Multinomial Choice Model

Tadao Hoshino （星野匡郎）

Econometrics II: ver. 2019 Spring Semester

Multinomial Choice Model

Example: Which OS to buy?

- Suppose that you are choosing the OS for your PC from Windows, Mac and Linux (e.g., Ubuntu).
- You obtain utilities U_W, U_M and U_L from Windows, Mac and Linux, respectively.
- For example, a Linux PC would be chosen if and only if (without considering ties) $U_L = \max\{U_W, U_M, U_L\}$.

Utility: U_L $>$ Utility: U_W

Utility: U_M

$U_L > U_M$
$U_L > U_W$
$\Rightarrow$ Linux PC will be chosen.

Example: Which OS to buy?

- Let X_i be a vector of observable individual characteristics, and ε_i be an unobservable random variable.
- Based on the random utility framework, i's utility of buying Windows, Mac and Linux PC can be written as

$$U_{W,i} = U_W(X_i, \varepsilon_{W,i}), \;\; U_{M,i} = U_M(X_i, \varepsilon_{M,i}), \;\; U_{L,i} = U_L(X_i, \varepsilon_{L,i})$$

  respectively.
- We assume that i will buy, for example, Windows PC if

$$U_W(X_i, \varepsilon_{W,i}) = \max\{U_W(X_i, \varepsilon_{W,i}), U_M(X_i, \varepsilon_{M,i}), U_L(X_i, \varepsilon_{L,i})\}$$

  or equivalently

$$U_W(X_i, \varepsilon_{W,i}) > U_M(X_i, \varepsilon_{M,i}) \;\&\; U_W(X_i, \varepsilon_{W,i}) > U_L(X_i, \varepsilon_{L,i})$$

Multinomial Logit Model

Multinomial Logit Model

- In general, there are $J+1$ alternatives to choose from: $\underbrace{\{0,\ldots,J\}}_{J+1 \text{ alternatives}}$.
- We assume the following linear utility functions:

$$
\begin{aligned}
U_0(X_i, \varepsilon_{0,i}) &= 0 \\
U_j(X_i, \varepsilon_{j,i}) &= X_i^\top \beta_j + \varepsilon_{j,i}, \;\; j = 1, ..., J
\end{aligned}
$$

  where $\beta_1, ..., \beta_J$ are vectors of unknown parameters to be estimated.
- It is important to note that X_i is a vector of individual characteristics, not the characteristics of the alternatives.
- The utility of choosing alternative 0 is normalized to zero.[1]

[1] Recall that in discrete choice models, only the difference of utilities matters. Thus, one of the alternatives must be fixed as a baseline of reference.

Multinomial Logit Model

- Let $D_{j,i}$ be a dummy variable that takes one when alternative j is chosen and zero otherwise.
- For simplicity of exposition, consider the case where $J = 2$ (i.e., three alternatives). Then, we can write

$$\begin{aligned} D_{1,i} &= \mathbf{1}(U_1(X_i, \varepsilon_{1,i}) > 0 \;\&\; U_1(X_i, \varepsilon_{1,i}) > U_2(X_i, \varepsilon_{2,i})) \\ &= \mathbf{1}(X_i^\top \beta_1 + \varepsilon_{1,i} > 0 \;\&\; X_i^\top \beta_1 + \varepsilon_{1,i} > X_i^\top \beta_2 + \varepsilon_{2,i}) \end{aligned}$$

- When $(\varepsilon_{1,i}, \varepsilon_{2,i})$ has a "joint" density function f, the probability of choosing alternative 1 can be written as

$$\Pr(D_{1,i} = 1) = \int\int \mathbf{1}(X_i^\top \beta_1 + \varepsilon_{1,i} > 0,\ X_i^\top \beta_1 + \varepsilon_{1,i} > X_i^\top \beta_2 + \varepsilon_{2,i}) f(\varepsilon_1, \varepsilon_2) \mathrm{d}\varepsilon_1 \mathrm{d}\varepsilon_2.$$

- How can we calculate the above multiple integration in practice?

Multinomial Logit Model

- An easiest way to solve this multiple integration is to assume that $(\varepsilon_{1,i}, \varepsilon_{2,i})$ are distributed "independent" and identically as extreme value distribution.

$$\Rightarrow \text{Multinomial Logit Model}$$

- In this case, the probability of choosing $j \in \{0, 1, 2\}$ is given by

$$\Pr(D_{0,i} = 1) = \frac{1}{1 + \exp(X_i^\top \beta_1) + \exp(X_i^\top \beta_2)}$$

and

$$\Pr(D_{j,i} = 1) = \frac{\exp(X_i^\top \beta_j)}{1 + \exp(X_i^\top \beta_1) + \exp(X_i^\top \beta_2)}$$

for $j = 1, 2$.

Multinomial Logit Model

- The above result holds for general J.
- Namely,

$$\Pr(D_{0,i} = 1) = \frac{1}{1 + \sum_{j=1}^{J} \exp(X_i^\top \beta_j)}$$

and

$$\Pr(D_{j,i} = 1) = \frac{\exp(X_i^\top \beta_j)}{1 + \sum_{j'=1}^{J} \exp(X_i^\top \beta_{j'})}$$

for $j \in \{1, \ldots, J\}$.
- One can see that the standard binary logit is a special case of the multinomial logit with $J = 1$.

Multinomial Logit Model

- Let $V_{j,i}$ be deterministic (i.e., observed) part of the utility of alternative j for individual i; that is,

$$
\begin{aligned}
V_{j,i} &= \mathrm{X}_i^\top \beta_j \quad \text{for} \quad j \in \{1, \ldots, J\} \\
V_{j,i} &= 0 \qquad\quad\; \text{for} \quad j = 0.
\end{aligned}
$$

- Then, in multinomial logit models, we can simply write

$$
\Pr(D_{j,i} = 1) = \frac{\exp(V_{j,i})}{\sum_{j'=0}^{J} \exp(V_{j',i})}, \quad \text{for all } j
$$

which states that $\Pr(D_{j,i} = 1)$ is equal to the ratio of the exponential of the observed utility for j to the sum of the exponentials of the observed utilities for all alternatives.

ML Estimation of Multinomial Logit Model

- Letting $\mathbf{D} = \{D_0, \ldots, D_J\}$, the likelihood function for $\mathbf{D}$ can be written as

$$\Pr(\mathbf{D}) = \prod_{j=0}^{J} \Pr(D_j = 1)^{D_j}.$$

- Further, let $B_0 = (\beta_1^\top, ..., \beta_J^\top)^\top$, and

$$L_{j,i}(B_0) = \frac{\exp(V_{j,i})}{\sum_{j'=0}^{J} \exp(V_{j',i})}.$$

ML Estimation of Multinomial Logit Model

- Suppose that the data of n independent observations $\{\mathbf{D}_1, X_1), ..., (\mathbf{D}_n, X_n)\}$ is available.
- Then, the likelihood function for $\{\mathbf{D}_1, ..., \mathbf{D}_n\}$ can be obtained by

$$\Pr(\mathbf{D}_1, ..., \mathbf{D}_n) = \prod_{i=1}^{n} \prod_{j=0}^{J} \Pr(D_{j,i} = 1)^{D_{j,i}} = \prod_{i=1}^{n} \prod_{j=0}^{J} L_{j,i}(B_0)^{D_{j,i}}$$

- Thus, the log-likelihood function is

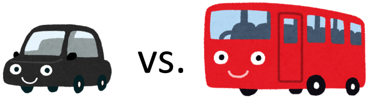
$$\ell_n(B) = \sum_{i=1}^{n} \sum_{j=0}^{J} D_{j,i} \log L_{j,i}(B),$$

and the MLE of B_0 can be obtained by

$$\widehat{B}_n = \underset{B}{\operatorname{argmax}}\, \ell_n(B).$$

Red Bus – Blue Bus Problem

- A traveler has a choice of going to work by car or taking a red bus.



- For simplicity, assume that $V_{\text{car}} = V_{\text{rb}}$, and then

$$\Pr(D_{\text{car}} = 1) = \Pr(D_{\text{rb}} = 1) = 1/2.$$

Red Bus – Blue Bus Problem

- Now suppose that a blue bus is newly introduced, which seems exactly like the red bus; that is $V_{\text{bb}} = V_{\text{rb}}$ holds.

- Then, since $V_{\text{car}} = V_{\text{rb}}$ by assumption, we must have

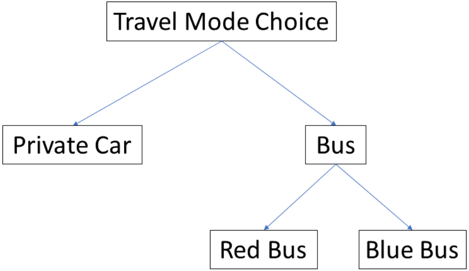
$$\text{(Multi. Logit)} \quad \Pr(D_{\text{car}} = 1) = \Pr(D_{\text{rb}} = 1) = \Pr(D_{\text{bb}} = 1) = 1/3.$$

- In reality, however, we would expect the probability of taking a car to remain the same even when a new bus is introduced that is exactly the same as the old bus:

$$\text{(Reality)} \quad \Pr(D_{\text{car}} = 1) = 1/2, \;\; \Pr(D_{\text{rb}} = 1) = \Pr(D_{\text{bb}} = 1) = 1/4.$$

Red Bus – Blue Bus Problem

- This property of multinomial logit models is called the IIA (independence from irrelevant alternatives).
- The IIA property of multinomial logit models is problematic in some practical applications.
- One way to overcome this problem is to use Nested Logit Models that present a hierarchy of decision making process.



Estimation of Multinomial Logit Models in **R**

- A practice data set: **DocVisits.csv**
  - Data on doctor visits of individuals, from the 1977–1978 Australian Health Survey.[2]
- The data csv file is available from my website or from **Course Navi**.
- Set your working directory appropriately, and import the csv file by `read.csv()`:

```
setwd("C:/Rdataset")
data <- read.csv("DocVisits.csv")
```

[2] This data is taken from the **R** package **AER** that contains various data sets for applied econometrics.

Estimation of Multinomial Logit Models in **R**

```
>  data <- read.csv("DocVisits.csv")
>  head(data)
  visits gender age income illness private freepoor freerepat
1   once      0  19   0.55       1       1        0         0
2   once      0  19   0.45       1       1        0         0
3   once      1  19   0.90       3       0        0         0
4   once      1  19   0.15       1       0        0         0
5   once      1  19   0.45       2       0        0         0
6   once      0  19   0.35       5       0        0         0
>  dim(data)
[1] 5190    8
```

Estimation of Multinomial Logit Models in **R**

Definitions of variables

Dependent variable (1st column)

- visits: number of doctor visits in past 2 weeks: [none, once, twice, mttwice (more than twice)].

Explanatory variables (2nd - 8th columns)

- gender: 1 = male, 0 = female .
- age: age in years.
- income: annual income in tens of thousands of dollars.
- illness: number of illnesses in past 2 weeks.
- private: 1 = the individual has private health insurance, 0 otherwise.
- freepoor: 1 = the individual has free government health insurance due to low income, 0 otherwise.
- freerepat: 1 = the individual has free government health insurance due to old age, disability or veteran status, 0 otherwise.

Estimation of Multinomial Logit Models in R

- The data class of the dependent variable is "factor" (i.e., text data). The levels in the factor are by default ordered alphabetically.

```
>  class(data$visits)
[1] "factor"
>  levels(data$visits)
[1] "mttwice" "none"    "once"    "twice"
```

- Thus, we need to re-order the data appropriately. Run the following command:

```
data$visits <- ordered(data$visits, levels = c("none",
                 "once", "twice", "mttwice"))
```

Then,

```
>  data$visits <- ordered(data$visits, levels = c("none",
+ "once", "twice", "mttwice"))
>  levels(data$visits)
[1] "none"    "once"    "twice"   "mttwice"
```

- To see the empirical distribution of factor data, it is convenient to use the `table()` function:

```
table(data$visits)

none     once    twice mttwice
4141      782      174      93
```

Estimation of Multinomial Logit Models in R

- Unfortunately, there is no built-in function in **R** for performing multinomial logit estimation. But there are several "packages" that can do this.
- The package used here is **nnet**.[3]
- Run the following code:

```
install.packages("nnet")
library(nnet)
```

  - The first line installs the package to your computer. Once the package is installed, this step can be skipped next time.
  - The second line loads the package.

[3] **nnet** is a package for neural networks. We will study neural networks in a later lecture.

Estimation of Multinomial Logit Models in R

- Once the **nnet** package is loaded, we can use the function `multinom()` for estimating a multinomial logit model.
- Multinomial logit estimation:

```
MLM <- multinom(visits ~ gender + age + income + illness +
       private + freepoor + freerepat, data)
summary(MLM)
```

* Alternatively, you may use a shorthand syntax:

```
MLM <- multinom(visits ~ ., data)
summary(MLM)
```

Estimation of Multinomial Logit Models in R

```
>  MLM <- multinom(visits ~ gender + age + income + illness +
+ private + freepoor + freerepat, data)
# weights:  36 (24 variable)
initial  value 7194.867734
iter  10 value 3445.683921
iter  20 value 3227.973953
iter  30 value 3160.026797
final  value 3159.580243
converged
>  summary(MLM)
Call:
multinom(formula = visits ~ gender + age + income + illness +
    private + freepoor + freerepat, data = data)

Coefficients:
        (Intercept)      gender          age      income   illness
once      -2.608607 -0.28529863  0.006843832  0.08482831 0.3236239
twice     -4.047962 -0.25572841 -0.003494325 -0.51883679 0.5047092
mttwice   -4.782943  0.04177185  0.014951810 -0.45379423 0.4962996
           private   freepoor  freerepat
once     0.2190271 -0.5075607  0.4849194
twice    0.6888452 -1.0284469  0.7944516
mttwice -0.3961391 -0.5435642 -0.5635826
```

- `final value` reports the value of the log-likelihood $\ell_n(\widehat{B}_n)$.
- The t-values are not automatically reported.

- Computation of the t-values:

```
coef  <- summary(MLM)$coefficients
se    <- summary(MLM)$standard.errors
tvals <- coef/se
tvals
```

Estimation of Multinomial Logit Models in R

```
>  coef  <- summary(MLM)$coefficients
>  se    <- summary(MLM)$standard.errors
>  tvals <- coef/se
>  tvals
        (Intercept)     gender        age     income   illness   private
once      -16.58557 -3.2156623  2.6406582  0.6488521 11.706545  2.000995
twice     -12.70876 -1.4617640 -0.6977197 -1.8904315  9.888329  2.964953
mttwice   -11.63618  0.1813828  2.2092411 -1.2508077  7.348468 -1.401619
         freepoor freerepat
once    -1.771079  3.256177
twice   -1.385174  2.631045
mttwice -0.867343 -1.511276
```

- As expected, `illness` has a strong positive impact on the number of doctor visits.
- The availability of insurances (`private` and `freerepat`) affects only on `[once, twice]`.
    - Individuals in the `[mttwice]` category, who (probably) have serious illness, need to see a doctor regardless of the availability of insurances.

Estimation of Multinomial Logit Models in R

- By using the `predict()` function, we can compute the predicted probability of choosing each alternative for each individual.

```
Pr_MLM <- predict(MLM, type = "probs")
```

- For example, the choice probabilities for the first six individuals are

```
> Pr_MLM <- predict(MLM, type = "probs")
> head(Pr_MLM)
       none      once      twice     mttwice
1 0.8324507 0.12585193 0.03372415 0.007973222
2 0.8315335 0.12465138 0.03548094 0.008334185
3 0.8000786 0.14374690 0.02884112 0.027333333
4 0.8885447 0.07841851 0.01722542 0.015811354
5 0.8488116 0.10620644 0.02332899 0.021652956
6 0.5816952 0.25345378 0.09884428 0.066006792
```

Ordered Choice Model

Ordered Choice Model

- In the above empirical analysis, there is a clear order among the dependent variables:

$$\text{(0 visits)} \leq \text{(once)} \leq \text{(twice)} \leq \text{(more than 2 visits)}$$

- On the other hand, in the example of the OS choice, there is no clear ordering between Windows, Mac and Linux.
- Remember that, in the multinomial logit model, the error terms $(\varepsilon_1, \ldots, \varepsilon_J)$ are assumed to be independent.
- This implies that the dependent variable $(D_0, D_1, \ldots, D_J)$ are also independent (after controlling for X).
  $\Rightarrow$ such assumption is unreasonable for ordered dependent variables.

Ordered Choice Model

- Suppose that there are J ordered alternatives $\{1, \ldots, J\}$, and let D_i denote the alternative chosen by individual i.
  - Here, D_i is not a dummy variable, but is a discrete random variable that takes a value from $\{1, \ldots, J\}$.
- Let Y_i^* a latent (unobservable) dependent variable defined by

$$Y_i^* = X_i^\top \beta + \varepsilon_i.$$

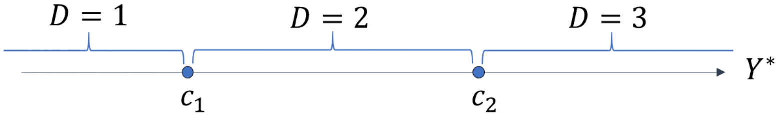
  The value of D_i is uniquely determined by Y_i^*.
- Examples:

  - Rating system: $D =$ [Excellent/Good/Fair/Poor], $Y^* =$ satisfaction level.
  - Course grades: $D =$ [A/B/C/D/F], $Y^* =$ raw test score.
  - Frequency of doctor visits: $D =$ [0 to once a month/once a week/ almost everyday], $Y^* =$ health status.

Ordered Choice Model

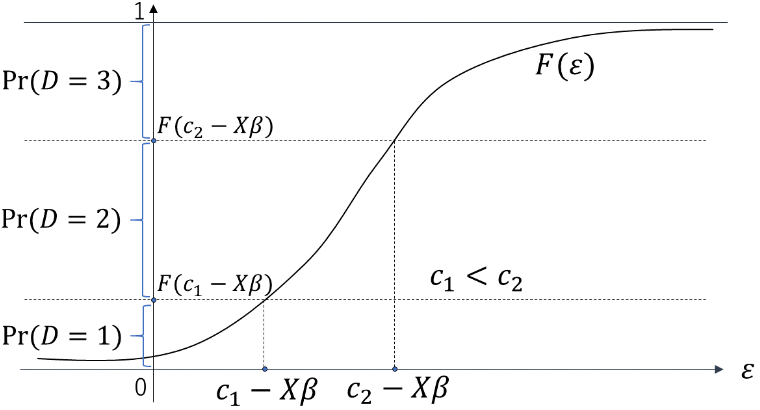
- Assume that there are threshold parameters $(c_1, \dots, c_{J-1})$ such that

$$
\begin{aligned}
D_i = 1 &\iff Y_i^* \leq c_1 \\
D_i = j &\iff c_{j-1} < Y_i^* \leq c_j \ \text{ for } j = 2, \dots, J-1 \\
D_i = J &\iff c_{J-1} < Y_i^*.
\end{aligned}
$$



- Then, letting the distribution function of ε be $F(\cdot)$, the probability that i will chose alternative j is

$$
\begin{aligned}
\Pr(D_i = 1) &= F(c_1 - X_i^\top \beta) \\
\Pr(D_i = j) &= F(c_j - X_i^\top \beta) - F(c_{j-1} - X_i^\top \beta) \ \text{ for } j = 2, \dots, J-1 \\
\Pr(D_i = J) &= 1 - F(c_{J-1} - X_i^\top \beta).
\end{aligned}
$$


1
$F(\varepsilon)$
$\Pr(D=3)$
$F(c_2 - X\beta)$
$\Pr(D=2)$
$F(c_1 - X\beta)$
$c_1 < c_2$
$\Pr(D=1)$
0
$c_1 - X\beta$
$c_2 - X\beta$
ε

ML Estimation of Ordered Choice Model

- The likelihood function for D_i can be written as

$$\Pr(D_i) = \prod_{j=1}^{J} \Pr(D_i = j)^{\mathbf{1}\{D_i = j\}}.$$

- Let $\theta_0 = (c_1, \ldots, c_{J-1}, \beta^\top)^\top$ be the vector of unknown parameters to be estimated. Further let $c_0 = -\infty$ and $c_J = \infty$, and define

$$L_{j,i}(\theta_0) = F(c_j - X_i^\top \beta) - F(c_{j-1} - X_i^\top \beta)$$

for $j = 1, \ldots, J$.

ML Estimation of Ordered Choice Model

- Suppose that the data of n independent observations $\{(D_1, X_1), ..., (D_n, X_n)\}$ is available.
- Then, the likelihood function for $\{D_1, ..., D_n\}$ can be obtained by

$$\Pr(D_1, ..., D_n) = \prod_{i=1}^{n}\prod_{j=1}^{J}\Pr(D_i = j)^{\mathbf{1}\{D_i=j\}} = \prod_{i=1}^{n}\prod_{j=1}^{J}L_{j,i}(\theta_0)^{\mathbf{1}\{D_i=j\}}$$

- Thus, the log-likelihood function is

$$\ell_n(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{J}\mathbf{1}\{D_i = j\}\log L_{j,i}(\theta),$$

and the MLE of θ_0 can be obtained by

$$\widehat{\theta}_n = \underset{\theta}{\operatorname{argmax}}\, \ell_n(\theta).$$

ML Estimation of Ordered Choice Model

- The distribution function $F(\cdot)$ is usually assumed to be either logistic or standard normal:
  - Logistic $F(\cdot)$ = Ordered Logit Model
  - Standard normal $F(\cdot)$ = Ordered Probit Model
- When estimating the threshold parameters $(c_1, \ldots, c_{J-1})$, the order restriction $c_{j-1} < c_j$ must be maintained. This introduces some computational difficulty.
- One way to mitigate the computational complexity is to re-parameterize c_j as $c_j = c_{j-1} + a_j$ and estimate a_j (instead of c_j) with a positivity constraint.

Estimation of Ordered Logit Models in R

Estimation of Ordered Logit Models in **R**

- We again use the dataset: **DocVisits.csv**
- If you have closed the console window, launch **R**, and read the csv file again.
- The estimation of the ordered logit model can be performed by `polr()` function[4] in the **MASS** package.
- Install and load the package by typing

```
install.packages("MASS")
library(MASS)
```

[4] `polr` stands for "proportional odds logistic regression".

Estimation of Ordered Logit Models in R

- Ordered logit estimation:

```
OLM <- polr(visits ~ gender + age + income + illness +
       private + freepoor + freerepat, data)
summary(OLM)
```

- Note that the variable `visits` needs to be ordered appropriately, as above.

Estimation of Ordered Logit Models in R

```
>  OLM <- polr(visits ~ gender + age + income + illness +
+ private + freepoor + freerepat, data)
>  summary(OLM)
Coefficients:
              Value Std. Error t value
gender    -0.230573   0.078565  -2.935
age        0.005889   0.002284   2.579
income    -0.090273   0.116937  -0.772
illness    0.376517   0.024438  15.407
private    0.241868   0.096560   2.505
freepoor  -0.589381   0.252500  -2.334
freerepat  0.394136   0.131747   2.992

Intercepts:
              Value   Std. Error t value
none|once      2.2852  0.1388    16.4612
once|twice     3.9163  0.1508    25.9763
twice|mttwice  5.0288  0.1730    29.0658
```

- The estimates for the thresholds c_j's are reported in `Intercepts`.
- The log-likelihood value $\ell_n(\widehat{\theta}_n)$ can be checked by `logLik(OLM)`, and it is equal to -3173.213.

Estimation of Ordered Logit Models in R

- Similarly as above, we can compute the predicted choice probabilities for each individual by the function `predict()`:

```
Pr_OLM <- predict(OLM, type = "probs")
```

```
> Pr_OLM <- predict(OLM, type = "probs")
> head(Pr_OLM)
       none       once      twice    mttwice
1 0.8326530 0.12950069 0.02508147 0.012764883
2 0.8313913 0.13043226 0.02529730 0.012879147
3 0.7950354 0.15693161 0.03171789 0.016315137
4 0.8850227 0.09018142 0.01650716 0.008288673
5 0.8444139 0.12077997 0.02309118 0.011714934
6 0.5798825 0.29593257 0.07965123 0.044533653
```

- Although MLM and OLM assume different decision-making rules, they gave very similar prediction results (not in general, but for this particular data set).