUTON*
- **Estimate a parameter:** $exp(\alpha + \beta_1 \times x_i + \beta_2 \times log(h_i))$

$\Rightarrow$ *If the exposure is the same for all units, we set it to 1 and ignore it.*

Interpretation: back and forth

Interpretation is relatively easy with all count models

Interpretation: back and forth

Interpretation is relatively easy with all count models

- Recoding (for estimation): we logtransform the mean of the y (within x-values)

Interpretation: back and forth

Interpretation is relatively easy with all count models

- Recoding (for estimation): we logtransform the mean of the y (within x-values)
- We back-transform (for interpretation): $exp(\lambda)$ is simply an approximation (with digits) of our counts!

Interpretation: back and forth

Interpretation is relatively easy with all count models

- Recoding (for estimation): we logtransform the mean of the y (within x-values)
- We back-transform (for interpretation): $exp(\lambda)$ is simply an approximation (with digits) of our counts!

Interpretation: effects

Interpretation is relatively easy

Interpretation: effects

Interpretation is relatively easy

- **Recoding (for estimation):** we logtransform the mean of the y (within x-values)

Interpretation: effects

Interpretation is relatively easy

- **Recoding (for estimation):** we logtransform the mean of the y (within x-values)
- **Back-transform** (for interpretation):
  - Predicted value: $exp(\hat{\lambda})$ is simply an approximation (with digits) of our counts

Interpretation: effects

Interpretation is relatively easy

- **Recoding (for estimation):** we logtransform the mean of the *y* (within x-values)
- **Back-transform** (for interpretation):
  - Predicted value: $exp(\hat{\lambda})$ is simply an approximation (with digits) of our counts
  - Marginal effect of β: $exp(\beta)$ is multiplicative of predicted $\hat{\lambda}$

Interpretation: effects

Interpretation is relatively easy

- **Recoding (for estimation):** we logtransform the mean of the *y* (within x-values)
- **Back-transform** (for interpretation):
  - Predicted value: $exp(\hat{\lambda})$ is simply an approximation (with digits) of our counts
  - Marginal effect of β: $exp(\beta)$ is multiplicative of predicted $\hat{\lambda} \rightarrow$ easy!
  - NB: the log-transformation means that the model has an inbuilt interaction effect, so all effects are proportional

$\Rightarrow$ *Make scenarios, predict, knock yourself out*

Dispersion

The main assumption of the Poisson model

The model assumes equidispersion: The spread equals the mean

- The y can be overdispersed, but not the $\hat{\lambda}$

The main assumption of the Poisson model

The model assumes equidispersion: The spread equals the mean

- The y can be overdispersed, but not the $\hat{\lambda}$ $\rightarrow$ *as in OLS*

The main assumption of the Poisson model

The model assumes equidispersion: The spread equals the mean

- The y can be overdispersed, but not the $\hat{\lambda} \rightarrow$ *as in OLS*

$\Rightarrow$ *The standerd errors will be too small*

Identifying overdispersion

- Formal tests: Using residuals and significance tests.

Identifying overdispersion

- Formal tests: Using residuals and significance tests.
- Poissonness plot

Identifying overdispersion

- Formal tests: Using residuals and significance tests.
- Poissonness plot
- Rootograms

Identifying overdispersion

- Formal tests: Using residuals and significance tests.
- Poissonness plot
- Rootograms

Visual identification of overdispersion

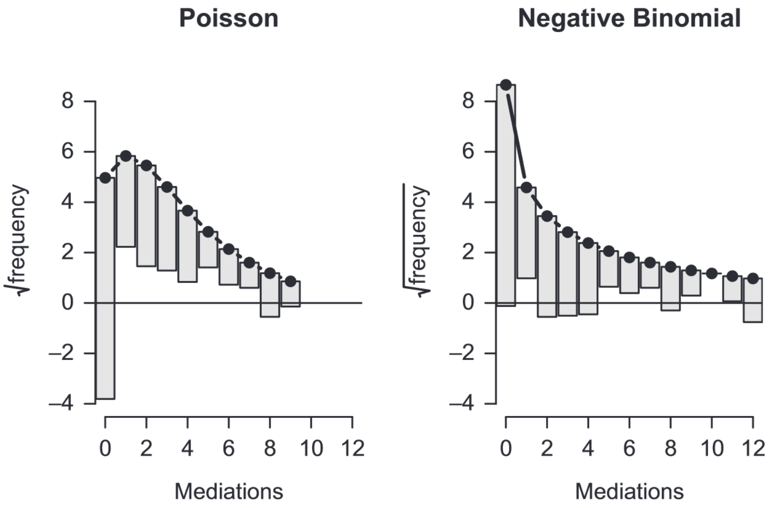

Figure 1: Hanging rootograms

Reasons for overdispersion

The poisson process assumes that each event (count) is independent. Overdispersion may be a symptom that your events are related

- Address the symptoms, not the problem:
  - Unexplained (random) variance → quasipoisson
  - Unexplained change in likelihood/functional form → negative binomial model
- Address the problem
  - Poor choice of variables → include more (also random intercepts)
  - Lack of exposure time
- Address the symptoms and – potentially – the problem:
  - Too many zeros
    - Two separate processes generate our data → zeroinflated
    - We need to arrive at a certain threshold before positive counts arrive → hurdle

Adressing overdispersion

The quasi-poisson model

- Adds an additional parameter, ϕ, to the variance estimation

The quasi-poisson model

- Adds an additional parameter, ϕ, to the variance estimation $\rightarrow$ *similar to robust standard errors*

The quasi-poisson model

- Adds an additional parameter, ϕ, to the variance estimation $\rightarrow$ *similar to robust standard errors*

$\Rightarrow$ β *remains the same, standard errors are larger*

The negative binomial model

The event is in fact generated by two processes

The negative binomial model

The event is in fact generated by two processes

- $\lambda_i = exp(\beta \times x_i + 1 \times u_i)$
- $v = exp(u_i)$ is in itself generated by a gamma distribution $v_i \sim f\Gamma(\alpha)$
- The latent variable is manipulated directly: the rate increases over y

The negative binomial model

The event is in fact generated by two processes

- $\lambda_i = exp(\beta \times x_i + 1 \times u_i)$
- $v = exp(u_i)$ is in itself generated by a gamma distribution $v_i \sim f\Gamma(\alpha)$
- The latent variable is manipulated directly: the rate increases over y

Exess zeros

Substantially that two data generating processes are at work.

Exess zeros

Substantially that two data generating processes are at work.

- One producing zeros

Exess zeros

Substantially that two data generating processes are at work.

- One producing zeros
- One producing (at least some) positive counts

Exess zeros

Substantially that two data generating processes are at work.

- One producing zeros
- One producing (at least some) positive counts

⇒ *We can model this in two parallel regressions with possibly different x or just an additional intercept.*

Hurdle models

Observations have a higher hurdle/threshold/distance to pass in order to obtain a positive count (from 0 to 1) than between positive counts (1 to 2, 2 to 3, etc)

Hurdle models

Observations have a higher hurdle/threshold/distance to pass in order to obtain a positive count (from 0 to 1) than between positive counts (1 to 2, 2 to 3, etc)

- Hurdle part: A binomial logit where success is $y > 0$

Hurdle models

Observations have a higher hurdle/threshold/distance to pass in order to obtain a positive count (from 0 to 1) than between positive counts (1 to 2, 2 to 3, etc)

- Hurdle part: A binomial logit where success is $y > 0$
- Count model: A zero-truncated poisson (or negative binomial) on all the positive counts.

Hurdle models

Observations have a higher hurdle/threshold/distance to pass in order to obtain a positive count (from 0 to 1) than between positive counts (1 to 2, 2 to 3, etc)

- Hurdle part: A binomial logit where success is $y > 0$
- Count model: A zero-truncated poisson (or negative binomial) on all the positive counts.

$\Rightarrow$ *Can accomodate under-dispersion too.*

Zero-inflated models

There are two sources of zeros, but only one of positive counts.

Zero-inflated models

There are two sources of zeros, but only one of positive counts.

- Zero-inflated part: A binomial logit where success is the “always zeros”.

Zero-inflated models

There are two sources of zeros, but only one of positive counts.

- Zero-inflated part: A binomial logit where success is the “always zeros”.
- Count model: A poisson or negative binomial that is not truncated.

Zero-inflated models

There are two sources of zeros, but only one of positive counts.

- Zero-inflated part: A binomial logit where success is the “always zeros”.
- Count model: A poisson or negative binomial that is not truncated.

⇒ *functions as a switch that is turned on/off after a threshold. The observation is then passed to the count-model group.*

Recap on GLMs

What are the criteria for model selection?

You can think of model selection as a set of criteria that should be met

Try out the model selection decision tree to see my mental map!

https://siljehermansen.github.io/teaching/choose_glm/