Binary Choice Model

Tadao Hoshino （星野匡郎）

Econometrics II: ver. 2019 Spring Semester

Random Utility

Example 1: Which OS to buy?

- Suppose that you are buying a new PC. You can choose the OS (Operating System) for your PC from Windows or Mac.
- You obtain utilities U_W and U_M from Windows and Mac, respectively.
- It would be natural to assume that a Windows PC is chosen if and only if $U_W > U_M$ (without considering a “tie”).

Utility: U_W

Utility: U_M

$U_W > U_M \Rightarrow$ Windows PC will be chosen
$U_W < U_M \Rightarrow$ Mac PC will be chosen

Example 1: Which OS to buy?

- The utility obtained from buying Windows/Mac may differ between individuals; for example
  - graphic designers tend to have strong preference on Mac PCs,
  - while heavy computer game lovers may have strong preference on Windows (because many PC games are not supported by Mac).
- For an individual i, let X_i be a vector of "observable" individual characteristics, and ε_i be an "unobservable" random variable.[1]
- Then, i's utility of buying Windows and Mac can be written as

$$U_{W,i} = U_W(X_i, \varepsilon_{W,i}) \text{ and } U_{M,i} = U_M(X_i, \varepsilon_{M,i}),$$

respectively. So i will choose Windows if

$$U_W(X_i, \varepsilon_{W,i}) > U_M(X_i, \varepsilon_{M,i}).$$

[1] ε is unobservable to researchers, but of course i knows its value.

Random Utility Model

- Unlike the traditional economic model of consumer demand, it is allowed that the researchers cannot fully observe the variables that determine the utility.
- The utility can depend on an unobservable random variable:

$$U = U(X, \varepsilon),\ X\text{: observable, } \varepsilon\text{: unobservable}$$

- This framework is called Random Utility Model (RUM).

Daniel McFadden: Based on the random utility framework, McFadden developed a set of econometric methods for analyzing discrete choice behavior. He was awarded the Nobel Prize in economics for this work.

Binary Choice Model

- Suppose there are two alternatives to choose from: 1 or 2.
- For individual i, the utility from alternative 1 and that from 2 are denoted by $U_1(X_i, \varepsilon_{1,i})$ and $U_2(X_i, \varepsilon_{2,i})$, respectively.
- For simplicity, we assume the following "linear" RUM framework:

$$U_1(X_i, \varepsilon_{1,i}) = X_i^\top \beta_1 + \varepsilon_{1,i}$$
$$U_2(X_i, \varepsilon_{2,i}) = X_i^\top \beta_2 + \varepsilon_{2,i}$$

where β_1 and β_2 are unknown parameters.[2]

NOTE: We cannot observe the utility itself. (If the utilities can be observed in one way or another, simple linear regression suffices our purpose.)

[2] β_1 and β_2 correspond to the marginal utility of X for alternative 1 and 2, respectively.

Binary Choice Model

- Let D_i be a dummy variable that takes one when alternative 1 is chosen, and zero when 2 is chosen.
- Since alternative 1 is chosen when $U_1(X_i, \varepsilon_{1,i}) > U_2(X_i, \varepsilon_{2,i})$, we can write

$$\begin{aligned} D_i &= \mathbf{1}(X_i^\top \beta_1 + \varepsilon_{1,i} > X_i^\top \beta_2 + \varepsilon_{2,i}) \\ &= \mathbf{1}(X_i^\top (\beta_1 - \beta_2) > \varepsilon_{2,i} - \varepsilon_{1,i}), \end{aligned}$$

where $\mathbf{1}(\cdot)$ is the indicator function.

- Further, if we define $\beta_0 = \beta_1 - \beta_2$ and $\varepsilon_i = \varepsilon_{2,i} - \varepsilon_{1,i}$, we have

$$D_i = \mathbf{1}(X_i^\top \beta_0 > \varepsilon_i).$$

- This is equivalent to setting the utility of 1 to $X_i^\top \beta_0 - \varepsilon_i$ and normalizing the utility of 2 to zero:

$$\begin{aligned} U_1(X_i, \varepsilon_{1,i}) &= X_i^\top \beta_1 + \varepsilon_{1,i} \\ U_2(X_i, \varepsilon_{2,i}) &= X_i^\top \beta_2 + \varepsilon_{2,i} \end{aligned} \iff \begin{aligned} U_1(X_i, \varepsilon_{1,i}) &= X_i^\top \beta_0 - \varepsilon_i \\ U_2(X_i, \varepsilon_{2,i}) &= 0 \end{aligned}$$

<u>Only the difference of utilities matters.</u>

  - This observation implies that we "cannot" estimate β_1 and β_2 separately. What we can estimate is the difference of the marginal utilities, β_0.

Binary Choice Model

- Suppose that ε_i has a continuous distribution function F. Then,

$$\Pr(D_i = 1) = \Pr(\varepsilon_i < X_i^\top \beta_0) = F(X_i^\top \beta_0),$$
$$\Pr(D_i = 0) = \Pr(\varepsilon_i \geq X_i^\top \beta_0) = 1 - F(X_i^\top \beta_0).$$

- In order to estimate the model, we need to specify the functional form of F.
- A simplest way is to assume that F is a linear function such that

$$F(X_i^\top \beta_0) = X_i^\top \gamma_0$$

for some parameter vector γ_0. Such a model is called the Linear Probability Model.

Binary Choice Model

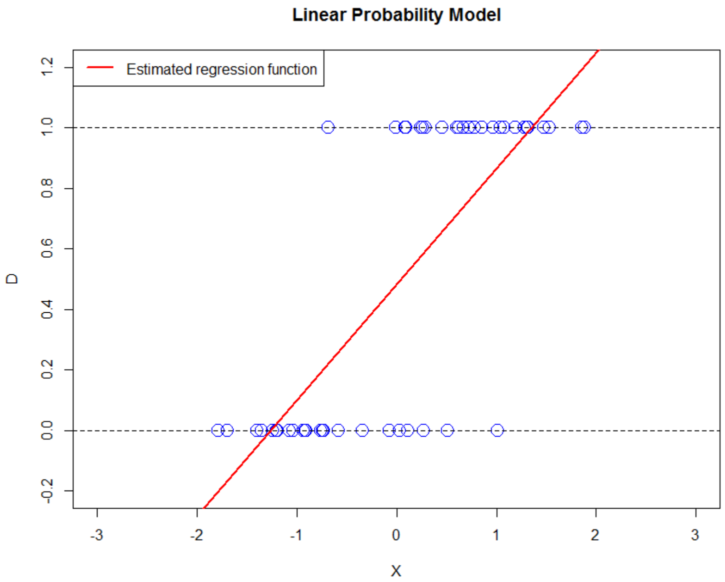
- The linear probability model can be estimated simply by regressing D on X, i.e., the OLS estimation with D being the dependent variable and X being the explanatory variables:

$$D_i = X_i^\top \gamma_0 + u_i \;\; (i = 1, \ldots, n).$$

- Although its implementation is easy, the linear probability model has a serious problem:

  the predicted "probability" of $D = 1$, i.e., $X^\top \widehat{\gamma}_n$, is not actually a probability — that is, it is not necessarily lying between 0 and 1.

Binary Choice Model



- Thus, we should specify F to be an increasing function that takes its value on $[0,1]$ (which is an requirement for F to be a distribution function).
- F is often assumed to be a logistic function:

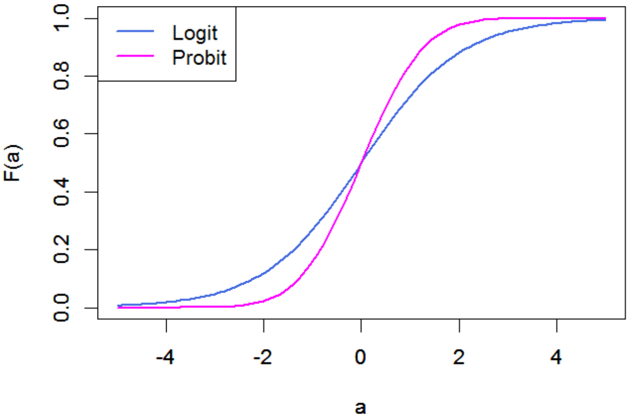
$$F(a) = \frac{\exp(a)}{1+\exp(a)} \text{ or equivalently, } F(a) = \frac{1}{1+\exp(-a)}$$

  In this case, the model is called the Logit Model.
- The standard normal distribution function is also often used for F:

$$F(a) = \int^{a} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

  In this case, the model is called the Probit Model.

Binary Choice Model



ML Estimation of Logit Models

Maximum Likelihood Estimation

- Suppose that we have data of n observations $\{(D_1, X_1), ..., (D_n, X_n)\}$ which are independent and identically distributed.
- As in the likelihood function for the Bernoulli distribution (recall the previous lecture), the likelihood function for $\{D_1, ..., D_n\}$ (conditional on X) is given by

$$\begin{aligned} \Pr(D_1, ..., D_n) &= \prod_{i=1}^{n} \Pr(D_i = 1)^{D_i} \Pr(D_i = 0)^{1-D_i} \\ &= \prod_{i=1}^{n} F(X_i^\top \beta_0)^{D_i} [1 - F(X_i^\top \beta_0)]^{1-D_i}. \end{aligned}$$

Maximum Likelihood Estimation

- Thus, the log-likelihood function is

$$\begin{aligned}\ell_n(\beta_0) &= \log \Pr(D_1, ..., D_n) \\ &= \sum_{i=1}^{n} \left\{ D_i \log F(X_i^\top \beta_0) + (1 - D_i) \log[1 - F(X_i^\top \beta_0)] \right\}\end{aligned}$$

- The MLE of β_0 can be obtained by

$$\widehat{\beta}_n = \operatorname*{argmax}_{\beta} \ell_n(\beta).$$

- Under the standard conditions, $\widehat{\beta}_n$ is a consistent estimator of β_0.
- How to compute the standard errors is beyond the scope of this class.

Interpretation of the Estimated Coefficients

Interpretation of the Estimated Coefficients

- Recall that in the case of linear regression model, the estimated coefficients represent the marginal effects of the corresponding explanatory variables on the dependent variable. Namely,

$$Y = X^\top \beta_0 + \varepsilon$$
$$\Rightarrow \quad \beta_0 = \frac{\partial Y}{\partial X}$$

- This interpretation of β_0 is <u>not</u> applied to discrete choice models.
    - $D_i = \mathbf{1}(X_i^\top \beta_0 > \varepsilon_i)$ is not differentiable with respect to X.
- In binary choice models, we measure the impacts of X by how a unit increase in X relates to a change in the probability $\Pr(D = 1)$.

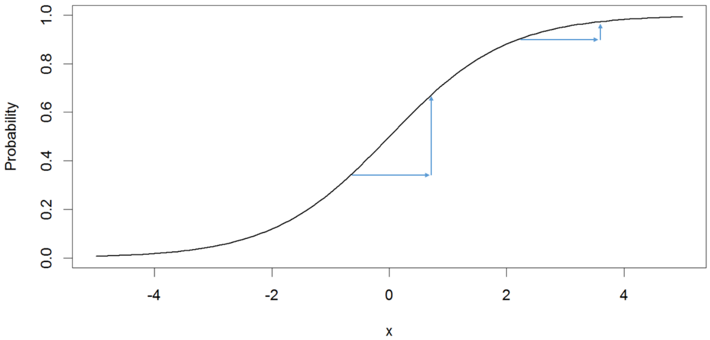
Interpretation of the Estimation Results

- Since $\Pr(D=1) = F(X^\top \beta_0)$, the marginal effect of X_j (j-th component of X) on the choice probability is given by

$$\frac{\partial F(X^\top \beta_0)}{\partial X_j} = f(X^\top \beta_0)\beta_{0,j}$$

  where $\beta_{0,j}$ is the j-th component of β_0, and $f(\cdot)$ is the density function of ε (= the derivative of $F(\cdot)$).
- This implies that the marginal effect is not constant but varies with the magnitude of $f(X^\top \beta_0)$.
  - Note that for logistic (logit) and standard normal distribution function (probit) the density $f(\cdot)$ takes the largest value at 0.
  - Hence, the marginal effect is biggest when $X^\top \beta_0 = 0$, i.e., when $\Pr(D=1) = \Pr(D=0) = 0.5$.

Interpretation of the Estimation Results



- For an individual who is wavering between $D = 1$ and $D = 0$, even a unit increase in X has a certain impact on her choice.
- On the other hand, for an individual who has a strong preference on $D = 1$, a unit increase in X has only a negligible impact.

Interpretation of the Estimation Results

- When the MLE $\widehat{\beta}_n$ of β_0 is available, the marginal effect of X_j for individual i can be estimated by

$$f(X_i^\top \widehat{\beta}_n)\widehat{\beta}_{n,j}.$$

- The sample average of the marginal effects

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i^\top \widehat{\beta}_n)\widehat{\beta}_{n,j}$$

  is called the Average Marginal Effect (AME) of X_j.
- AME = to what extent a unit increase in X_j affects the probability $\Pr(D=1)$ on average.

Goodness of Fit

Likelihood Ratio Index

- In a linear regression model, we can use the R^2 statistic (coefficient of determination) to evaluate "goodness of fit" of the estimated regression model.
- For discrete choice models, we often use the Likelihood Ratio Index (LRI), often referred to as McFadden's pseudo R^2, to evaluate how well the estimated model fits the data:

$$\rho = 1 - \frac{\ell_n(\widehat{\beta}_n)}{\ell_{n,\mathbf{0}}},$$

where $\ell_{n,\mathbf{0}}$ is the log-likelihood when all the parameters, excluding a constant term, are restricted to zero.

NOTE: $\ell_{n,\mathbf{0}}$ corresponds to the worst case log-likelihood; the model's explanatory variables have no explanatory power at all.

Likelihood Ratio Index

- The interpretation of the likelihood ratio index ρ is similar to that of R^2 statistic.
- Since $\widehat{\beta}_n$ is the "maximum likelihood" estimate, $\ell_n(\widehat{\beta}_n)$ cannot be less than $\ell_{n,\mathbf{0}}$. Here, note that the log-likelihood is always negative (because log-probability is negative).

$$\underbrace{\ell_{n,\mathbf{0}}}_{\text{worst case}} \leq \ell_n(\widehat{\beta}_n) \leq \underbrace{0}_{\text{best case}}$$

  Thus, the lower bound of ρ is 0.
- When the estimated model predicts the data with 100% accuracy, the value of the likelihood function (probability) is one. Since the log of one is zero, $\ell_n(\widehat{\beta}_n) = 0$. Thus, the upper bound of ρ is 1.
  $=> 0 \leq \rho \leq 1$

Estimation of Binary Choice Models in **R**

- A practice data set: **flabor.csv**
  - Data on married women's labor force participation in the US.[3]
- 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("flabor.csv")
```

[3] The data is taken from: Mroz, T. (1987) The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. Econometrica.

Estimation of Binary Choice Models in R

ファイル　編集　その他　パッケージ　ウインドウ　ヘルプ

```
> data <- read.csv("flabor.csv")
> head(data)
  labor kids6 kids18 age educ husage huseduc huswage
1     1     1      0  32   12     34      12  4.0288
2     1     0      2  30   12     30       9  8.4416
3     1     1      3  35   12     40      12  3.5807
4     1     0      3  34   12     53      10  3.5417
5     1     1      2  31   14     32      12 10.0000
6     1     0      0  54   12     57      11  6.7106
> dim(data)
[1] 753   8
> 
```

Definitions of variables

Dependent variable (1st column)

- labor: labor-force participation dummy; 1 if yes, 0 otherwise.

Explanatory variables (2nd - 8th columns)

- kids6: number of children younger than 6 years
- kids18: number of children 6 - 18 years old
- age: wife's age in years
- educ: wife's education in years.
- husage: husband's age in years
- huseduc: husband's education in years
- huswage: husband's hourly wage

Estimation of Binary Choice Models in R

- To estimate a logit (and also probit) model in **R**, we can use the `glm()` function.
- `glm()` can be used in a similar way as `lm()`.

```
logit_full <- glm(labor ~ kids6 + kids18 + age + educ + husage +
    huseduc + huswage, data, family = binomial(link = "logit"))
summary(logit_full)
```

Alternatively, you can use the following shorthand syntax:

```
logit_full <- glm(labor ~ ., data, family = binomial(link = "logit"))
```

- If you replace `"logit"` with `"probit"`, then a probit model will be estimated.

```
>  logit_full <- glm(labor ~ kids6 + kids18 + age + educ + husage +
+ huseduc + huswage, data, family = binomial(link = "logit"))
>  summary(logit_full)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.47909    0.83782   1.765  0.07750 .
kids6       -1.50567    0.19883  -7.573 3.65e-14 ***
kids18      -0.09086    0.06734  -1.349  0.17723
age         -0.04143    0.02197  -1.886  0.05928 .
educ         0.27503    0.04719   5.828 5.62e-09 ***
husage      -0.02642    0.02199  -1.201  0.22957
huseduc     -0.05517    0.03543  -1.557  0.11948
huswage     -0.05746    0.02112  -2.720  0.00652 **
```

- Significant at the 1% level: `kids6` $\ominus$, `educ` $\oplus$, `huswage` $\ominus$.
- Significant at the 10% level: `age` $\ominus$.

Estimation of Binary Choice Models in R

- In order to improve the interpretability of the estimated coefficients, we compute the AME for each explanatory variable. (Unfortunately, `glm()` does not return the AME.)
- The value of $X_i^\top \widehat{\beta}_n$ for each i can be obtained by using the `predict()` command:

```
Xb <- predict(logit_full)
```

- Then, compute the AME as follows:

```
AME <- mean(dlogis(Xb))*logit_full$coef
```

  - `dlogis()`: logistic density function
  - `logit_full$coef`: $\widehat{\beta}_n$

Estimation of Binary Choice Models in R

```
>  Xb   <- predict(logit_full)
>  AME  <- mean(dlogis(Xb))*logit_full$coef
>  AME
 (Intercept)        kids6       kids18          age
 0.310928726 -0.316517692 -0.019099925 -0.008710080
        educ       husage      huseduc      huswage
 0.057816284 -0.005553466 -0.011597087 -0.012078399
```

- A one-point increase in `kids6` decreases the probability of labor force participation by about 32%.
- A one-year increase in `educ` increases the probability of labor force participation by about 6%.
- A one USD increase in `huswage` decreases the probability of labor force participation by about 1%.

- Finally, we check the value of Likelihood Ratio Index (LRI).
- The log-likelihood of the estimated model, $\ell_n(\widehat{\beta}_n)$, can be obtained by `logLik(logit_full)`.
- In order to compute the LRI, we need to estimate a model with only a constant term:

```
logit_0 <- glm(labor ~ 1, data, family = binomial(link = "logit"))
```

- Then, by the definition of LRI

```
LRI <- 1 - logLik(logit_full)/logLik(logit_0)
```

Estimation of Binary Choice Models in R

```
>  logit_0 <- glm(labor ~ 1, data, family = binomial(link = "logit"))
>  LRI <- 1 - logLik(logit_full)/logLik(logit_0)
>  LRI
'log Lik.' 0.1114908 (df=8)
```

- The value of LRI is not sufficiently high.
- Here, `df` means the degree of freedom (the number of explanatory variables including the intercept term).