7.3 Count Data Model (計数データモデル)

Poisson distribution:

$$\mathrm{P}(X = x) = f(x) = \frac{e^{-\lambda}\lambda^x}{x!},$$

for $x = 0, 1, 2, \cdots$.

In the case of Poisson random variable X, the expectation of X is:

$$\mathrm{E}(X) = \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} \lambda\frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda\sum_{x'=0}^{\infty} \frac{e^{-\lambda}\lambda^{x'}}{x'!} = \lambda.$$

Remember that $\sum_x f(x) = 1$, i.e., $\sum_{x=0}^{\infty} e^{-\lambda}\lambda^x/x! = 1$.

Therefore, the probability function of the count data y_i is taken as the Poisson distribution with parameter λ_i.

In the case where the explained variable y_i takes 0, 1, 2, $\cdots$ (discrete numbers), assuming that the distribution of y_i is Poisson, the logarithm of λ_i is specified as a linear function, i.e.,

$$\mathrm{E}(y_i) = \lambda_i = \exp(X_i\beta).$$

Note that λ_i should be positive.

Therefore, it is better to avoid the specification: $\lambda = X_i\beta$.

The joint distribution of $y_1, y_2, \cdots, y_n$ is:

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

where $\lambda_i = \exp(X_i\beta)$.

The log-likelihood function is:

$$\begin{aligned}\log L(\beta) &= -\sum_{i=1}^{n} \lambda_i + \sum_{i=1}^{n} y_i \log \lambda_i - \sum_{i=1}^{n} y_i! \\ &= -\sum_{i=1}^{n} \exp(X_i\beta) + \sum_{i=1}^{n} y_i X_i\beta - \sum_{i=1}^{n} y_i!.\end{aligned}$$

The first-order condition is:

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

$\Longrightarrow$ Nonlinear optimization procedure

[Review] Nonlinear Optimization Procedures:

Note that the Newton-Raphson method (one of the nonlinear optimization procedures) is:

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

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

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

and β and β^* are replaced by $\beta^{(j+1)}$ and $\beta^{(j)}$, respectively.

An alternative nonlinear optimization procedure is known as the method of scoring, which is shown as:

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

where $\left(\dfrac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'}\right)$ is replaced by $\mathrm{E}\left(\dfrac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'}\right)$.

[End of Review]

In this case, we have the following iterative procedure:

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

The Newton-Raphson method is equivalent to the scoring method in this count model, because any random variable is not included in the expectation.

Zero-Inflated Poisson Count Data Model: In the case of too many zeros, we have to modify the estimation procedure.

Suppose that the probability of $y_i = j$ is decomposed of two regimes.

$\longrightarrow$ We have the case of $y_i = j$ and Regime 1, and that of $y_i = j$ and Regime 2.

Consider $P(y_i = 0)$ and $P(y_i = j)$ separately as follows:

$$P(y_i = 0) = P(y_i = 0|\text{Regime 1})P(\text{Regime 1}) + P(y_i = 0|\text{Regime 2})P(\text{Regime 2})$$

$$P(y_i = j) = P(y_i = j|\text{Regime 1})P(\text{Regime 1}) + P(y_i = j|\text{Regime 2})P(\text{Regime 2}),$$

for $j = 1, 2, \cdots$.

Assume:

- $P(y_i = 0|\text{Regime 1}) = 1$ and $P(y_i = j|\text{Regime 1}) = 0$ for $j = 1, 2, \cdots,$

- $P(\text{Regime 1}) = F_i$ and $P(\text{Regime 2}) = 1 - F_i,$

- $P(y_i = j|\text{Regime 2}) = \dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$ for $j = 0, 1, 2, \cdots,$

where $F_i = F(Z_i\alpha)$ and $\lambda_i = \exp(X_i\beta)$. $\Longrightarrow$ w_i and X_i are exogenous variables.

Under the first assumption, we have the following equations:

$$P(y_i = 0) = P(\text{Regime 1}) + P(y_i = 0|\text{Regime 2})P(\text{Regime 2})$$
$$P(y_i = j) = P(y_i = j|\text{Regime 2})P(\text{Regime 2}),$$

for $j = 1, 2, \cdots.$

Combining the above two equations, we obtain the following:

$$P(y_i = j) = P(\text{Regime 1})I_i + P(y_i = j|\text{Regime 2})P(\text{Regime 2}),$$

for $j = 0, 1, 2, \cdots$,

where the indicator function I_i is given by $I_i = 1$ for $y_i = 0$ and $I_i = 0$ for $y_i \neq 0$.

F_i denotes a cumulative distribution function of $Z_i\alpha$. $\Longrightarrow$ We have to assume F_i.

Including the other two assumptions, we obtain the distribution of y_i as follows:

$$P(y_i = j) = F_i I_i + \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}(1 - F_i), \qquad j = 0, 1, 2, \cdots$$

where $F_i \equiv F(Z_i\alpha)$, $\lambda_i = \exp(X_i\beta)$, and the indicator function I_i is given by $I_i = 1$ for $y_i = 0$ and $I_i = 0$ for $y_i \neq 0$.

Therefore, the log-likelihood function is:

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

where $F_i \equiv F(Z_i\alpha)$ and $\lambda_i = \exp(X_i\beta)$.

$\log L(\alpha,\beta)$ is maximized with respect to α and β.

$\Longrightarrow$ The Newton-Raphson method or the method of scoring is utilized for maximization.

Example of Poisson Regression:

bike	自転車事故死者数（2012 年）
lowland	低地面積（平方キロ，2012 年）
dwellings	居住用宅地面積（平方キロ，2012 年）
pop	人口（2010 年）

	pref	bike	lowland	dwellings	pop
北海道	1	11	9794	543	5504
青森	2	6	1237	193	1374
岩手	3	7	1261	216	1326
宮城	4	4	1757	259	2352
秋田	5	2	2453	170	1085
山形	6	5	1393	163	1167
福島	7	5	1437	255	2021
茨城	8	20	1647	454	2887
栃木	9	17	752	289	1990
群馬	10	17	585	272	2005
埼玉	11	42	1414	487	6373
千葉	12	30	1452	489	5560
東京	13	34	274	421	15576
神奈川	14	17	575	418	8254
新潟	19	5	2775	274	2375
富山	20	4	987	145	1091
石川	15	5	656	116	1172
福井	16	2	932	93	807
山梨	17	4	343	115	855

長野	21	7	751	307	2149
岐阜	22	12	1174	226	1998
静岡	23	22	1155	338	3760
愛知	24	44	1148	521	7521
三重	18	8	1031	207	1820
滋賀	25	6	935	132	1363
京都	26	15	820	149	2668
大阪	27	47	610	318	9281
兵庫	28	23	1604	346	5348
奈良	29	4	273	110	1260
和歌山	30	7	316	93	983
鳥取	31	4	411	70	589
島根	32	3	495	94	718
岡山	33	14	1141	216	1943
広島	34	12	559	232	2869
山口	35	2	461	173	1444
徳島	36	7	551	88	783
香川	37	17	474	117	998
愛媛	38	9	557	146	1433
高知	39	6	327	70	763
福岡	40	18	1224	400	5078
佐賀	41	6	645	103	852
長崎	42	1	339	141	1423
熊本	43	14	958	225	1810
大分	44	6	595	140	1197
宮崎	45	6	764	163	1136
鹿児島	46	5	771	258	1704
沖縄	47	1	151	98	1392

```
----

. poisson bike lowland dwellings pop

Iteration 0:   log likelihood = -156.83031
Iteration 1:   log likelihood = -153.97721
Iteration 2:   log likelihood = -153.97403
Iteration 3:   log likelihood = -153.97403

Poisson regression                                Number of obs   =         47
                                                  LR chi2(3)      =     286.85
                                                  Prob > chi2     =     0.0000
Log likelihood = -153.97403                       Pseudo R2       =     0.4823

------------------------------------------------------------------------------
        bike |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lowland |  -.0001559   .0000368    -4.23   0.000    -.0002281   -.0000837
   dwellings |   .0042478    .000447     9.50   0.000     .0033716    .0051239
         pop |   .0000519   .0000146     3.56   0.000     .0000234    .0000804
       _cons |   1.309844   .1051302    12.46   0.000     1.103793    1.515896
------------------------------------------------------------------------------

. gen llland=log(lowland)

. gen ldwellings=log(dwellings)

. gen lpop=log(pop)
```

```
. poisson bike llland ldwellings lpop

Iteration 0:   log likelihood = -156.15686
Iteration 1:   log likelihood =  -155.6255
Iteration 2:   log likelihood = -155.62489
Iteration 3:   log likelihood = -155.62489

Poisson regression                                Number of obs   =         47
                                                  LR chi2(3)      =     283.54
                                                  Prob > chi2     =     0.0000
Log likelihood = -155.62489                       Pseudo R2       =     0.4767

------------------------------------------------------------------------------
        bike |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      llland |  -.1028579   .0800629    -1.28   0.199    -.2597784    .0540625
  ldwellings |   .4817018   .2171779     2.22   0.027      .056041    .9073626
        lpop |   .5715923   .1220733     4.68   0.000      .332333    .8108517
       _cons |   -3.93974    .559487    -7.04   0.000    -5.036315   -2.843166
------------------------------------------------------------------------------
```

8 Panel Data

8.1 Some Formulas of Matrix Algebra — Review

1. Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lk} \end{pmatrix} = [a_{ij}]$,

   which is a $l \times k$ matrix, where a_{ij} denotes ith row and jth column of A.

   The **transposed matrix (転置行列)** of A, denoted by A', is defined as:

$$A' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{l1} \\ a_{12} & a_{22} & \cdots & a_{l2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{lk} \end{pmatrix} = [a_{ji}],$$

where the ith row of A' is the ith column of A.

2. $(Ax)' = x'A'$,

   where A and x are a $l \times k$ matrix and a $k \times 1$ vector, respectively.

3. $a' = a$,

   where a denotes a scalar.

4. $\dfrac{\partial a'x}{\partial x} = a$,

   where a and x are $k \times 1$ vectors.

5. $\dfrac{\partial x'Ax}{\partial x} = (A + A')x$,

   where A and x are a $k \times k$ matrix and a $k \times 1$ vector, respectively.

   Especially, when A is symmetric,

$$\frac{\partial x'Ax}{\partial x} = 2Ax.$$

6. Let A and B be $k \times k$ matrices, and I_k be a $k \times k$ **identity matrix (単位行列)** (one in the diagonal elements and zero in the other elements).

   When $AB = I_k$, B is called the **inverse matrix (逆行列)** of A, denoted by $B = A^{-1}$.

   That is, $AA^{-1} = A^{-1}A = I_k$.

7. Let A be a $k \times k$ matrix and x be a $k \times 1$ vector.

   If A is a **positive definite matrix (正値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax > 0.$$

   If A is a **positive semidefinite matrix (非負値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax \geq 0.$$

If A is a **negative definite matrix (負値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax < 0.$$

If A is a **negative semidefinite matrix (非正値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax \leq 0.$$

Trace, Rank and etc.: $A : k \times k, \qquad B : n \times k, \qquad C : k \times n.$

1. The **trace (トレース)** of A is: $\mathrm{tr}(A) = \sum_{i=1}^{k} a_{ii}$, where $A = [a_{ij}]$.

2. The **rank (ランク，階数)** of A is the maximum number of linearly independent column (or row) vectors of A, which is denoted by rank(A).

3. If A is an **idempotent matrix** (べき等行列), $A = A^2$.

4. If A is an idempotent and symmetric matrix, $A = A^2 = A'A$.

5. A is idempotent if and only if the eigen values of A consist of 1 and 0.

6. If A is idempotent, rank(A) =tr(A) .

7. tr(BC) =tr(CB)

Distributions in Matrix Form:

1. Let X, μ and Σ be $k \times 1$, $k \times 1$ and $k \times k$ matrices.

When $X \sim N(\mu, \Sigma)$, the density function of X is given by:

$$f(x) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\Big).$$

$\mathrm{E}(X) = \mu$ and $\mathrm{V}(X) = \mathrm{E}\Big((X-\mu)(X-\mu)'\Big) = \Sigma$

The moment-generating function: $\phi(\theta) = \mathrm{E}\Big(\exp(\theta' X)\Big) = \exp(\theta'\mu + \frac{1}{2}\theta'\Sigma\theta)$

(*) In the univariate case, when $X \sim N(\mu, \sigma^2)$, the density function of X is:

$$f(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big).$$

2. If $X \sim N(\mu, \Sigma)$, then $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2(k)$.

   Note that $X'X \sim \chi^2(k)$ when $X \sim N(0, I_k)$.

3. X: $n \times 1$, Y: $m \times 1$, $X \sim N(\mu_x, \Sigma_x)$, $Y \sim N(\mu_y, \Sigma_y)$

X is independent of Y, i.e., $\mathrm{E}\Big((X-\mu_x)(Y-\mu_y)'\Big) = 0$ in the case of normal random variables.

$$\frac{(X-\mu_x)'\Sigma_x^{-1}(X-\mu_x)/n}{(Y-\mu_y)'\Sigma_y^{-1}(Y-\mu_y)/m} \sim F(n,m)$$

4. If $X \sim N(0, \sigma^2 I_n)$ and A is a symmetric idempotent $n \times n$ matrix of rank G, then $X'AX/\sigma^2 \sim \chi^2(G)$.

   Note that $X'AX = (AX)'(AX)$ and $\mathrm{rank}(A) = \mathrm{tr}(A)$ because A is idempotent.

5. If $X \sim N(0, \sigma^2 I_n)$, A and B are symmetric idempotent $n \times n$ matrices of rank G and K, and $AB = 0$, then

$$\frac{X'AX}{G\sigma^2}\bigg/\frac{X'BX}{K\sigma^2} = \frac{X'AX/G}{X'BX/K} \sim F(G,K).$$