CSSS/POLS 510 Maximum Likelihood Estimation: Lab 7

Multinomial Logistic Regression

Kenya Amano

2020-11-20

Agenda

1. Recap lectures
2. Multinomial Logistic Regression
3. Homework: HW4 & HW5

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. Multinomial logit model

Review the lecture materials to understand the concept

Notation for Nominal Regression Models

Welcome to subscript hell

Observations: $1, \ldots i, \ldots N$

Unordered Categories: $1, \ldots j, \ldots M$

Category 1 will be the "reference category",
so we will often speak of the remaining categories $2, \ldots M$

Covariates: $x_1, \ldots x_k, \ldots x_P$

Parameters: There may be a parameter for each combination of a non-reference category and a covariate, plus an intercept for non-reference category

That is, the systematic component for the jth category of the ith observation is

$$\mu_{ij} = \beta_{j0} + \sum_{k=1}^{P} \beta_{jk} \times x_{ik}$$

Note the lack of a j subscript on x . . . this will change later

Notation for Nominal Regression Models

For comparison:

in ordinary logit, we estimate $P + 1$ parameters

in ordered probit, we estimate $P + M - 1$ parameters

in multinomial logit, we estimate $(M - 1) \times (P + 1)$ parameters

$\rightarrow$ MNL gobbles up degrees of freedom fast

$\rightarrow$ Compared to ordinary logit, no big deal (we have more data)

$\rightarrow$ Compared to ordered probit, or linear regression, MNL is “expensive”

If the assumptions of ordered probit fit your data, use it

But the assumptions don’t always fit . . .

Likelihood for Multinomial Logit (MNL)

By definition, the likelihood is proportional to the probability, p_{ij}, of observing the value of y that is ultimately observed (the probabilities for all unobserved categories are irrelevant to the likelihood)

$$\mathcal{L}(\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M | \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{N} \prod_{j=1}^{M} {p_{ij}}^{y_{ij}}$$

Substituting for p_{ij}, we have

$$\mathcal{L}(\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M | \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{N} \prod_{j=1}^{M} \left[\frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_j)}{1 + \sum_{\ell=2}^{M} \exp(\mathbf{x}_i \boldsymbol{\beta}_\ell)} \right]^{y_{ij}}$$

Taking logs, we end up with

$$\log \mathcal{L}(\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M | \mathbf{y}, \mathbf{X}) = \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} \log \frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_j)}{1 + \sum_{\ell=2}^{M} \exp(\mathbf{x}_i \boldsymbol{\beta}_\ell)}$$

which we can maximize with `optim()`

Calculating Expected Values in MNL

After estimating an MNL model, calculating expected probabilities is straightforward.

For a given counterfactual level of the explanatory variables, $\mathbf{x}_c$, and a three category multinomial logit, we have

$$\Pr(y = 1 | \mathbf{x}_c, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) = \frac{1}{1 + \exp(\mathbf{x}_c \hat{\boldsymbol{\beta}}_2) + \exp(\mathbf{x}_c \hat{\boldsymbol{\beta}}_3)}$$

$$\Pr(y = 2 | \mathbf{x}_c, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) = \frac{\exp(\mathbf{x}_c \hat{\boldsymbol{\beta}}_2)}{1 + \exp(\mathbf{x}_c \hat{\boldsymbol{\beta}}_2) + \exp(\mathbf{x}_c \hat{\boldsymbol{\beta}}_3)}$$

$$\Pr(y = 3 | \mathbf{x}_c, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) = \frac{\exp(\mathbf{x}_c \hat{\boldsymbol{\beta}}_3)}{1 + \exp(\mathbf{x}_c \hat{\boldsymbol{\beta}}_2) + \exp(\mathbf{x}_c \hat{\boldsymbol{\beta}}_3)}$$

Simulating from the MNL is also simple:

Just draw the $\hat{\boldsymbol{\beta}}$'s from the multivariate normal,

then plug them into the above equations to get a matrix of simulates

Notation for Nominal Regression Models

And each column corresponds to a set of parameters for the same covariate:

$$\boldsymbol{\beta} = \beta_{jk} = \begin{pmatrix} \beta_{10} & \beta_{11} & \ldots & \beta_{1k} & \ldots & \beta_{1P} \\ \beta_{20} & \beta_{21} & \ldots & \beta_{2k} & \ldots & \beta_{2P} \\ \vdots & & \ddots & & & \vdots \\ \beta_{j0} & & & \beta_{jk} & & \beta_{jP} \\ \vdots & & & & \ddots & \vdots \\ \beta_{M0} & \beta_{M1} & \ldots & \beta_{Mk} & \ldots & \beta_{MP} \end{pmatrix}$$

Notation for Nominal Regression Models

For identification, we set one row of coefficients to 0:

$$\boldsymbol{\beta} = \beta_{jk} = \begin{pmatrix} 0 & 0 & \dots & 0 & \dots & 0 \\ \beta_{20} & \beta_{21} & \dots & \beta_{2k} & \dots & \beta_{2P} \\ \vdots & & \ddots & & & \vdots \\ \beta_{j0} & & & \beta_{jk} & & \beta_{jP} \\ \vdots & & & & \ddots & \vdots \\ \beta_{M0} & \beta_{M1} & \dots & \beta_{Mk} & \dots & \beta_{MP} \end{pmatrix}$$

That is, we treat one category (here, “1”) as the *reference category*

That leaves $(M-1) \times (P+1)$ parameters to estimate

2 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()...`

2 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. Homework: Question HW4

- Due on Nov 24
- Email subject: **MLE510HW4**
- File name: **MLE510HW4KenyaAmano**
- *slack chanel: #hw4
- One common problem when knitting: the math mode environment doesn't like white space or empty line + Try `\begin{alinged}` instead of `\begin{split}` + `R Markdown` guide is here