

Intermediate Econometrics

17/12/2025 - Vincenzo Gioia

Discrete choice model

Introduction

- **Cases where the response is the choice of an alternative among a set of mutually exclusive alternatives**
- This choice can be modeled in a utility maximization framework, which means that we hypothesize that the individual chooses the alternative which corresponds to the maximum level of utility
- Namely, denoting with $j = 1, \dots, J$ the set of alternatives, we denote with U_{ij} the level of utility of individual i if he/she chooses alternative j and the j^{th} alternative will be chosen if $U_{ij} > U_{ik} \forall k \neq j$
- The level of utility will depend on some observable covariates and on some other unobservables whose effect will be summarized in a variable ϵ_{nj} that will be considered, from the researcher's point of view, as the realization of a random variable
- Two key questions for these models are:
 1. how the covariates enter the utility function
 2. what is the assumed distribution of the error

Discrete choice model

Data Management

- Data sets used for discrete choice models estimation concern some individuals, who make one or a sequential choice of one alternative among a set of mutually exclusive alternatives
- The determinants of these choices are covariates that can depend on the alternative and the choice situation, only on the alternative or only on the choice situation
- Data sets can have two different shapes:
 1. a *wide* shape (one row for each choice situation)
 2. a *long* shape (one row for each alternative and, therefore, as many rows as there are alternatives for each choice situation)
- **mlogit** (the R package we are working with) deals with both formats
- It depends on the **dfidx** package with two main arguments
 1. The first argument is the `data.frame` and returns a `dfidx` object, which is a `data.frame` in “long” format with a special data frame column which contains the indexes
 2. The second argument is called `idx`. In its simple use, it should be a list (or a vector) of two characters containing the choice situation and the alternative indexes

Discrete choice model

Wide format

- `dutch_railways` is an example of a *wide* data set
- This data set contains data about a stated preference survey in the Netherlands in 1987
- Each individual has responded to several (up to 16) scenarios

```
1 library(miscr)
2 dutch_railways <- as.data.frame(miscr.data::dutch_railways)
3 head(dutch_railways)
```

```
  id choiceid choice  price_A  price_B  time_A  time_B  change_A  change_B
1  1         1      A 10.89073 18.15121 2.500000 2.500000          0          0
2  1         2      A 10.89073 14.52097 2.500000 2.166667          0          0
3  1         3      A 10.89073 18.15121 1.916667 1.916667          0          0
4  1         4      B 18.15121 14.52097 2.166667 2.500000          0          0
5  1         5      B 10.89073 14.52097 2.500000 2.500000          0          0
6  1         6      B 18.15121 10.89073 1.916667 2.166667          0          0
  comfort_A  comfort_B
1          1          1
2          1          1
3          1          0
4          1          0
5          1          0
6          1          0
```

Discrete choice model

Data description

- For every scenario, two train trips are proposed to the user, with different combinations of four attributes:
 1. **price** (the price in euros)
 2. **time** (travel time in minutes)
 3. **change** (the number of changes)
 4. **comfort** (the class of comfort, 0, 1 or 2; 0 being the most comfortable class)
- This “wide” format is suitable to store choice situation (or individual specific) variables because, in this case, they are stored only once in the data
- It is cumbersome for alternative-specific variables because there are as many columns for such variables as there are alternatives

Discrete choice model

Data management

- For such a wide data set, the `shape` argument of `dfidx` is mandatory, as its default value is "long"
- The alternative-specific variables are indicated with the `varying` argument which is a numeric vector that indicates their position in the data frame
- As the names of the variables are of the form `price_A`, one must add `sep = "_"` (the default value being ".")
- To take the panel dimension into account, `idx` is a list of length 1 (the choice situation) containing a vector of length 2 with `choiceid` and `id`
- The `idnames` is used to give a relevant name for the second index, the `NA` in first position indicating that the name of the first index is unchanged
- Note the use of the `opposite` argument for the four covariates: we expect negative coefficients for all of them, taking the opposite of the covariates will lead to expected positive coefficients

```
1 library(dfidx)
2 Tr <- dfidx(dutch_railways, shape = "wide", varying = 4:11, sep = "_",
3           idx = list(c("choiceid", "id")), idnames = c(NA, "alt"),
4           opposite = c("price", "time", "change", "comfort"))
```

Discrete choice model

```
1 head(Tr)
```

```
~~~~~
```

```
first 10 observations out of 5858
```

```
~~~~~
```

	choice	price	time	change	comfort	idx
1	A	-10.89073	-2.500000	0	-1	1:A
2	A	-18.15121	-2.500000	0	-1	1:B
3	A	-10.89073	-2.500000	0	-1	2:A
4	A	-14.52097	-2.166667	0	-1	2:B
5	A	-10.89073	-1.916667	0	-1	3:A
6	A	-18.15121	-1.916667	0	0	3:B
7	B	-18.15121	-2.166667	0	-1	4:A
8	B	-14.52097	-2.500000	0	0	4:B
9	B	-10.89073	-2.500000	0	-1	5:A
10	B	-14.52097	-2.500000	0	0	5:B

```
~~~ indexes ~~~~
```

```
choiceid id alt
```

Discrete choice model

Long format

- `toronto_montreal` is an example of a data set in long format
- It presents the choice of individuals for a transport mode for the Toronto-Montreal corridor in 1989
- There are four transport modes (`air`, `train`, `bus` and `car`) and most of the variables are alternative-specific (`cost` for monetary cost, `ivt` for in-vehicle time, `ovt` for out-vehicle time, `freq` for frequency)
- The only choice situation-specific variables are `dist` (distance of the trip), `income` (household income), `urban` (a dummy for trips which have a large city at the origin or the destination) and `noalt` (the number of available alternatives)
- The advantage of this shape is that there are much fewer columns than in the wide format, the caveat being that values of `dist`, `income` and `urban` are repeated up to four times

```
1 toronto_montreal <- as.data.frame(micsr.data::toronto_montreal)
2 head(toronto_montreal)
```

	case	alt	choice	dist	cost	ivt	ovt	freq	income	urban	noalt
1	1	train	0	83	28.25	50	66	4	45	0	2
2	1	car	1	83	15.77	61	0	0	45	0	2
3	2	train	0	83	28.25	50	66	4	25	0	2
4	2	car	1	83	15.77	61	0	0	25	0	2
5	3	train	0	83	28.25	50	66	4	70	0	2
6	3	car	1	83	15.77	61	0	0	70	0	2

Discrete choice model

Long format

- For data in “long” format, the `shape` and the `choice` arguments are no longer mandatory
- To replicate published results, we use only a subset of the choice situations, namely those for which the four alternatives are available
- This can be done using the `subset` function with the `subset` argument set to `noalt == 4` while estimating the model. This can also be done within `dfidx`, using the `subset` argument

```
1 MC <- dfidx(toronto_montreal, subset = noalt == 4,  
2           idx = c("case", "alt"), drop.index = FALSE)
```

Discrete choice model

Long format

- The information about the structure of the data can be explicitly indicated using choice situations and alternative indexes (respectively `case` and `alt` in this data set) or, in part, guessed by the `dfidx` function
- Here, after subsetting, we have 2779 choice situations with 4 alternatives, and the rows are ordered first by choice situation and then by alternative (`train`, `air`, `bus`, and `car` in this order)

```
1 head(MC)
```

```
~~~~~
```

```
first 10 observations out of 11116
```

```
~~~~~
```

	case	alt	choice	dist	cost	ivt	ovt	freq	income	urban	noalt	idx
1	109	train	0	377	58.25	215	74	4	45	0	4	109:rain
2	109	air	1	377	142.80	56	85	9	45	0	4	109:air
3	109	bus	0	377	27.52	301	63	8	45	0	4	109:bus
4	109	car	0	377	71.63	262	0	0	45	0	4	109:car
5	110	train	0	377	58.25	215	74	4	70	0	4	110:rain
6	110	air	1	377	142.80	56	85	9	70	0	4	110:air
7	110	bus	0	377	27.52	301	63	8	70	0	4	110:bus
8	110	car	0	377	71.63	262	0	0	70	0	4	110:car
9	111	train	0	377	58.25	215	74	4	35	0	4	111:rain
10	111	air	1	377	142.80	56	85	9	35	0	4	111:air

~~~ indexes ~~~~  
..

# Discrete choice model

## Long format

- Standard formulas are not very practical to describe random utility models, as these models may use different sets of covariates
- Working with random utility models, one has to consider at most three sets of covariates:
  1. alternative- and choice situation-specific covariates  $x_{ij}$  with generic coefficients  $\beta$  and alternative-specific covariates  $t_j$  with a generic coefficient  $\nu$
  2. choice situation-specific covariates  $z_i$  with alternative-specific coefficients  $\gamma_j$
  3. alternative- and choice situation-specific covariates  $w_{ij}$  with alternative-specific coefficients  $\delta_j$
- The covariates enter the observable part of the utility which can be written, for alternative  $j$ :

$$V_{ij} = \alpha_j + \beta x_{ij} + \nu t_j + \gamma_j z_i + \delta_j w_{ij}$$

- As the absolute value of utility is irrelevant, only utility differences are useful to modelize the choice for one alternative. For two alternatives  $j$  and  $l$ , we obtain:

$$V_{ij} - V_{il} = (\alpha_j - \alpha_l) + \beta(x_{ij} - x_{il}) + \nu(t_j - t_l) + (\gamma_j - \gamma_l)z_i + (\delta_j w_{ij} - \delta_l w_{il})$$

# Discrete choice model

## Model specification

- It is clear from the previous expression that coefficients of choice situation-specific variables (the intercept being one of those) should be alternative-specific; otherwise they would disappear in the differentiation
- Moreover, only differences of these coefficients are relevant and can be identified
- For example, with three alternatives 1, 2 and 3, the three coefficients  $\gamma_1, \gamma_2, \gamma_3$  associated with a choice situation-specific variable cannot be identified, but only two linear combinations
- Therefore, one has to make a choice of normalization, and the simplest one is just to set  $\gamma_1 = 0$
- Coefficients for alternative and choice situation-specific variables may (or may not) be alternative-specific
  - For example, transport time is alternative-specific, but 10 mn in public transport may not have the same impact on utility than 10 mn in a car. In this case, alternative-specific coefficients are relevant
  - Monetary cost is also alternative-specific, but in this case, one can consider that \$1 is \$1 however it is spent for the use of a car or in public transports. In this case, a generic coefficient is relevant.
- The treatment of alternative-specific variables doesn't differ much from the alternative and choice situation-specific variables with a generic coefficient
- However, if some of these variables are introduced, the  $\nu$  parameter can only be estimated in a model without intercepts to avoid perfect multicollinearity

# Discrete choice model

## Model specification

- The `mlogit` package provides objects of class `mFormula` which are built upon `Formula` objects provided by the `Formula` package.
- To illustrate the use of `mFormula` objects, we use again the `toronto_montreal` data set and consider three sets of covariates that will be indicated in a three-part formula, which refers to the three items mentioned above
  - Part 1: `cost` (monetary cost) is an alternative-specific covariate with a generic coefficient
  - Part 2: `income` and `urban` are choice situation-specific covariates
  - Part 3: `ivt` (in-vehicle travel time) is alternative-specific and alternative-specific coefficients are expected

```
1 library(Formula)
2 f <- Formula(choice ~ cost | income + urban | ivt)
```

# Discrete choice model

## Model specification

- A `model.frame` method is provided for `dfidx` objects
- It differs from the `formula` method by the fact that the returned object is an object of class `dfidx` and not an ordinary data frame, which means that the information about the structure of the data is not lost
- Defining a specific `model.frame` method for `dfidx` objects implies that the first argument of the function should be a `dfidx` object, which results in an unusual order of the arguments in the function (the data first, and then the formula)
- Moreover, as the model matrix for random utility models has specific features, we add a supplementary argument called `pkg` to the `dfidx` function so that the returned object has a specific class (and inherits the `dfidx` class)

```
1 library(mlogit)
2 MC <- dfidx(toronto_montreal, subset = noalt == 4, pkg = "mlogit")
3 class(MC)
```

```
[1] "dfidx_mlogit" "dfidx"          "data.frame"
```

```
1 f <- Formula(choice ~ cost | income | ivt)
2 mf <- model.frame(MC, f)
3 class(mf)
```

```
[1] "dfidx_mlogit" "dfidx"          "data.frame"
```

# Discrete choice model

## Model specification

- Using `mf` as the argument of `model.matrix` enables the construction of the relevant model matrix for random utility model, as a specific `model.matrix` method for `dfidx_mlogit` objects is provided
  - The model matrix contains  $J - 1$  columns for every choice situation-specific variables (`income` and the intercept), which means that the coefficient associated with the first alternative (`train`) is set to 0
  - It contains only one column for `cost` because we want a generic coefficient for this variable
  - It contains  $J$  columns for `ivt`, because it is an alternative specific variable for which we want alternative specific coefficients

```
1 head(model.matrix(mf), 4)
```

```
(Intercept):air (Intercept):bus (Intercept):car cost income:air income:bus
1              0              0              0 58.25              0              0
2              1              0              0 142.80             45              0
3              0              1              0  27.52              0             45
4              0              0              1  71.63              0              0
income:car ivt:train ivt:air ivt:bus ivt:car
1          0       215      0      0      0
2          0         0     56      0      0
3          0         0      0     301      0
4         45         0      0      0     262
```

# Discrete choice model

## Random utility model

- The utility for alternative  $l$  is written as:  $U_l = V_l + \epsilon_l$  where
  - $V_l$  is a function of some observable covariates and unknown parameters to be estimated
  - $\epsilon_l$  is a random deviate which contains all the unobserved determinants of the utility
- Alternative  $l$  is therefore chosen if  $\epsilon_j < (V_l - V_j) + \epsilon_l \forall j \neq l$
- The probability of choosing this alternative is then:

$$P(\epsilon_1 < V_l - V_1 + \epsilon_l, \epsilon_2 < V_l - V_2 + \epsilon_l, \dots, \epsilon_J < V_l - V_J + \epsilon_l)$$

- From the joint probability, we can derive the conditional (to the error) probability  $P_l | \epsilon_l$  and the unconditional probability  $P_l$

# Discrete choice model

## Distribution of the error terms

- The multinomial logit model is a special case of this setting
- It is based on **three hypotheses**:
  1. **independence of the errors**
  2. **each  $\epsilon$  follows a Gumbel distribution**
  3. **the errors are identically distributed (homoskedasticity hypothesis: the scale parameter of the Gumbel distribution is the same for all the alternatives)**
- The probabilities have then very simple, closed forms, which correspond to the logit transformation of the deterministic part of the utility

$$P_l = \frac{e^{V_l}}{\sum_{j=1}^J e^{V_j}}$$

# Discrete choice model

## Gumbel distribution

- The 2nd hypothesis states that each  $\epsilon$  follows a Gumbel (maximum) distribution, whose density and cumulative distribution functions are

$$f_Z(z) = \frac{1}{\theta} e^{-\frac{z-\mu}{\theta}} e^{-e^{-\frac{z-\mu}{\theta}}} \text{ and } F_Z(z) = \int_{-\infty}^z f(t) dt = e^{-e^{-\frac{z-\mu}{\theta}}},$$

where  $\mu$  is the location parameter and  $\theta$  the scale parameter

- The first two moments of the Gumbel distribution are  $E(Z) = \mu + \theta\gamma$ , where  $\gamma$  is the Euler-Mascheroni constant ( $\approx 0.57721$ ) and  $V(Z) = \frac{\pi^2}{6}\theta^2$
- The mean of  $\epsilon_j$  is not identified if  $V_j$  contains an intercept. We can then, without loss of generality suppose that  $\mu_j = 0, \forall j$
- Moreover, the overall scale of utility is not identified. Therefore, only  $J - 1$  scale parameters may be identified, and a natural choice of normalization is to impose that one of the  $\theta_j$  is equal to 1

# Discrete choice model

## Indipencece of Irrilevant alternative (IIA)

- If we consider the probabilities of choice for two alternatives  $l$  and  $m$ , we have  $P_l = e^{V_l} / \sum_j e^{V_j}$  and  $P_m = e^{V_m} / \sum_j e^{V_j}$
- The ratio of these two probabilities is

$$\frac{P_l}{P_m} = \frac{e^{V_l}}{e^{V_m}} = e^{V_l - V_m}.$$

- **This probability ratio for the two alternatives depends only on the characteristics of these two alternatives and not on those of other alternatives: This is called the independence of irrelevant alternatives (IIA) property**
- IIA relies on the hypothesis that the errors are identical and independent
- It is not a problem in itself and may even be considered as a useful feature for a well-specified model
- However, this hypothesis may be in practice violated, especially if some important variables are omitted

# Discrete choice model

## Marginal effects

- The marginal effects are the derivatives of the probabilities with respect to the covariates, which can be choice situation-specific ( $z_i$ ) or alternative-specific ( $x_{ij}$ ):

$$\frac{\partial P_{il}}{\partial z_i} = P_{il}(\beta_l - \sum_j P_{ij}\beta_j)$$

$$\frac{\partial P_{il}}{\partial x_{il}} = \gamma P_{il}(1 - P_{il})$$

$$\frac{\partial P_{il}}{\partial x_{ik}} = -\gamma P_{il}P_{ik}$$

- **For a choice situation-specific variable, the sign of the marginal effect is not necessarily the sign of the coefficient.**

Actually, the sign of the marginal effect is given by  $(\beta_l - \sum_j P_{ij} \beta_j)$ , which is positive if the coefficient for alternative  $l$  is greater than a weighted average of the coefficients for all the alternatives, the weights being the probabilities of choosing the alternatives. In this case, the sign of the marginal effect can be established with no ambiguity only for the alternatives with the lowest and the greatest coefficients

- **For an alternative-specific variable, the sign of the coefficient can be directly interpreted.** The marginal effect is obtained by multiplying the coefficient by the product of two probabilities which is at most 0.25. The rule of thumb is therefore to divide the coefficient by 4 in order to have an upper bound of the marginal effect

# Discrete choice model

## Marginal effects

- Note that the last equation can be rewritten:

$$\frac{\partial P_{il}}{\partial x_{ik}} = -\gamma P_{il} P_{ik} \implies \frac{dP_{nl}/P_{nl}}{dx_{nk}} = -\gamma P_{nk}$$

- Therefore, when a characteristic of alternative  $k$  changes, the relative changes of the probabilities for every alternative except  $k$  are the same, which is a consequence of the IIA property

# Discrete choice model

## Marginal rates of substitution

- Coefficients are marginal utilities, which cannot be interpreted
- However, ratios of coefficients are marginal rates of substitution
- For example, if the observable part of utility is:  $V = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ , joint variations of  $x_1$  and  $x_2$  which ensure the same level of utility are such that:  $dV = \beta_1 dx_1 + \beta_2 dx_2 = 0$  so that:

$$-\frac{dx_2}{dx_1} \Big|_{dV=0} = \frac{\beta_1}{\beta_2}.$$

- For example, if  $x_2$  is transport cost (in dollars),  $x_1$  transport time (in hours),  $\beta_1 = 1.5$  and  $\beta_2 = 0.05$ ,  $\frac{\beta_1}{\beta_2} = 30$  is the marginal rate of substitution of time in terms of dollars and the value of 30 means that, to reduce the travel time of 1 hour, the individual is willing to pay at most \$30 more. Stated more simply, time value is \$30 per hour

# Discrete choice model

## Consumer surplus

- Consumer's surplus has a very simple expression for multinomial logit models
- The level of utility attained by an individual is  $U_j = V_j + \epsilon_j$ ,  $j$  being the chosen alternative
- The expected utility is then:  $E(\max_j U_j)$ , where the expectation is taken over the values of all the error terms
- Its expression is simply, up to an additive unknown constant, the log of the denominator of the logit probabilities, often called the "log-sum"

$$E(U) = \ln \sum_{j=1}^J e^{V_j} + C.$$

- If the marginal utility of income ( $\alpha$ ) is known and constant, the expected surplus is simply  $\frac{E(U)}{\alpha}$

# Example

## mlogit

- Random utility models are fitted using the `mlogit` function
- Basically, only two arguments are mandatory, `formula` and `data`, if an `dfidx` object (and not an ordinary `data.frame`) is provided
- We use the `toronto_montreal` data set, which was already coerced to a `dfidx` object (called `MC`) previously
- The same model can then be estimated using as `data` argument this `dfidx` object or a `data.frame` (In this latter case, further arguments that will be passed to `dfidx` should be indicated)

```
1 MC <- dfidx(toronto_montreal, subset = noalt == 4)
2 ml.MC1 <- mlogit(choice ~ cost + freq + ovt | income | ivt, MC)
3 ml.MC1b <- mlogit(choice ~ cost + freq + ovt | income | ivt,
4                 toronto_montreal, subset = noalt == 4,
5                 idx = c("case", "alt"))
```

# Example

## mlogit

- `mlogit` provides two further useful arguments:
  - `reflevel` indicates which alternative is the “reference” alternative, i.e., the one for which the coefficients of choice situation-specific covariates are set to 0
  - `alt.subset` indicates a subset of alternatives on which the estimation has to be performed; in this case, only the lines that correspond to the selected alternatives are used, and all the choice situations where unselected alternatives have been chosen are removed
- We estimate the model on the subset of three alternatives (we exclude `bus` whose market share is negligible in our sample) and we set `car` as the reference alternative
- Moreover, we use a total transport time variable computed as the sum of the in-vehicle and out-vehicle time variables

```
1 MC$time <- MC$ivt + MC$ovt
2 ml.MC1 <- mlogit(choice ~ cost + freq | income | time, MC,
3               alt.subset = c("car", "train", "air"), reflevel = "car")
```

# Example

## mlogit

- The main results of the model are computed and displayed using the `summary` method:
  - The frequencies of the different alternatives in the sample are first indicated
  - Next, some information about the optimization is displayed: the Newton-Raphson method (with analytical gradient and hessian) is used, as it is the most efficient method for this simple model for which the log-likelihood function is globally concave. Note that very few iterations and computing times are required to estimate this model
  - Then the usual table of coefficients is displayed, followed by some goodness-of-fit measures: the value of the log-likelihood function, which is compared to the value when only intercepts are introduced, which leads to the computation of the McFadden  $R^2$  and to the likelihood ratio test

# Example

```
1 summary(ml.MC1)
```

Call:

```
mlogit(formula = choice ~ cost + freq | income | time, data = MC,  
        alt.subset = c("car", "train", "air"), reflevel = "car",  
        method = "nr")
```

Frequencies of alternatives:choice

```
      car      train      air  
0.45757 0.16721 0.37523
```

nr method

6 iterations, 0h:0m:0s

$g'(-H)^{-1}g = 6.94E-06$

successive function values within tolerance limits

Coefficients :

|  | Estimate | Std. Error | z-value | Pr(> z ) |
|--|----------|------------|---------|----------|
|--|----------|------------|---------|----------|

# Example

## mlogit

- The `fitted` method can be used either to obtain the probability of actual choices (`type = "outcome"`) or the probabilities for all the alternatives (`type = "probabilities"`)

```
1 head(fitted(ml.MC1, type = "outcome"))
```

```
      109      110      111      112      113      114
0.1909475 0.3399941 0.1470527 0.3399941 0.3399941 0.2440011
```

```
1 head(fitted(ml.MC1, type = "probabilities"), 4)
```

```
      car      train      air
109 0.4206404 0.3884120 0.1909475
110 0.3696476 0.2903582 0.3399941
111 0.4296769 0.4232704 0.1470527
112 0.3696476 0.2903582 0.3399941
```

# Example

## mlogit

- Note that the log-likelihood is the sum of the log of the fitted outcome probabilities and that, as the model contains intercepts, the average fitted probabilities for every alternative equals the market shares of the alternatives in the sample

```
1 sum(log(fitted(ml.MC1, type = "outcome")))
```

```
[1] -1951.344
```

```
1 logLik(ml.MC1)
```

```
'log Lik.' -1951.344 (df=9)
```

```
1 apply(fitted(ml.MC1, type = "probabilities"), 2, mean)
```

```
      car      train      air  
0.4575659 0.1672084 0.3752257
```

# Example

## mlogit

- Predictions can be made using the `predict` method
- If no data is provided, predictions are made for the sample mean values of the covariates
- Assume, for example, that we wish to predict the effect of a reduction of train transport time of 20%. We first create a new `data.frame` simply by multiplying train transport time by 0.8 and then using the `predict` method with this new `data.frame`

```
1 NMC <- MC
2 NMC$time[NMC$idx$salt == "train"] <- 0.8* NMC$time[NMC$idx$salt == "train"]
3 Oprob <- fitted(ml.MC1, type = "probabilities")
4 Nprob <- predict(ml.MC1, newdata = NMC)
5 rbind(old = apply(Oprob, 2, mean), new = apply(Nprob, 2, mean))
```

|     | car       | train     | air       |
|-----|-----------|-----------|-----------|
| old | 0.4575659 | 0.1672084 | 0.3752257 |
| new | 0.4039891 | 0.2643906 | 0.3316203 |

# Example

## mlogit

- (Illustration of the IIA property) If, for the first individuals in the sample, we compute the ratio of the probabilities of the air and the car mode, we obtain
- If train time changes, it changes the probabilities of choosing air and car, but not their ratio.

```
1 head(Nprob[, "air"] / Nprob[, "car"])
```

```
      109      110      111      112      113      114  
0.4539448 0.9197791 0.3422401 0.9197791 0.9197791 0.6021092
```

```
1 head(Oprob[, "air"] / Oprob[, "car"])
```

```
      109      110      111      112      113      114  
0.4539448 0.9197791 0.3422401 0.9197791 0.9197791 0.6021092
```

# Example

## mlogit

- We next compute the surplus for individuals of the sample induced by train time reduction
- This requires the computation of the log-sum term (also called inclusive value or inclusive utility) for every choice situation, which is:

$$iv_n = \ln \sum_{j=1}^J e^{\beta^\top x_{nj}}.$$

- For this purpose, we use the `logsum` function, which works on a vector of coefficients and a model matrix. The basic use of `logsum` consists of providing as unique argument (called `coef`) a `mlogit` object
- In this case, the `model.matrix` and the `coef` are extracted from the same model

```
1 ivbefore <- logsum(ml.MC1)
```

# Example

## mlogit

- To compute the log-sum after train time reduction, we must provide a model matrix which is not the one corresponding to the fitted model
- This can be done using the `X` argument which is a matrix or an object from which a model matrix can be extracted
- This can also be done by filling the `data` argument (a data frame or an object from which a data frame can be extracted using `model.frame`), and eventually the `formula` argument (a formula or an object for which the `formula` method can be applied)
- If no formula is provided, but if `data` is a `dfidx` object, the formula is extracted from it

```
1 ivafter <- logsum(ml.MC1, data = NMC)
```

# Example

## mlogit

- Surplus variation is then computed as the difference of the log-sums divided by the opposite of the cost coefficient which can be interpreted as the marginal utility of income
- Consumer surplus variations range from 0.6 to 31 Canadian dollars, with a median value of about \$4

```
1 surplus <- - (ivafter - ivbefore) / coef(ml.MC1) ["cost"]
2 summary(surplus)
```

| Min.   | 1st Qu. | Median | Mean   | 3rd Qu. | Max.    |
|--------|---------|--------|--------|---------|---------|
| 0.5853 | 2.8633  | 3.9395 | 4.7385 | 5.8623  | 31.6600 |

# Example

## mlogit

- Marginal effects are computed using the `effects` method
- By default, they are computed at the sample mean, but a `data` argument can be provided
- The variation of the probability and the covariate can be either absolute or relative
- This is indicated with the `type` argument which is a combination of two `a` (as absolute) and `r` (as relative) characters
- For example, `type = "ar"` means that what is measured is an absolute variation of the probability for a relative variation of the covariate
- The results indicate that, for a 100% increase of income, the probability of choosing `air` increases by 33 percentage points, as the probabilities of choosing `car` and `train` decrease by 18 and 15 percentage points

```
1 effects(ml.MC1, covariate = "income", type = "ar")
```

```
      car      train      air  
-0.1822177 -0.1509079  0.3331256
```

# Example

## mlogit

- For an alternative specific covariate, a matrix of marginal effects is displayed.
- The cell in the  $l^{\text{th}}$  row and the  $c^{\text{th}}$  column indicates the change of the probability of choosing alternative  $c$  when the cost of alternative  $l$  changes. As `type = "rr"`, elasticities are computed. For example, a 10% change of train cost increases the probabilities of choosing car and air by 3.36%. Note that the relative changes of the probabilities of choosing one of these two modes are equal, which is a consequence of the IIA property. Finally, in order to compute travel time valuation, we divide the coefficients of travel times (in minutes) by the coefficient of monetary cost (in dollars)
- The value of travel time ranges from 23 for a train to 37 Canadian dollars per hour for a plane

```
1 effects(ml.MC1, covariate = "cost", type = "rr")
```

|       | car        | train      | air        |
|-------|------------|------------|------------|
| car   | -0.9131273 | 0.9376923  | 0.9376923  |
| train | 0.3358005  | -1.2505014 | 0.3358005  |
| air   | 1.2316679  | 1.2316679  | -3.1409703 |

```
1 coef(ml.MC1)[grep("time", names(coef(ml.MC1)))] /  
2   coef(ml.MC1)["cost"] * 60
```

| time:car | time:train | time:air |
|----------|------------|----------|
| 29.52728 | 23.09447   | 36.95360 |

# Discrete choice model

## Logit models relaxing the iid hypothesis

- We assumed that the error terms were iid (identically and independently distributed), i.e., uncorrelated and homoskedastic. Extensions of the basic multinomial logit model have been proposed by relaxing one of these two hypotheses while maintaining the hypothesis of a Gumbel distribution
- Using `mlogit`, the heteroskedastic logit model is obtained by setting the `heterosc` argument to `TRUE`

```
1 ml.MC <- mlogit(choice ~ freq + cost + ivt + ovt |
2               urban + income, MC, refllevel = 'car',
3               alt.subset = c("car", "train", "air"))
4 hl.MC <- mlogit(choice ~ freq + cost + ivt + ovt |
5               urban + income, MC, refllevel = 'car',
6               alt.subset = c("car", "train", "air"),
7               heterosc = TRUE)
```

# Discrete choice model

## Logit models relaxing the iid hypothesis

- Two supplementary coefficients (`sp.train` and `sp.air`) are estimated ( $\theta_j$ ), the third for the reference modality being set to 1
- The variance of the error terms of train and air are respectively higher and lower than the variance of the error term of car (set to 1)

```
1 hl.MC
```

Call:

```
mlogit(formula = choice ~ freq + cost + ivt + ovt | urban + income, data = MC,
alt.subset = c("car", "train", "air"), reflevel = "car", heterosc = TRUE)
```

Coefficients:

|                   |                 |             |            |
|-------------------|-----------------|-------------|------------|
| (Intercept):train | (Intercept):air | freq        | cost       |
| 0.6783934         | 0.6567544       | 0.0639247   | -0.0269615 |
| ivt               | ovt             | urban:train | urban:air  |
| -0.0096808        | -0.0321655      | 0.7971316   | 0.4454726  |
| income:train      | income:air      | sp.train    | sp.air     |
| -0.0125979        | 0.0188600       | 1.2371829   | 0.5403239  |

# Discrete choice model

## Logit models relaxing the iid hypothesis

- Note that the z-values and p-values of the output are not particularly meaningful, as the hypothesis that the coefficient is zero (and not 1) is tested.
- The homoskedasticity hypothesis can be tested using any of the three tests. For the likelihood ratio and the Wald test, one can use only the fitted heteroskedastic model as argument. In this case, it is guessed that the hypothesis that the user wants to test is the homoskedasticity hypothesis.
- The homoskedasticity hypothesis is therefore strongly rejected using the Wald test, but only at 5% level for the likelihood ratio test

```
1 lrtest(hl.MC, ml.MC)
```

Likelihood ratio test

```
Model 1: choice ~ freq + cost + ivt + ovt | urban + income
```

```
Model 2: choice ~ freq + cost + ivt + ovt | urban + income
```

```
  #Df  LogLik  Df  Chisq Pr(>Chisq)
1   12 -1838.1
2   10 -1841.6 -2  6.8882  0.03193 *
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
1 waldtest(hl.MC, heterosc = FALSE)
```

Wald test

data: homoscedasticity

chisq = 25.196, df = 2, p-value = 3.38e-06

# Discrete choice model

## Nested logit model

- It is a generalization of the multinomial logit model that is based on the idea that some alternatives may be joined in several groups
- The error terms may then present some correlation in the same nest, whereas error terms of different nests are still uncorrelated. Denoting  $m = 1, \dots, M$ , the nests and  $B_m$  the set of alternatives belonging to nest  $m$ , the cumulative distribution of the errors is:

$$\exp\left(-\sum_{m=1}^M \left(\sum_{j \in B_m} e^{-\epsilon_j/\lambda_m}\right)^{\lambda_m}\right).$$

- The marginal distributions of the  $\epsilon$ s are still univariate extreme values, but there is now some correlation within nests.  $1 - \lambda_m$  is a measure of the correlation, i.e.,  $\lambda_m = 1$  implies no correlation

# Discrete choice model

## Nested logit model

- In the special case where  $\lambda_m = 1 \forall m$ , the errors are iid Gumbel errors and the nested logit model reduce to the multinomial logit model
- It can then be shown that the probability of choosing alternative  $j$  that belongs to nest  $l$  is:

$$P_j = \frac{e^{V_j/\lambda_l} \left( \sum_{k \in B_l} e^{V_k/\lambda_l} \right)^{\lambda_l - 1}}{\sum_{m=1}^M \left( \sum_{k \in B_m} e^{V_k/\lambda_m} \right)^{\lambda_m}},$$

and that this model is a random utility model if all the  $\lambda$  parameters are in the 0 – 1 interval

- Let us now write the deterministic part of the utility of alternative  $j$  as the sum of two terms: the first ( $Z_j$ ) being specific to the alternative and the second ( $W_l$ ) to the nest it belongs to:

$$V_j = Z_j + W_l.$$

# Discrete choice model

## Nested logit model

- We can then rewrite the probabilities as follow:

$$\begin{aligned} P_j &= \frac{e^{(Z_j+W_l)/\lambda_l}}{\sum_{k \in B_l} e^{(Z_k+W_l)/\lambda_l}} \times \frac{\left( \sum_{k \in B_l} e^{(Z_k+W_l)/\lambda_l} \right)^{\lambda_l}}{\sum_{m=1}^M \left( \sum_{k \in B_m} e^{(Z_k+W_m)/\lambda_m} \right)^{\lambda_m}} \\ &= \frac{e^{Z_j/\lambda_l}}{\sum_{k \in B_l} e^{Z_k/\lambda_l}} \times \frac{\left( e^{W_l/\lambda_l} \sum_{k \in B_l} e^{Z_k/\lambda_l} \right)^{\lambda_l}}{\sum_{m=1}^M \left( e^{W_m/\lambda_m} \sum_{k \in B_m} e^{Z_k/\lambda_m} \right)^{\lambda_m}}. \end{aligned}$$

- Then denote  $I_l = \ln \sum_{k \in B_l} e^{Z_k/\lambda_l}$  which is often called the log-sum, the inclusive value or the inclusive utility

# Discrete choice model

## Nested logit model

- We then can write the probability of choosing alternative  $j$  as:

$$P_j = \frac{e^{Z_j/\lambda_l}}{\sum_{k \in B_l} e^{Z_k/\lambda_l}} \times \frac{e^{W_l + \lambda_l I_l}}{\sum_{m=1}^M e^{W_m + \lambda_m I_m}}.$$

- The first term  $P_{j|l}$  is the conditional probability of choosing alternative  $j$  if nest  $l$  is chosen. It is often referred to as the *lower model*
- The second term  $P_l$  is the marginal probability of choosing nest  $l$  and is referred to as the *upper model*.  $W_l + \lambda_l I_l$  can be interpreted as the expected utility of choosing the best alternative in  $l$ ,  $W_l$  being the expected utility of choosing an alternative in this nest (whatever this alternative is) and  $\lambda_l I_l$  being the expected extra utility gained by being able to choose the best alternative in the nest
- The inclusive values link the two models. It is then straightforward to show that IIA applies within nests, but not for two alternatives in different nests

# Discrete choice model

## Nested logit model

- A consistent but inefficient way of estimating the nested logit model is to estimate separately its two components
  - The coefficients of the lower model are first estimated, which enables the computation of the inclusive values  $I_j$
  - The coefficients of the upper model are then estimated, using  $I_j$  as covariates
- Maximizing directly the likelihood function of the nested model leads to a more efficient estimator.

# Nested logit model

## Example

- To illustrate the estimation of the nested logit model, we use the [telephone](#) data set
- A total of 428 households were surveyed in 1984 about their choice of a local telephone service, which typically involves the choice between a flat service (a fixed monthly charge for an unlimited calls within a specified geographical area) and a measured (a reduced fixed monthly charge for a limited number of calls plus usage charges for additional calls) service. Households had the choice between five services:
  - *budget measured* ([budget](#)): no fixed monthly charge; usage charges apply to each call made
  - *standard measured* ([standard](#)): a fixed monthly charge covers up to a specified dollar amount (greater than the fixed charge) of local calling, after which usage charges apply to each call made
  - *local flat* ([local](#)): a greater monthly charge that may depend upon residential location; unlimited free calling within a local calling area; usage charges apply to calls made outside local calling area
  - *extended area flat* ([extended](#)): a further increase in the fixed monthly charge to permit unlimited free calling within an extended area
  - *metro area flat* ([metro](#)): the greatest fixed monthly charge that permits unlimited free calling within the entire metropolitan area
- The first two services are measured, and the last three are flat services. There is therefore an obvious nesting structure for this example

# Nested logit model

## Example

- We first estimate the multinomial logit model, with the log of cost as the unique covariate. We then update this model in order to introduce nests, using the `nests` argument. It is a list of characters that contains the alternatives for the different nests. It is advisable to use a named list (we use here `"measured"` and `"flat"` as names of the nests)

```
1 telephone <- as.data.frame(micsr.data::telephone)
2 head(telephone)
```

|   | choice | service  | household | cost    |
|---|--------|----------|-----------|---------|
| 1 | FALSE  | budget   | 1         | 1.7613  |
| 2 | FALSE  | extended | 1         | 13.8160 |
| 3 | FALSE  | local    | 1         | 2.5455  |
| 4 | FALSE  | metro    | 1         | 3.1476  |
| 5 | TRUE   | standard | 1         | 1.7544  |
| 6 | FALSE  | budget   | 2         | 1.2585  |

```
1 ml_tel <- mlogit(choice ~ log(cost), telephone,
2                 idx = c("household", "service"))
3 nl_tel <- mlogit(choice ~ cost, telephone,
4                 idx = c("household", "service"),
5                 nests = list(measured = c("budget", "standard"),
6                               flat = c("local", "metro", "extended")))
```

```
7 coef(nl_tel)
```

```
(Intercept):extended      (Intercept):local      (Intercept):metro
          1.2255254          1.2716312          1.7836981
(Intercept):standard      cost                    iv:measured
          0.3782248        -1.4899641          0.4847705
          iv:flat
          0.4362081
```

# Nested logit model

## Example

- Two supplementary coefficients are estimated, `iv:measured` and `iv:flat`
- The two values are in the 0-1 interval and close to each other
- The `un.nest.el` argument enables to estimate a unique supplementary coefficient for the two nests

```
1 coef(nl_tel)
```

```
(Intercept):extended      (Intercept):local      (Intercept):metro
                1.2255254                1.2716312                1.7836981
(Intercept):standard      cost                    iv:measured
                0.3782248                -1.4899641                0.4847705
                iv:flat
                0.4362081
```

```
1 nl_tel_u <- update(nl_tel, un.nest.el = TRUE)
```

# Nested logit model

## Example

- We then have three nested models and the hypothesis of a unique parameter can be tested using any of the three tests
- The three tests conclude that a unique parameter can be estimated.

```
1 scoretest(nl_tel_u, un.nest.el = FALSE)
```

```
score test
```

```
data: un.nest.el = FALSE  
chisq = 0.10891, df = 1, p-value = 0.7414  
alternative hypothesis: unique nest elasticity
```

```
1 lrtest(nl_tel, nl_tel_u)
```

```
Likelihood ratio test
```

```
Model 1: choice ~ cost
```

```
Model 2: choice ~ cost
```

|   | #Df | LogLik  | Df | Chisq  | Pr(>Chisq) |
|---|-----|---------|----|--------|------------|
| 1 | 7   | -473.22 |    |        |            |
| 2 | 6   | -473.29 | -1 | 0.1367 | 0.7116     |

```
1 waldtest(nl_tel, un.nest.el = TRUE)
```

Wald test

```
data: unique nest elasticity  
chisq = 0.16024, df = 1, p-value = 0.6889
```

# Nested logit model

## Example

- Then, we can test whether this parameter is 1, in which case the nested logit model reduces to the multinomial logit model
- Based on the Wald and the likelihood ratio, the preferred specification is the nested logit model (but note that the p-value for the score test is slightly higher than 5%)

```
1 scoretest(ml_tel,  
2           nests = list(measured = c("budget", "standard"),  
3                         flat = c("local", "metro", "extended")))
```

score test

```
data: measured, flat  
chisq = 5.6853, df = 2, p-value = 0.05827  
alternative hypothesis: nested model
```

```
1 lrtest(nl_tel_u, ml_tel)
```

Likelihood ratio test

```
Model 1: choice ~ cost  
Model 2: choice ~ log(cost)
```

```
#Df  LogLik Df  Chisq Pr(>Chisq)
1    6 -473.29
2    5 -481.88 -1 17.177  3.406e-05 ***
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
1 waldtest(nl_tel_u, nests = NULL)
```

Wald test

data: no nests

chisq = 25.541, df = 1, p-value = 4.33e-07

# Discrete choice model

Multinomial probit model

- Random utility:  $U_j = V_j + \varepsilon_j$ ,  $\varepsilon \sim \mathcal{N}(0, \Omega)$

- Choice rule: Alternative  $l$  is chosen if

$$U_l > U_j \forall j \neq l \Leftrightarrow \varepsilon_j - \varepsilon_l < -(V_j - V_l)$$

- Choice probability:

$$P_l = \Pr(\varepsilon^l < -V^l), \quad \text{cov}(\varepsilon^l) = \Omega^l = M^l \Omega M^{l\top}$$

- $(J - 1)$ -dimensional multivariate normal integral
- Identification:
  - Utility scale not identified  $\rightarrow$  one variance normalized
  - Only functions of  $\Omega$  are identified
  - With  $J$  alternatives:  $J(J - 1)/2 - 1$  identified parameters
- Estimation:
  - No closed-form probabilities
  - Simulation using Cholesky decomposition
- Key points:
  - Flexible substitution patterns
  - No IIA (unlike multinomial logit)