Econometrics 2 (2018) TA session 5*

Kenji Hatakenaka †

8 November 2018

Aim(目的)

- Review the previous lectures with some additional contents.
- Introduce an empirical example of the Poisson Model.

1 Count Data Model

As far, $y_i = 0$ or 1. However, sometimes we are interested in how many times an event occurs (e.g. traffic accidents). In this section, $y_i = 0, 1, 2, \ldots$.

1.1 Poisson Count Data Model

Consider the following Poisson distribution:

$$P(X = x) = f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for} \quad x = 0, 1, 2, \ldots. \tag{1}$$

This Poisson distribution is obtained as a limit of binomial distribution(二項分布). Consider that an event happens with probability $p = \lambda/n$, and we consider the probability of the number of the event happens as follows:

$$P(X = x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} = \frac{n!}{(n-x)!x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}. \tag{2}$$

* All comments welcome!

† E-mail: u626530i@ecs.osaka-u.ac.jp, Room 501

Then, we can transform the equation as follows:

$$P(X=x) = \frac{\overbrace{n \times (n-1) \times \cdots \times (n-x+1)}^{x \text{ times multiplied}}}{x!} \frac{\lambda^x}{n^x} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \tag{3}$$

$$= \frac{n}{n}\frac{(n-1)}{n}\frac{(n-2)}{n}\cdots\frac{(n-x+1)}{n}\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x}. \tag{4}$$

By taking limit with n, with sustaining $p = \lambda/n$, we obtain

$$\lim_{n\to\infty} P(X=x) = \lim_{n\to\infty} \underbrace{\frac{n}{n}}_{1} \underbrace{\frac{(n-1)}{n}}_{1} \cdots \underbrace{\frac{(n-x+1)}{n}}_{1} \frac{\lambda^x}{x!} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{e^{-\lambda}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{1}. \tag{5}$$

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{6}$$

Expectation of this distribution is

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{\infty} \lambda \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda \underbrace{\sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}}_{1} = \lambda. \tag{7}$$

Variance is

$$V(X) = E(X^2) - E(X)^2 = \lambda(\lambda+1) - \lambda^2 = \lambda. \tag{8}$$

This is because the second moment of X is calculated as follows:

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} = \lambda \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda \left(\sum_{x^*=0}^{\infty} (x^*+1) \frac{e^{-\lambda}\lambda^{x^*}}{x^*!} \right)$$
$$= \lambda E[x^*+1] = \lambda(\lambda+1). \tag{9}$$

1.1.1 Model

$y_i \in \{0, 1, 2, \cdots\}$ (discrete numbers) and $y_i \sim Poi(\lambda)$. Poisson count data model is represented as

$$E[y_i] = \lambda_i = \exp(X_i\beta) \tag{10}$$

where $\lambda_i > 0$. ($X_i\beta$ can be negative. So you should avoid the specification $\lambda = X_i\beta$.) To estimate β, we use MLE. Likelihood function is

$$L(\beta) = f(y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \tag{11}$$

where $\lambda_i = \exp(X_i\beta)$. By taking natural logarithm, we obtain the following equation:

$$\log L(\beta) = -\sum_{i=1}^{n} \lambda_i + \sum_{i=1}^{n} y_i \log \lambda_i - \sum_{i=1}^{n} \log(y_i!) \tag{12}$$

$$= -\sum_{i=1}^{n} \exp(X_i\beta) + \sum_{i=1}^{n} y_i X_i \beta - \sum_{i=1}^{n} \log(y_i!) \tag{13}$$

Then, by differentiating with β, we obtain

$$\frac{\partial \log L(\beta)}{\partial \beta} = -\sum_{i=1}^{n} X_i' \exp(X_i\beta) + \sum_{i=1}^{n} X_i' y_i = 0. \tag{14}$$

Obtaining β is difficult. Hence, we use Newton Raphson method or method of Scoring. The procedure is shown below:

$$\beta^{(j+1)} = \beta^{(j)} - \left(-\sum_{i=1}^{n} X_i' X_i \exp(X_i\beta^{(j)})\right)^{-1} \left(-\sum_{i=1}^{n} X_i' \exp(X_i\beta^{(j)}) + \sum_{i=1}^{n} X_i' y_i\right) \tag{15}$$

1.1.2 Review: Non-linear Optimization Procedure

Note that the Newton-Raphson method (one of the non-linear optimization procedure) is described as follows:

$$\beta^{(j+1)} = \beta^{(j)} - \left(\frac{\partial^2 \log L(\beta^{(j)})}{\partial \beta \partial \beta'}\right)^{-1} \frac{\partial \log L(\beta^{(j)})}{\partial \beta}$$

This equation comes from the first-order Taylor series expansion around $\beta = \beta^*$:

$$0 = \frac{\partial \log L(\beta)}{\partial \beta} \approx \frac{\partial \log L(\beta^*)}{\partial \beta} + \frac{\partial^2 \log L(\beta^*)}{\partial \beta \partial \beta'}(\beta - \beta^*)$$

Then we obtain

$$\frac{\partial^2 \log L(\beta^*)}{\partial\beta\partial\beta'}(\beta - \beta^*) = -\frac{\partial \log L(\beta^*)}{\partial\beta}$$

$$\beta - \beta^* = -\left(\frac{\partial^2 \log L(\beta^*)}{\partial\beta\partial\beta'}\right)^{-1} \frac{\partial \log L(\beta^*)}{\partial\beta}$$

This yields the above equation. If we take expectation on second derivative of likelihood function, the method is known as the method of Scoring(スコア法).

1.2 Example

Here, we supply an example of Poisson regression from Stata 13 manual.

In a famous age-specific study of coronary disease deaths among male British doctors, Doll and Hill (1966) reported the following data (reprinted in Rothman, Greenland, and Lash [2008, 264]):

Smokers			Nonsmokers		
Age	Deaths	person-years	Age	Deaths	person-years
35–44	32	52407	35–44	2	18790
45–54	104	43248	45–54	12	10673
55–64	206	28612	55–64	28	5710
65–74	186	12663	65–74	28	2585
75–84	102	5317	75–84	31	1462

We can use this dataset by typing the following command:

`use http://www.stata-press.com/data/r13/dollhill3, clear`

We conduct a Poisson regression the model shown below by the following command:

`poisson deaths smokes i.agecat, exposure(pyears) irr`

The model is

$$E[Deaths_j] = EXP_j \exp(I_j^*\beta^* + \beta_0 + I_j^{45}\beta_1 + I_j^{55}\beta_2 + I_j^{65}\beta_3 + I_j^{75}\beta_4) \quad (16)$$

$$= \exp(\ln EXP_j + I_j^*\beta^* + \beta_0 + I_j^{45}\beta_1 + I_j^{55}\beta_2 + I_j^{65}\beta_3 + I_j^{75}\beta_4) \quad (17)$$

$$I_j^* = \begin{cases} 1 & \text{if } j \text{ is smoke} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

where $Deaths_j$ is number of deaths, EXP_j is exposure term, and I_j^k is indicator function if Age is between k and $k+9$, 1; otherwise 0. The result is below.

```
. poisson deaths smokes i.agecat, exposure(pyears) irr

Iteration 0:   log likelihood = -33.823284
Iteration 1:   log likelihood = -33.600471
Iteration 2:   log likelihood = -33.600153
Iteration 3:   log likelihood = -33.600153

Poisson regression                              Number of obs   =         10
                                                LR chi2(5)      =     922.93
                                                Prob > chi2     =     0.0000
Log likelihood = -33.600153                     Pseudo R2       =     0.9321

------------------------------------------------------------------------------
      deaths |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smokes |   1.425519   .1530638     3.30   0.001     1.154984    1.759421
             |
      agecat |
      45-54  |   4.410584   .8605197     7.61   0.000     3.009011    6.464997
      55-64  |    13.8392   2.542638    14.30   0.000     9.654328    19.83809
      65-74  |   28.51678   5.269878    18.13   0.000     19.85177    40.96395
      75-84  |   40.45121   7.775511    19.25   0.000     27.75326    58.95885
             |
       _cons |   .0003636   .0000697   -41.30   0.000     .0002497    .0005296
  ln(pyears) |          1  (exposure)
------------------------------------------------------------------------------
```

1.3 Zero Inflated Poisson Count Data Model

The model is the same so far:

$$E[y_i] = \lambda_i = \exp(X_i\beta),$$

but we observe too many zeros. We have to modify the estimation procedure because the estimation often becomes bad. So we assume that there are two regimes below:

- **Regime 1**: $y_i = 0 \quad w.p.1$
  Regime1 is chosen with probability "$F(Z_i\alpha)$"
- **Regime 2**: $y_i \sim Poi(\lambda_i)$
  Regime2 is chosen with probability "$1 - F(Z_i\alpha)$".

So the probability of observing $y_i = j$ is

$$P(y_i = 0) = P(y_i = 0 \mid R_1)P(R_1) + P(y_i = 0 \mid R_2)P(R_2),$$
$$P(y_i = j) = P(y_i = j \mid R_1)P(R_1) + P(y_i = j \mid R_2)P(R_2).$$

This can be rewritten as

$$P(y_i = j) = P(R_1)I_i + P(y_i = j \mid R_2)P(R_2)$$

From some assumption above,

$$P(R_1) = F(Z_i\alpha)$$
$$P(R_2) = 1 - F(Z_i\alpha)$$
$$P(y_i = j|R_2) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$$

So probability function can be rewritten as

$$P(y_i = j) = F(Z_i\alpha)I_{y_i=0} + \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}(1 - F(Z_i\alpha)), \quad j = 0, 1, 2, \cdots .$$

where Likelihood function is

$$\log L(\alpha, \beta) = \sum_{i=1}^{n} \log P(y_i = j) = \sum_{i=1}^{n} \log \left(F(Z_i\alpha)I_{y_i=0} + \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}(1 - F(Z_i\alpha)) \right)$$

This log-likelihood function is maximized with respect to α and β by using Newton-Raphson Method or Method of Scoring.

2 Summary of Qualitative Dependent Variable

Here, we shortly summarize what we have learned by now. When the dependent variable is discrete, the regular method (OLS) is not adequate because the prediction does not work. In most cases, estimation is conducted by maximum likelihood method (最尤法) (including numerical non-linear estimation).

- Discrete Choice Model(離散選択モデル): Choices are discrete.
  - **Binary Choice Model**(二値選択モデル):

$$y_i = \begin{cases} 1 & y_i^* > 0 \\ 0 & y_i^* \leq 0 \end{cases} \tag{19}$$

    Choices are restricted to 0 and 1.

    → we have Probit and Logit Model. The difference is distribution function.
  - **Ordered Probit or Logit Model**:

$$y_i = \begin{cases} 1, & \text{if } -\infty < y_i^* \leq a_1 \\ 2, & \text{if } a_1 < y_i^* \leq a_2 \\ \vdots & \\ m, & \text{if } a_{m-1} < y_i^* < \infty \end{cases} \tag{20}$$

- **Multinominal Logit Model**: Choices are restricted to 0 to m.
  y_i isn't ordered.
- **Nested Logit Model**: Choices have nested (tree) structure

- Limited Dependent Variable Model(制限従属変数モデル): unobservable dependent variable or exact value of dependent variable is unknown
  - **Truncated Regression Model**:

$$y_i = X_i\beta + u_i, \quad u_i \sim N(0, \sigma^2) \quad \text{when } y_i > a, \tag{21}$$

    $\rightarrow$ we have to use $f(y_i|y_i > a)$ to make likelihood function.
  - **Censored Regression Model or Tobit Model**: When the dependent variable is censored: we can know the existence of observation but we cannot know exact value of it.
    $\rightarrow$ we have to consider the fully observable probability

$$y_i = \begin{cases} X_i\beta, & \text{if } y_i > a \\ a, & \text{otherwise} \end{cases} \tag{22}$$

    Likelihood function is

$$L(\beta) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} f(y_i)^{I(y_i > a)} P(y_i = a)^{I(y_i = a)}. \tag{23}$$

- Count Data Model(計数データモデル):

$$y_i \in \{0, 1, 2, \cdots\}, \quad y_i \sim Poi(\lambda_i) \tag{24}$$

  Dependent variable is count data $(0, 1, \cdots)$
  - **Poisson Model**: dependent variable counts rare event.
  - **Zero Inflated Poisson Model**: dependent variable counts rare event and contains **too much zeros**.