Discrete Choice Analysis

Taro Mieno

AECN 896-003: Applied Econometrics

Discrete Choice Analysis

- Focus on understanding choices that are discrete (not continuous)
  - Whether you own a car or not (binary choice)
  - Whether you use an iPhone, Android, or other types of cell phones (Multinomial choice)
  - Which recreation sites you visit this winter (multinomial)
- Linear models we have seen are often not appropriate

Binary Response Model

Binary Response

$y = 0$ (if you do not own a car)

$y = 1$ (if you own at least one car)

Question we would like to answer

How do independent variables $x_1, \ldots, x_k$ affect the status of y (the choice of whether to own at least one car or not)?

Binary Response Model

We try to model the probability of $y = 1$ (own at least one car)

$$Pr(y = 1|x_1, \dots, x_k) = f(x_1, \dots, x_k)$$

as a function of independent variables.

Linear probability model

$$Pr(y = 1|x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Binary Response Model

We try to model the probability of $y = 1$ (own at least one car)

$$Pr(y = 1|x_1, \ldots, x_k) = f(x_1, \ldots, x_k)$$

as a function of independent variables.

So, how about

$$Pr(y = 1|x_1, \ldots, x_k) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

such that $G()$ is a function taking on values strictly between zero and once: $0 < G(z) < 1$ for all real numbers z?

Notes

Different choices of $G()$ lead to different models

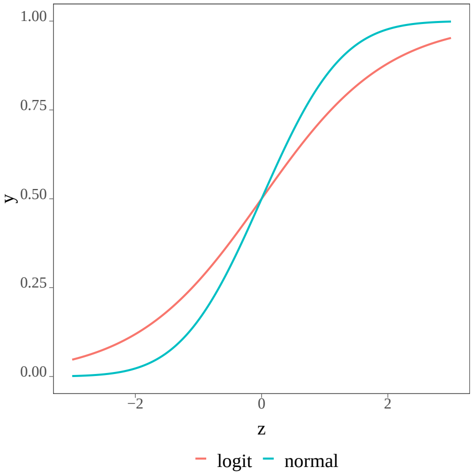
Logit model

$$G(z) = exp(z)/[1 + exp(z)] = \frac{e^z}{1 + e^z}$$

Probit model

$$G(z) = \Phi(z)$$

where $\Phi(z)$ is the standard normal cumulative distribution function


1.00
0.75
0.50
0.25
0.00
y
−2
0
2
z
logit
normal

Interpretation

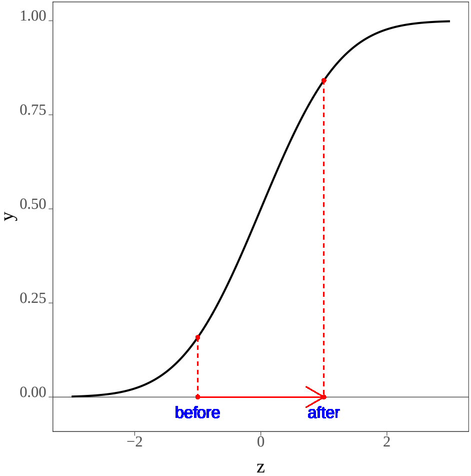
$$Pr(y = 1|x_1, \ldots, x_k) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

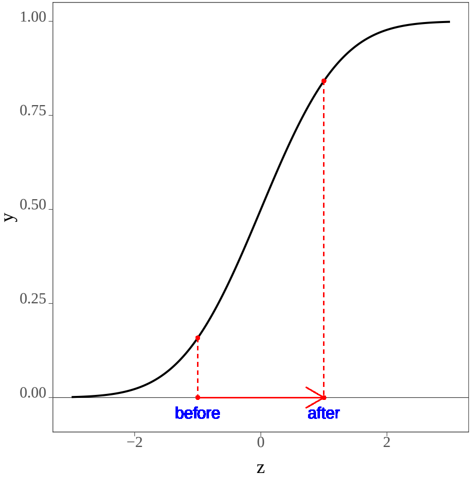
- What do βs measure?
- How do we interpret them?

Before and after

Before: $x_1 = 0, \ldots, x_k = 0 \Rightarrow z = \beta_0$

After: $x_1 = 1$ and $x_2 = 0, \ldots, x_k = 0 \Rightarrow z = \beta_0 + \beta_1 x_1$





- βs measure how far you move along the x-axis
- βs does not directly measure how independent variables influence the probility of $y = 1$

Marginal impact of x_k (continuous) on $Prob(y = 1)$

$$Pr(y = 1|x_1, \ldots, x_k) = G(z)$$
$$z = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Differentiating both sides with respect to x_k,

$$\begin{aligned}\frac{\partial Pr(y = 1|x_1, \ldots, x_k)}{\partial x_k} &= G'(z) \times \frac{\partial z}{\partial x_k} \\ &= G'(z) \times \beta_k\end{aligned}$$

Marginal impact of x_k (continuous) on $Prob(y = 1)$

$$Pr(y = 1|x_1, \ldots, x_k) = G(z)$$
$$z = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Differentiating both sides with respect to x_k,

$$\begin{aligned} \frac{\partial Pr(y = 1|x_1, \ldots, x_k)}{\partial x_k} &= G'(z) \times \frac{\partial z}{\partial x_k} \\ &= G'(z) \times \beta_k \end{aligned}$$

Notes

- The marginal impact of an independent variable depends on the values of all the independent variables: $G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$
- Since $G'()$ is always positive, the sign of the marginal impact of an independent variable on $Prob(y = 1)$ is always the same as the sign of its coefficient

Estimation of Binary Choice Models

- Linear models: OLS
- Binary choice models: Maximum Likelihood Estimation (MLE)

OLS

Find parameters that makes the sum of residuals squared the smallest

MLE (very loosely put)

Find parameters (βs) that makes what we observed (collection of binary decisions made by different individuals) most likely (Maximum Likelihood)

Decisions made by two individuals

- Individual 1: $y = 1$ (own at least one car)
- Individual 2: $y = 0$ (does not own a car)

Decisions made by two individuals

- Individual 1: $y = 1$ (own at least one car)
- Individual 2: $y = 0$ (does not own a car)

Probability of individual decisions

Individual 1 :$Prob(y_1 = 1|\mathbf{x_1}) = G(z_1)$

Individual 2 :$Prob(y_2 = 0|\mathbf{x_2}) = 1 - G(z_2)$

Decisions made by two individuals

- Individual 1: $y = 1$ (own at least one car)
- Individual 2: $y = 0$ (does not own a car)

Probability of individual decisions

$$\text{Individual 1 :} Prob(y_1 = 1|\mathbf{x_1}) = G(z_1)$$

$$\text{Individual 2 :} Prob(y_2 = 0|\mathbf{x_2}) = 1 - G(z_2)$$

Probability of a collection of decisions

The probability that we observe a collection of choices made by them (if their decisions are independent)

$$Prob(y_1 = 1|\mathbf{x_1}) \times Prob(y_2 = 0|\mathbf{x_2}) = G(z_1) \times [1 - G(z_2)],$$

which we call likelihood function.

Probability of individual decisions

Individual 1 :$Prob(y_1 = 1|\mathbf{x_1}) = G(z_1)$

Individual 2 :$Prob(y_2 = 0|\mathbf{x_2}) = 1 - G(z_2)$

Probability of a collection of decisions

The probability that we observe a collection of choices made by them (if their decisions are independent)

$$Prob(y_1 = 1|\mathbf{x_1}) \times Prob(y_2 = 0|\mathbf{x_2}) = G(z_1) \times [1 - G(z_2)],$$

which we call likelihood function.

MLE

$$Max_{\beta_1,\ldots,\beta_k} \;\; G(z_1) \times [1 - G(z_2)]$$

MLE of Binary Choice Model in General

Maximize the likelihood function:

$$Max_{\beta_1,\ldots,\beta_k} \;\; L$$

where $L = \Pi_{i=1}^{n}\Big[y_i \times G(z_i) + (1-y_i) \times (1-G(z_i))\Big]$ is the likelihood function.

Log-likelihood function

$$\begin{aligned} LL &= log\Big(\Pi_{i=1}^{n}\Big[y_i \times G(z_i) + (1-y_i) \times (1-G(z_i))\Big]\Big) \\ &= \sum_{i=1}^{n} log\Big(y_i \times G(z_i) + (1-y_i) \times (1-G(z_i))\Big) \end{aligned}$$

MLE with LL

$$argmax_{\beta_1,\ldots,\beta_k} \;\; L \equiv argmax_{\beta_1,\ldots,\beta_k} \;\; LL$$

Implementation in R with an example

Participation of females in labor force:

$$Pr(inlf = 1|\mathbf{x}) = G(z)$$

where

$$z = \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \\ + \beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6$$

- $inlf$: 1 if in labor force in 1975, 0 otherwise
- $nwifeinc$: earning as a family if she does not work
- $kidslt6$: # of kids less than 6 years old
- $kidsge6$: # of kids who are 6-18 year old

R code: importing the data

```
#--- import the data ---#
data <- read.dta13('MROZ.dta') %>%
  mutate(exper2=exper^2)

#--- take a look ---#
dplyr::select(data,inlf,nwifeinc,kidslt6,kidsge6,educ) %>%
  head()
```

```
  inlf  nwifeinc kidslt6 kidsge6 educ
1    1 10.910060       1       0   12
2    1 19.499981       0       2   12
3    1 12.039910       1       3   12
4    1  6.799996       0       3   12
5    1 20.100058       1       2   14
6    1  9.859054       0       0   12
```

Estimating binary choice model using R

You can use the $glm()$ function (no new packages installation necessary) when using cross-sectional data

- glm refers to Generalized Linear Model, which encompass linear models we have been using
- you specify the $family$ option to tell what kind of model you are estimating

R code: Probit model estimation using glm()

```
probit_lf <- glm(

#--- formula ---#
inlf~nwifeinc+educ+exper+exper2+age+kidslt6+kidsge6,

#--- data ---#
data = data,

#--- models ---#
family=binomial(link='probit')
)
```

$family$ option

- $binomial()$: tells R that your dependent variable is binary
- $link =' probit'$: tells R that you want to use the cumulative distribution function of the standard normal distribution as $G()$ in $Prob(y = 1|\mathbf{x}) = G(z)$

R code: Probit model estimation using glm()

```
summary(probit_lf)$coef

                Estimate    Std. Error    z value     Pr(>|z|)
(Intercept)  0.270073573 0.5080781657  0.5315591 5.950314e-01
nwifeinc    -0.012023637 0.0049391713 -2.4343430 1.491885e-02
educ         0.130903969 0.0253987284  5.1539576 2.550456e-07
exper        0.123347168 0.0187586870  6.5754692 4.850000e-11
exper2      -0.001887067 0.0005999272 -3.1454942 1.658065e-03
age         -0.052852442 0.0084623619 -6.2455899 4.222037e-10
kidslt6     -0.868324680 0.1183772702 -7.3352315 2.213386e-13
kidsge6      0.036005611 0.0440302624  0.8177469 4.135017e-01
```

R code: Logit model estimation using glm()

```
logit_lf <- glm(

#--- formula ---#
inlf~nwifeinc+educ+exper+exper2+age+kidslt6+kidsge6,

#--- data ---#
data = data,

#--- models ---#
family=binomial(link='logit')
)
```

$family$ option

- $binomial()$: tells R that your dependent variable is binary
- $link =' logit'$: tells R that you want to use the function as $G()$ in $Prob(y = 1|\mathbf{x}) = G(z)$

Logit model estimation results

```r
summary(logit_lf)$coef
```

```
                Estimate  Std. Error    z value     Pr(>|z|)
(Intercept)  0.425452376 0.860364519  0.4945025 6.209514e-01
nwifeinc    -0.021345174 0.008421380 -2.5346410 1.125626e-02
educ         0.221170370 0.043439281  5.0914832 3.552734e-07
exper        0.205869531 0.032056713  6.4220411 1.344591e-10
exper2      -0.003154104 0.001016107 -3.1041065 1.908546e-03
age         -0.088024375 0.014572890 -6.0402826 1.538446e-09
kidslt6     -1.443354143 0.203582842 -7.0897632 1.343417e-12
kidsge6      0.060112222 0.074789293  0.8037544 4.215388e-01
```

```r
stargazer(probit_lf,logit_lf,type='latex',no.space=TRUE,table.layout='-ldc-t-s-')
```

Table

	Dependent variable:	
	inlf	
nwifeinc	-0.012^{**}	-0.021^{**}
	(0.005)	(0.008)
educ	0.131^{***}	0.221^{***}
	(0.025)	(0.043)
exper	0.123^{***}	0.206^{***}
	(0.019)	(0.032)
exper2	-0.002^{***}	-0.003^{***}
	(0.001)	(0.001)
age	-0.053^{***}	-0.088^{***}
	(0.008)	(0.015)
kidslt6	-0.868^{***}	-1.443^{***}
	(0.118)	(0.204)
kidsge6	0.036	0.060
	(0.044)	(0.075)
Constant	0.270	0.425
	(0.508)	(0.860)
Observations	753	753
Log Likelihood	−401.302	−401.765
Akaike Inf. Crit.	818.604	819.530

Post-estimation operations and diagnostics

Log-likelihood

$$LL = \sum_{i=1}^{n} log\Big(y_i \times G(\hat{z}_i) + (1 - y_i) \times (1 - G(\hat{z}_i))\Big)$$

- $\hat{z}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$
- $G(\hat{z}_i)$ is the fitted value of $Prob(y = 1|\mathbf{x})$

(The greater (less negative) the LL, the better the fit of the regression)

McFadden's pseudo-R^2

A measure of how much better your model is compared to the model with only the intercept

$$pseudo - R^2 = 1 - LL/LL_0$$

where LL_0 is the log-likelihood when you include only the intercept

R code: pseudo-R^2

```
#--- estimate the model with only the intercept ---#
logit_lf_0 <- glm(inlf~1,data = data,family=binomial(link='logit'))

#--- extract LL using the logLik() function ---#
LL0 <- logLik(logit_lf_0)
LL0
```

```
'log Lik.' -514.8732 (df=1)
```

```
#--- extract LL using the logLik() function from your preferred model ---#
LL <- logLik(logit_lf)
LL
```

```
'log Lik.' -401.7652 (df=8)
```

```
#--- pseudo R2 ---#
pR2 <- 1-LL/LL0
pR2
```

```
'log Lik.' 0.2196814 (df=8)
```

R code: alternative (easier) calculation of pseudo-R^2

```
  #--- or more easily ---#
  1-logit_lf$deviance/logit_lf$null.deviance

[1] 0.2196814

  #--- what are deviances? ---#
  logit_lf$null.deviance # = -2*LL0

[1] 1029.746

  logit_lf$deviance # = -2*LL

[1] 803.5303
```

Testing: joint significance

You can do Likelihood Ratio (LR) test:

$$LR = 2(LL_{unrestricted} - LL_{restricted}) \sim \chi^2_{df\_restrictions}$$

where $df\_restrictions$ is the number of restrictions

Testing: joint significance

You can do Likelihood Ratio (LR) test:

$$LR = 2(LL_{unrestricted} - LL_{restricted}) \sim \chi^2_{df\_restrictions}$$

where $df\_restrictions$ is the number of restrictions

An example

- H_0 : the coefficients on $exper$, $exper2$, and age are 0
- H_1 : H_0 is false

R code: LR test for joint significance

```r
#--- unrestricted ---#
logit_ur <- glm(
  inlf~nwifeinc+educ+exper+exper2+age+kidslt6+kidsge6,
  data = data,family=binomial(link='logit')
  )

#--- restricted ---#
logit_r <- glm(
  inlf~nwifeinc+educ+kidslt6+kidsge6,
  data = data,family=binomial(link='logit')
  )

#--- LR test using lrtest() from the lmtest package ---#
library(lmtest)
lrtest(logit_r,logit_ur)
```

```
Likelihood ratio test

Model 1: inlf ~ nwifeinc + educ + kidslt6 + kidsge6
Model 2: inlf ~ nwifeinc + educ + exper + exper2 + age + kidslt6 + kidsge6
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   5 -464.92
2   8 -401.77  3 126.32  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Prediction

After estimating a binary choice model, you can easily predict the following two

- $\hat{z} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$
- $\widehat{Prob(y=1}|\mathbf{x}) = G(\hat{z}) = G(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k)$

R code: LR test for joint significance

```
#--- z hat ---#
z <- predict(probit_lf,type='link')
head(z)
```

```
        1         2         3         4         5         6
0.5071349 0.6624576 0.5116317 0.7423429 0.1972781 0.8837878
```

```
#--- G(z) hat ---#
Gz <- predict(probit_lf,type='response')
head(Gz)
```

```
        1         2         3         4         5         6
0.6939699 0.7461610 0.6955456 0.7710602 0.5781950 0.8115946
```

Marginal effect of an independent variable

- Coefficient estimates across different models (probit and logit) are not meaningful because the same value of a coefficient estimate means different things
- They are the estimates of βs, not the direct impact of the independent variables on the $Prob(y = 1)$

Calculating the marginal effect of an independent variable

$$\frac{\partial Pr(y = 1|x_1, \ldots, x_k)}{\partial x_k} = G'(z) \times \beta_k$$

- the marginal impact depends on the current levels of all the independent variables
- we typically report one of the two types of marginal impacts
  - the marginal impact at the mean (average person): when all the independent variables take on their respective means
  - the average of the marginal impacts calculated for each of all the individuals observed

Marginal impact at the mean

$$\frac{\partial Pr(y=1|\bar{x_1},\ldots,\bar{x_k})}{\partial x_k} = G'(\beta_0+\beta_1\bar{x_1}+\cdots+\beta_k\bar{x_k}) \times \beta_k$$

Mean marginal impact

$$\sum_{i=1}^{n} \frac{\partial Pr(y_i=1|x_{\bar{i},1},\ldots,x_{\bar{i},k})}{\partial x_k} = \sum_{i=1}^{n} G'(z_i) \times \beta_k$$

R code: marginal impact of *educ* at the mean

```
#--- get the coef ---#
probit_lf$coef

(Intercept)     nwifeinc         educ        exper       exper2
0.270073573 -0.012023637  0.130903969  0.123347168 -0.001887067
        age      kidslt6      kidsge6
-0.052852442 -0.868324680  0.036005611

#--- get the mean ---#
means <- summarize(data,mean(nwifeinc),mean(educ),mean(exper),mean(exper2),
  mean(age),mean(kidslt6),mean(kidsge6))

#--- calculate z ---#
z <- probit_lf$coef[1] + sum(probit_lf$coef[-1]*means)

#--- marignal impact ---#
dnorm(z)*probit_lf$coef['educ']

(Intercept)
 0.05112843
```

R code: marginal impact of $educ$ at the mean

```
#--- get the coef ---#
probit_lf$coef
```

```
(Intercept)     nwifeinc         educ        exper       exper2
0.270073573 -0.012023637  0.130903969  0.123347168 -0.001887067
        age      kidslt6      kidsge6
-0.052852442 -0.868324680  0.036005611
```

```
#--- get the mean ---#
means <- summarize(data,mean(nwifeinc),mean(educ),mean(exper),mean(exper2),
  mean(age),mean(kidslt6),mean(kidsge6))

#--- calculate z ---#
z <- probit_lf$coef[1] + sum(probit_lf$coef[-1]*means)

#--- marignal impact ---#
dnorm(z)*probit_lf$coef['educ']
```

```
(Intercept)
 0.05112843
```

Interpretation

If your eduction goes up by 1 year, you are 5% more likely to be in the labor force when you are an average person

R code: mean marginal impact of $educ$

```
#--- get z for all the individuals ---#
z <- predict(probit_lf,type='link')

#--- get G'(z) ---#
Gz_indiv <- dnorm(z)

#--- mean marignal impact of eduction ---#
mean(Gz_indiv)*probit_lf$coef['educ']
```

```
      educ
0.03937009
```

Regression Models for Count Data

Count as the dependent variable

Count variables take non-negative discrete values $(0, 1, \ldots,)$

- the number of times people get arrested in a year
- the number of cars owned by a family
- the number of patents applied for by a firm in a year
- the number of kids in a family

Poisson regression

By far the most pupular choice to analyze count variables is Poisson regression

- The outcome (count) variable is assumed to be Poisson distributed
- The mean of the Poisson distribution is assumed to be a function of some variables you believe matter

Poisson regression

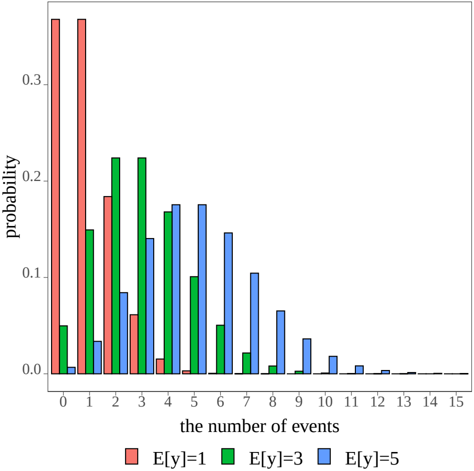
By far the most pupular choice to analyze count variables is Poisson regression

- The outcome (count) variable is assumed to be Poisson distributed
- The mean of the Poisson distribution is assumed to be a function of some variables you believe matter

Poisson distribution

Poisson distribution is a discrete probability distribution that describes the probability of the number of events that occur in a fixed interval of time and/or space

$$Prob(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, \text{ where } \lambda = E[y]$$



Poisson regression

We try to learn what and how variables affect the expected value (the expected number of events conditional on independent variables)

Expected number of events conditional on independent variables

$$E[y|\mathbf{x}] = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

- This is exactly the same modeling framework we used
  - Linear model: $G(z) = z$
  - Probit model: $G(z) = \Phi(z)$

Expected number of events conditional on independent variables

$$E[y|\mathbf{x}] = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

- This is exactly the same modeling framework we used
  - Linear model: $G(z) = z$
  - Probit model: $G(z) = \Phi(z)$

A popular choice of $G()$

$$G(z) = exp(z)$$

- ensures that the expected value conditional on $\mathbf{x}$ is always positive

The number of events for two individuals

- Individual 1: $y = 3$ (own three cars)
- Individual 2: $y = 1$ (own one car)

The number of events for two individuals

- Individual 1: $y = 3$ (own three cars)
- Individual 2: $y = 1$ (own one car)

Expected number of events observed

$$\text{Individual 1 :} \lambda_1 = exp(z_1)$$
$$\text{Individual 2 :} \lambda_2 = exp(z_2)$$

The number of events for two individuals

- Individual 1: $y = 3$ (own three cars)
- Individual 2: $y = 1$ (own one car)

Expected number of events observed

Individual 1 :$\lambda_1 = exp(z_1)$

Individual 2 :$\lambda_2 = exp(z_2)$

Probability of observing the number of events we observed

Individual 1 :$Prob(y = 3|\mathbf{x_1}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!}$

Individual 2 :$Prob(y = 1|\mathbf{x_2}) = \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$

Probability of observing the number of events we observed

$$\text{Individual 1 :} Prob(y = 3|\mathbf{x_1}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!}$$

$$\text{Individual 2 :} Prob(y = 1|\mathbf{x_2}) = \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

Probability of observing a series of events by all individuals

The probability that we observe a collection of choices made by them (if their events are independent)

$$L = Prob(y_1 = 3|\mathbf{x_1}) \times Prob(y_2 = 1|\mathbf{x_2}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!} \times \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

which we call likelihood function.

Probability of observing a series of events by all individuals

The probability that we observe a collection of choices made by them (if their events are independent)

$$L = Prob(y_1 = 3|\mathbf{x_1}) \times Prob(y_2 = 1|\mathbf{x_2}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!} \times \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

which we call likelihood function.

Log-likelihood function

$$LL = log(L) = log(\frac{\lambda_1^3 e^{-\lambda_1}}{3!}) + log(\frac{\lambda_2^1 e^{-\lambda_2}}{1!})$$

(Remember $\lambda_i = exp(\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k})$)

Probability of observing a series of events by all individuals

The probability that we observe a collection of choices made by them (if their events are independent)

$$L = Prob(y_1 = 3|\mathbf{x_1}) \times Prob(y_2 = 1|\mathbf{x_2}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!} \times \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

which we call likelihood function.

Log-likelihood function

$$LL = log(L) = log(\frac{\lambda_1^3 e^{-\lambda_1}}{3!}) + log(\frac{\lambda_2^1 e^{-\lambda_2}}{1!})$$

(Remember $\lambda_i = exp(\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k})$)

MLE

$$Max_{\beta_1,\ldots,\beta_k} \;\; LL$$

Implementation in R with an example

The number of times a man is arrested during 1986:

$$Pr(narr86|\mathbf{x}) = G(z)$$

where

$$\begin{aligned} z =&\beta_0 + \beta_1 pcnv + \beta_2 tottime + \beta_3 qemp86 + \beta_4 inc86 \\ &+ \beta_5 black + \beta_6 hispan \end{aligned}$$

- $narr86$: # of times arrested in 1986
- $pcnv$: proportion of prior conviction
- $tottime$: time in prison since 18
- $qemp86$: # quarters employed in 1986
- $inc86$: legal income in 1986 (in $\$100$)

R code: importing the data

```
998         0 0.00    3.0 133.1
999         0 1.00    4.0  95.6
1000        0 0.25    4.0 170.7
1001        0 1.00    4.0  54.3
1002        0 1.00    4.0 152.7
1003        0 0.60    0.0   0.0
1004        1 0.00    2.0   2.5
1005        0 0.67    2.0  20.0
1006        0 1.00    4.0 107.3
1007        1 0.33    0.0   0.0
1008        2 0.20    0.0   0.0
1009        0 1.00    4.0  67.1
1010        5 0.25    4.0  61.4
1011        2 0.25    1.0   5.6
1012        0 0.00    1.0   9.6
1013        0 1.00    4.0 175.6
1014        0 0.25    2.0  13.7
1015        0 0.60    0.0   0.0
1016        0 0.00    4.0  28.6
1017        0 0.00    3.0  54.5
1018        0 0.40    2.0   8.2
1019        3 0.56    0.0   0.0
1020        0 0.00    3.0  51.8
1021        0 0.00    0.0   0.0
1022        2 0.56    2.0   4.3
1023        0 0.33    1.0   7.5
1024        2 0.50    3.0  44.5
1025        0 1.00    2.0  27.3
1026        1 0.50    2.0   2.5
```

R code: Poisson model estimation using glm()

```
pois_lf <- glm(

#--- formula ---#
narr86~pcnv+tottime+qemp86+inc86+black+hispan,

#--- data ---#
data = data,

#--- models ---#
family=poisson(link='log')
)
```

$family$ option

- $poisson()$: tells R that your dependent variable is Poisson distributed
- $link =' log'$: tells R that you want to use $exp()$ (the inverse of $log()$) as $G()$ in $E(y = 1|\mathbf{x}) = G(z)$

R code: summary

```
summary(pois_lf)$coef
```

```
                Estimate  Std. Error     z value     Pr(>|z|)
(Intercept) -0.665805562 0.063749782 -10.4440446 1.560185e-25
pcnv        -0.432102017 0.085403627  -5.0595277 4.202962e-07
tottime     -0.001206742 0.005504698  -0.2192204 8.264784e-01
qemp86      -0.009956885 0.028588565  -0.3482821 7.276284e-01
inc86       -0.008395041 0.001039638  -8.0749648 6.749612e-16
black        0.644485369 0.073922145   8.7184344 2.820750e-18
hispan       0.473213044 0.073774349   6.4143303 1.414432e-10
```

```r
stargazer(pois_lf,type='latex',no.space=TRUE,table.layout='-ldc-t-s-')
```

Table

	Dependent variable:
	narr86
pcnv	-0.432^{***}
	(0.085)
tottime	-0.001
	(0.006)
qemp86	-0.010
	(0.029)
inc86	-0.008^{***}
	(0.001)
black	0.644^{***}
	(0.074)
hispan	0.473^{***}
	(0.074)
Constant	-0.666^{***}
	(0.064)
Observations	2,725
Log Likelihood	$-2,264.668$
Akaike Inf. Crit.	4,543.337

Calculating the marginal impact of an independent variable

$$\frac{\partial E(y|x_1, \ldots, x_k)}{\partial x_k} = \beta_k \times exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

- the marginal impact depends on the current levels of all the independent variables
- we typically report one of the two types of marginal impacts
  - the marginal impact at the mean (average person): when all the independent variables take on their respective means
  - the average of the marginal impacts calculated for each of all the individuals observed

R code: mean marginal impact of income

```
#--- get z for all the individuals ---#
z <- predict(pois_lf,type='link')

#--- get G'(z) ---#
Gz_indiv <- exp(z)

#--- mean marignal impact of eduction ---#
mean(Gz_indiv)*pois_lf$coef['inc86']
```

```
       inc86
-0.003394985
```

R code: mean marginal impact of income

```r
#--- get z for all the individuals ---#
z <- predict(pois_lf,type='link')

#--- get G'(z) ---#
Gz_indiv <- exp(z)

#--- mean marignal impact of eduction ---#
mean(Gz_indiv)*pois_lf$coef['inc86']
```

```
      inc86
-0.003394985
```

Interpretation

If your income goes up by $\$100$, the expected number of getting arrested declines by 0.0034

Notes

- Poisson regression model is under the same econometric modeling framework: GLM
- Codes for testing are exactly the same as those we saw for the binary response models

Multinomial Choice

Multinomial Choice

Instead of two options, you may be picking one option out of more than two options

- which carrier?
  - Verizon
  - Sprint
  - AT&T
  - T-mobile
- which transportation means to commute?
  - own car
  - Uber
  - bus
  - train
  - bike

Multinomial logit model

The most popular model to analyze multinomial choice

- environmental evaluation
- tranposrtation
- marketing

Understanding multinomial logit model through an example

Choice of trains

1. 10 euros, 30 minutes travel time, one change
2. 20 euros, 20 minutes travel time, one change
3. 22 euros, 22 minutes travel time, no change

Associated utility

1. $V_1 = \alpha_1 + \beta 10 + \gamma 30 + \rho 1 + v_1$
2. $V_2 = \alpha_2 + \beta 20 + \gamma 20 + \rho 1 + v_2$
3. $V_3 = \alpha_3 + \beta 22 + \gamma 22 + \rho 0 + v_3$

Choice probability

Logit model assumes that the probability of choosing an alternative is the following:

1. $P_1 = \dfrac{e^{V_1}}{e^{V_1} + e^{V_2} + e^{V_3}}$
2. $P_2 = \dfrac{e^{V_2}}{e^{V_1} + e^{V_2} + e^{V_3}}$
3. $P_3 = \dfrac{e^{V_3}}{e^{V_1} + e^{V_2} + e^{V_3}}$

Choice probability

Logit model assumes that the probability of choosing an alternative is the following:

1. $P_1 = \dfrac{e^{V_1}}{e^{V_1} + e^{V_2} + e^{V_3}}$
2. $P_2 = \dfrac{e^{V_2}}{e^{V_1} + e^{V_2} + e^{V_3}}$
3. $P_3 = \dfrac{e^{V_3}}{e^{V_1} + e^{V_2} + e^{V_3}}$

Notes

- $0 < P_j < 1, \ ^{\forall} j = 1, 2, 3$
- $\displaystyle\sum_{j=1}^{3} = 1$

Probability of observing individual i choosing what i chose

$$P_i = \Pi_{j=1}^{3} y_{i,j} \times P_j$$

where $y_{i,j} = 1$ if i chose j, 0 otherwise

An example

$$y_{i,1} = 0, \;\; y_{i,2} = 1, \;\; y_{i,3} = 0$$
$$P_i = \Pi_{j=1}^{3} y_{i,j} \times P_j = 0 \times P_1 + 1 \times P_2 + 0 \times P_3$$

Probability of observing a series of chocies made by all the subjects

If choices made by the subjects are independent with each other,

$$LL = \Pi_{i=1}^{n} P_i = \Pi_{i=1}^{n} \Pi_{j=1}^{3} y_{i,j} \times P_j$$

MLE

$$Max_{\beta,\gamma,\rho} \;\; log(LL)$$

Interpreatation of the coefficients

Model in general

$$V_{i,j} = \alpha_j + \beta_1 x_{1,i,j} + \cdots + \beta_k x_{k,i,j}$$

$$P_{i,j} = \frac{e^{V_{i,j}}}{\sum_{k=1}^{J} e^{V_{i,k}}}$$

Interpreatation of the coefficients

$$\frac{\partial P_{i,j}}{\partial x_{k,i,j}} = \beta_k P_{i,j}(1 - P_{i,j})$$

- A marginal change in kth variable for alternative j would change the probability of choosing alternative j by $\beta_k P_{i,j}(1 - P_{i,j})$
- the sign of the impact is the same as the sign of the coefficient

Implementation in R

You can use $mlogit$ package to estimate multinomial logit models

- format your data in a specific manner
- convert your data using $mlogit.data()$
- estimate using $mlogit()$

R code: multinomial logit model

```
#--- library ---#
library(mlogit)

#--- get the heating data from the mlogit package ---#
data('TravelMode',package='AER')

#--- take a look at the data ---#
# first 10 rows
head(TravelMode,10)
```

```
   individual  mode choice wait vcost travel gcost income size
1           1   air     no   69    59    100    70     35    1
2           1 train     no   34    31    372    71     35    1
3           1   bus     no   35    25    417    70     35    1
4           1   car    yes    0    10    180    30     35    1
5           2   air     no   64    58     68    68     30    2
6           2 train     no   44    31    354    84     30    2
7           2   bus     no   53    25    399    85     30    2
8           2   car    yes    0    11    255    50     30    2
9           3   air     no   69   115    125   129     40    1
10          3 train     no   34    98    892   195     40    1
```

R code: data preparation

```
#--- convert the data ---#
TM <- mlogit.data(TravelMode,
  shape='long', # what format is the data in?
  choice='choice', # name of the variable that indicates choice made
 chid.var='individual', # name of the variable that indicates who made choices
  alt.var='mode' # the name of the variable that indicates options
  )

#--- take a look at the data ---#
# first 10 rows
head(TM,10)
```

```
        individual  mode choice wait vcost travel gcost income size
1.air            1   air  FALSE   69    59    100    70     35    1
1.train          1 train  FALSE   34    31    372    71     35    1
1.bus            1   bus  FALSE   35    25    417    70     35    1
1.car            1   car   TRUE    0    10    180    30     35    1
2.air            2   air  FALSE   64    58     68    68     30    2
2.train          2 train  FALSE   44    31    354    84     30    2
2.bus            2   bus  FALSE   53    25    399    85     30    2
2.car            2   car   TRUE    0    11    255    50     30    2
3.air            3   air  FALSE   69   115    125   129     40    1
3.train          3 train  FALSE   34    98    892   195     40    1
```

R code: multinomial logit model estimation

```
  #--- estimate ---#
  ml_reg <- mlogit(choice~wait+vcost+travel,data=TM)

  #--- summary ---#
  summary(ml_reg)

Call:
mlogit(formula = choice ~ wait + vcost + travel, data = TM, method = "nr")

Frequencies of alternatives:
    air   train     bus     car
0.27619 0.30000 0.14286 0.28095

nr method
5 iterations, 0h:0m:0s
g'(-H)^-1g = 0.000192
successive function values within tolerance limits

Coefficients :
                     Estimate  Std. Error z-value  Pr(>|z|)
train:(intercept) -0.78666667  0.60260733 -1.3054   0.19174
bus:(intercept)   -1.43363372  0.68071345 -2.1061   0.03520 *
car:(intercept)   -4.73985647  0.86753178 -5.4636 4.665e-08 ***
wait              -0.09688675  0.01034202 -9.3683 < 2.2e-16 ***
vcost             -0.01391160  0.00665133 -2.0916   0.03648 *
travel            -0.00399468  0.00084915 -4.7043 2.547e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -192.89
McFadden R^2:  0.32024
Likelihood ratio test : chisq = 181.74 (p.value = < 2.22e-16)
```

R code: understanding the results

```
#--- coefficient ---#
summary(ml_reg)$coef
```

```
train:(intercept)   bus:(intercept)   car:(intercept)              wait
     -0.786666672      -1.433633718      -4.739856473      -0.096886747
            vcost            travel
     -0.013911604      -0.003994681
attr(,"names.sup.coef")
character(0)
attr(,"fixed")
train:(intercept)   bus:(intercept)   car:(intercept)              wait
            FALSE             FALSE             FALSE             FALSE
            vcost            travel
            FALSE             FALSE
attr(,"sup")
character(0)
```

Notes

- intercept for air is dropped (air is the base)
  - train:(intercept) is -0.786 means that train is less likely to be chosen if all the other included variables are the same
- the greater the travel time, the less likely the option is chosen