EPFL
ENAC TRANSP-OR
Prof. M. Bierlaire

Mathematical Modeling of Behavior
Fall 2020

EXERCISE SESSION 11 (Solution)

Exercise 1 The data about internal trips for a city of 300'000 individuals is described in Table 1. Three different transportation modes are considered to travel inside the city: car, bus and slow modes (walk and bike). The population is segmented in three different groups according to their age: young (age ≤ 21), adult ($21 <$ age ≤ 65) and retired (age > 65).

	Age			
Mode	Young	Adult	Retired	
Car	10'000	120'000	15'000	145'000
Bus	15'000	30'000	25'000	70'000
Slow mode	25'000	50'000	10'000	85'000
	50'000	200'000	50'000	300'000

Table 1: Description of the population of the city

Table 2 contains the data for a sample of 350 individuals of the full population.

	Age			
Mode	Young	Adult	Retired	
Car	20	120	15	155
Bus	30	30	25	85
Slow mode	50	50	10	110
	100	200	50	350

Table 2: Description of the sample

1. Calculate the sampling probabilities. Which sampling strategy has been used to collect the data?

   Solution: The sampling probabilities can be calculated by dividing the population for each segment (young, adult, retired) in Table 2 by the population for the corresponding segment in Table 1. We can check that for each stratum the market share has been multiplied by the corresponding sampling probability.

- $R_{\text{young}} = \frac{100}{50000} = \frac{1}{500}$
- $R_{\text{adult}} = \frac{200}{200000} = \frac{1}{1000}$
- $R_{\text{retired}} = \frac{50}{50000} = \frac{1}{1000}$

Since the probability of being drawn only depends on variables other than the choice (in this case the age), the sampling strategy is Exogeneous Stratified Sample (XSS).

2. Which estimation procedure can be used?

   Solution: ML (Maximum Likelihood) and ESML (Exogenous Sampling Maximum Likelihood). In general ML is impossible to handle, but it can be simplified if the sampling is exogeneous by using the ESML.

3. Express (mathematically) how to calculate the market shares in the population for each alternative using the sample (Table 2) and the choice probabilities $P(i|x_n)$.
   Hint: Consider the weight of each segment W_s ($W_s = \frac{\text{\# persons in segment } s \text{ in the population}}{\text{\# persons in segment } s \text{ in the sample}}$) and the indicator I_{ns}, which is 1 if individual n belongs to segment s and 0 otherwise.

   Solution: First we need to calculate the weight of each segment:

$$W_{\text{young}} = \frac{50000}{100} = 500$$
$$W_{\text{adult}} = \frac{200000}{200} = 1000$$
$$W_{\text{retired}} = \frac{50000}{50} = 1000$$

Then, the number of people choosing transportation mode i is estimated by:

$$\hat{N}(i) = \sum_{\text{n}\in\text{sample}} \sum_{\text{segment } s} W_s I_{ns} P(i|x_n)$$

The market share for each transportation mode is then obtained by dividing by the size of the full population (N):

$$\hat{W}(i) = \frac{\hat{N}(i)}{N}.$$

Exercise 2 Consider the population described in Table 1.

1. Characterize a sample with the following sampling probabilities:

   Solution:

   - $R(\text{Car}) = \frac{1}{1000}$
   - $R(\text{Bus}) = \frac{1}{500}$
   - $R(\text{Slow Modes}) = \frac{1}{1000}$

	Age			
Mode	Young	Adult	Retired	
Car	10	120	15	145
Bus	30	60	50	140
Slow mode	25	50	10	85
	65	230	75	370

Table 3: Description of the sample with the given sampling probabilities

2. Which sample strategy have you used?

   Solution: Pure-choice based sampling as each alternative in the choice set corresponds to a separate stratum (in this case car, bus and slow mode).

3. Which estimation procedures are used for the following models with this sampling strategy?

   Solution:

   - **Logit model:** ESML and correct the constants
   - **MEV model:** requires to estimate the bias from data, it is implemented on Biogeme

Exercise 3 For this exercise you will have to use material from the textbook[1] that you have not seen in the lectures. Make sure that you have read and understood section 10.5.3 about the incremental logit before doing the following exercise.

1. In a Stated Preferences (SP) mode choice experiment, respondents are exposed to the following three alternatives: (1) biking, (2) walking and (3) taking the metro. You have estimated a model with the following utility functions:

[1] Ben-Akiva, M., Bierlaire, M., McFadden, D. & Walker, J. Discrete Choice Analysis.

$$
\begin{aligned}
U_{1n} =& \beta_1 \cdot t_{1n} + \varepsilon_{1n} \\
U_{2n} =& ASC_2 + \beta_2 \cdot t_{2n} + \varepsilon_{2n} \\
U_{3n} =& ASC_3 + \beta_3 \cdot t_{3n} + \varepsilon_{3n}
\end{aligned}
$$

where t_{in} is the travel time of respondent n using mode i, $\varepsilon_{1n}, \varepsilon_{2n}, \varepsilon_{3n} \overset{iid}{\sim}$ EV$(0,1)$ and β_i, ASC$_i$ are the parameters that have been estimated, with $i \in \{1,2,3\}$. The population of interest consists of exactly *four* individuals: $n = 1$ with travel time t_{11}, $n = 2$ and $n = 3$ with equal travel times $t_{12} = t_{13}$ and $n = 4$ with travel time t_{14}. Fig. 1 illustrates the probability of choosing alternative 1 (bicycle), as predicted by this model, for each value of travel time.

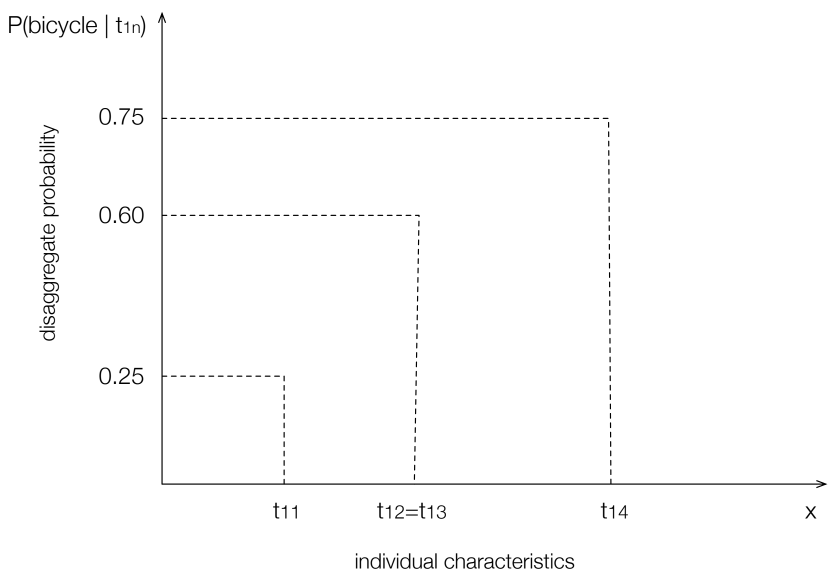

Figure 1: $P(\text{bicycle}|t_{1n})$

Compute the fraction of the population that is expected to choose alternative 1.

Solution:

The probabilities of choosing alternative 1 for each of the four individuals are the following:

(a) $n = 1$: $P_1(1) = 0.25$

(b) $n = 2$: $P_2(1) = 0.60$

(c) $n = 3$: $P_3(1) = 0.60$

(d) $n = 4$: $P_4(1) = 0.75$

The fraction of the population choosing alternative 1 is the following:

$$W(1) = \frac{1}{N_T} \sum_{n=1}^{N_T} P(i|x_n) = E(P(i|x_n) = \frac{\sum_{n=1}^{4} P_n(1)}{4} = \frac{(0.25 + 2 \cdot 0.6 + 0.75)}{4} = 0.55$$

Thus, 55% of the population is choosing Alternative 1.

2. Consider now individual $n = 1$ with travel time by bicycle t_{11}. Your *linear-in-parameters* logit model from the previous sub-question has predicted the probability with which this individual will choose each alternative as shown in column $P_1(i|t_{11})$ of Table 4. Due to an increase in the number of bike lanes in the city, there is a change in travel time from t_{11} to t'_{11}. The change in utility for the three alternatives associated with this individual change as indicated in column ΔV_i.

Mode	$P_1(i\|t_{11})$	ΔV_i	$P_1(i\|t'_{11})$
Bicycle	0.25	0.5	
Walk	0.45	0	
Metro	0.30	0	

Table 4: Choice probabilities before and after an increase in number of bike lanes

Complete Table 4 by computing the new choice probabilities $P_1(i|t'_{11})$ for the three alternatives.

Solution: The new choice probabilities are calculated as follows:

$$P_1(i|t'_{11}) = \frac{P_1(i|t_{11}) \cdot e^{\Delta V_i}}{\sum_{j \in C} P_1(j|t_{11}) \cdot e^{\Delta V_j}}$$

$$\text{Bike: } 0.25 \cdot e^{0.5}/(0.25 \cdot e^{0.5} + 0.45 + 0.30) = 0.35$$
$$\text{Walk: } 0.45/(0.25 \cdot e^{0.5} + 0.45 + 0.30) = 0.39$$
$$\text{Metro: } 0.30/(0.25 \cdot e^{0.5} + 0.45 + 0.30) = 0.26$$

The table is updated as follows:

Mode	$P_1(i\|t_{11})$	ΔV_i	$P_1(i\|t'_{11})$
Bicycle	0.25	0.5	0.35
Walk	0.45	0	0.39
Metro	0.30	0	0.26

---

jp/th/rk/no/mpp/mw