CSSS/POLS 510 Maximum Likelihood Estimation: Lab 8

Count Data

Kenya Amano

2020-12-4

Agenda

1. Count Data
2. Closing

1. Recap

Where are we at right now?

1. Learn distribution and MLE $\rightarrow$ HW1 & HW2

2. Logit model $\rightarrow$ HW3

3. Ordered Probit model $\rightarrow$ HW4

4. Multinomial logit $\rightarrow$ HW5

5. Count data $\rightarrow$ HW5

2. Count Data

Review the lecture materials to understand the concept

A TAXONOMY OF COUNT REGRESSION MODELS

Binomial

Beta-Binomial

Quasibinomial

Hurdle Binomial
Hurdle Beta-Binomial

Zero-inflated Binomial
Zero-inflated Beta-Binomial

Poisson

Negative Binomial

Quasipoisson

Hurdle Poisson
Hurdle Negative Binomial

Zero-inflated Poisson
Zero-inflated Negative Binomial

the models we've discussed imply other models

A TAXONOMY OF COUNT REGRESSION MODELS

Binomial	Poisson
Beta-Binomial	Negative Binomial
Quasibinomial	Quasipoisson
Hurdle Binomial	Hurdle Poisson
Hurdle Beta-Binomial	Hurdle Negative Binomial
Zero-inflated Binomial	Zero-inflated Poisson
Zero-inflated Beta-Binomial	Zero-inflated Negative Binomial

models that assume a maximum count

A TAXONOMY OF COUNT REGRESSION MODELS

Binomial	Poisson
Beta-Binomial	Negative Binomial
Quasibinomial	Quasipoisson
Hurdle Binomial Hurdle Beta-Binomial	Hurdle Poisson Hurdle Negative Binomial
Zero-inflated Binomial Zero-inflated Beta-Binomial	Zero-inflated Poisson Zero-inflated Negative Binomial

models that allow for unbounded counts

A TAXONOMY OF COUNT REGRESSION MODELS

Binomial	Poisson
Beta-Binomial	Negative Binomial
Quasibinomial	Quasipoisson
Hurdle Binomial	Hurdle Poisson
Hurdle Beta-Binomial	Hurdle Negative Binomial
Zero-inflated Binomial	Zero-inflated Poisson
Zero-inflated Beta-Binomial	Zero-inflated Negative Binomial

models that assume independent events

A TAXONOMY OF COUNT REGRESSION MODELS

Binomial	Poisson
Beta-Binomial	Negative Binomial
Quasibinomial	Quasipoisson
Hurdle Binomial	Hurdle Poisson
Hurdle Beta-Binomial	Hurdle Negative Binomial
Zero-inflated Binomial	Zero-inflated Poisson
Zero-inflated Beta-Binomial	Zero-inflated Negative Binomial

models that allow events to be correlated

A TAXONOMY OF COUNT REGRESSION MODELS

Binomial

Beta-Binomial

Quasibinomial

Hurdle Binomial
Hurdle Beta-Binomial

Zero-inflated Binomial
Zero-inflated Beta-Binomial

Poisson

Negative Binomial

Quasipoisson

Hurdle Poisson
Hurdle Negative Binomial

Zero-inflated Poisson
Zero-inflated Negative Binomial

models that avoid distributional assumptions

A TAXONOMY OF COUNT REGRESSION MODELS

Binomial

Beta-Binomial

Quasibinomial

Hurdle Binomial
Hurdle Beta-Binomial

Zero-inflated Binomial
Zero-inflated Beta-Binomial

Poisson

Negative Binomial

Quasipoisson

Hurdle Poisson
Hurdle Negative Binomial

Zero-inflated Poisson
Zero-inflated Negative Binomial

models that allow higher barriers to the initial event

A TAXONOMY OF COUNT REGRESSION MODELS

Binomial	Poisson
Beta-Binomial	Negative Binomial
Quasibinomial	Quasipoisson
Hurdle Binomial	Hurdle Poisson
Hurdle Beta-Binomial	Hurdle Negative Binomial
Zero-inflated Binomial	Zero-inflated Poisson
Zero-inflated Beta-Binomial	Zero-inflated Negative Binomial

models that allow some cases to be structural zeroes

A TAXONOMY OF COUNT REGRESSION MODELS

Binomial	Poisson
Beta-Binomial	Negative Binomial
Quasibinomial	Quasipoisson
Hurdle Binomial	Hurdle Poisson
Hurdle Beta-Binomial	Hurdle Negative Binomial
Zero-inflated Binomial	Zero-inflated Poisson
Zero-inflated Beta-Binomial	Zero-inflated Negative Binomial

models you might actually use for social science data

3. Last words

1. Statistics (and programming) should be intuitive
   - If I can't explain something in a simple manner, I don't understand it. e.g. coin flip
   - Always go back to first principles and simple analogies
   - Statistics is a tool; your reserach design is first
   - You run the model; don't let the model run you

3. Last words

2. Computers are powerful yet stupid
   - They execute what you instruct them, *literally*
   - When mistakes happen, it is usually us who make mistakes
   - No replacement of sound statistical judgement
   - Don't be held hostage to particular functions or packages
   - "No default, all manual" is a virtue of `simcf` and `tile`
   - Run incrementally when you face new `loop` and `function`: Reading ability is also critical

3. Last words

3. Simulations will be your best friends
    - Understand the assumed DGP
    - Solve probability problems
    - Evaluate estimators
    - Transform statistical results into QoI

Simulating QoI

1. Estimate: MLE $\hat{\beta}_{(M+1)\times(P+1)}$ and its variance $\hat{V}(\hat{\beta}_{(M+1)\times(P+1)})$
   $\rightarrow$ `optim()`, `multinom()`
2. Simulate estimation uncertainty from a multivariate normal distribution:
   Draw $\tilde{\beta} \sim MVN\left[\hat{\beta}, \hat{V}(\hat{\beta})\right]$
   $\rightarrow$ `MASS::mvrnorm()`
3. Create hypothetical scenarios of your substantive interest:
   Choose valuese of X: X_c
   $\rightarrow$ `simcf::cfmake()`, `cfchange()` ...

Simulating QoI

4. Calculate expected values:
   $\tilde{\pi_c} = g(X_c, \tilde{\beta})$
5. Compute EVs, First Differences or Relative Risks
   EV: $\mathbb{E}(y = j | X_{c1}, \tilde{\beta})$
   $\rightarrow$ `simcf::mlogitsimev()` ...
   FD: $\mathbb{E}(y = j | X_{c2}, \tilde{\beta},) - \mathbb{E}(y = j | X_{c1}, \tilde{\beta})$
   $\rightarrow$ `simcf::mlogitsimfd()` ...
   RR: $\frac{\mathbb{E}(y=j|X_{c2},\tilde{\beta})}{\mathbb{E}(y=j|X_{c1},\tilde{\beta})}$
   $\rightarrow$ `simcf::mlogitsimrr()` ...

3. Last words

4. Model results are unintelligible unless...

- You interpret and communicate them in meaningful ways
- *Substantively meaningful* quantities of interest (QoI) and counterfactual scenarios
- Visualization is critical (CS&SS 569)
  1) Avoid WYGWYS (What you get is what you see)
  2) LaTeX

3. Homework 5 and Feedback

- Due on Dec 10
- Email subject: **MLE510HW5**
- File name: **MLE510HW5KenyaAmano**
- *slack chanel: #hw5
- Evaluation URL:https://uw.iasystem.org/survey/231765