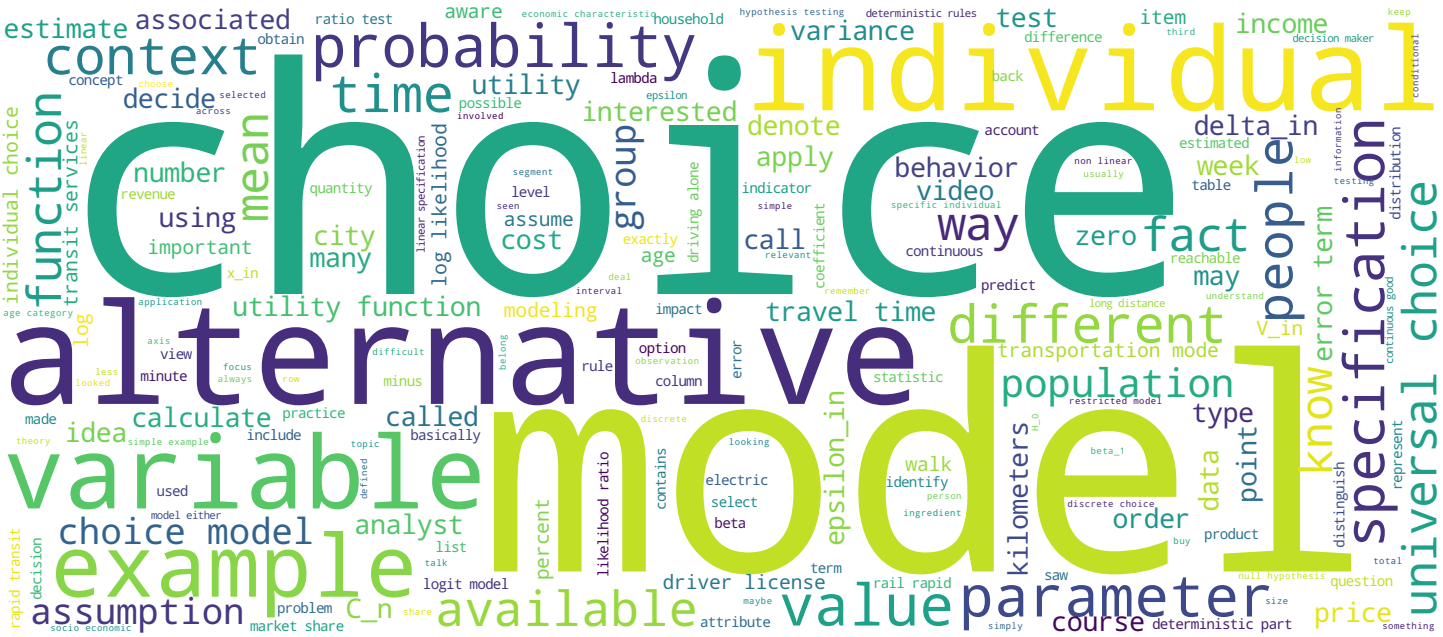
Choice with multiple alternatives

Derivation of the logit model

Michel Bierlaire

Introduction to choice models

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE



EPFL

Search MOOC

Video

Hello. This week we will look at the choice model involving more than two alternatives. And we will focus on the logit model. We saw the logit model in the context of the binary choice. We will now see how we generalize it to more than two alternatives.

Notes

Summary

0m 04s

The error term

For all $i \in \mathcal{C}_n$

$$U_{in} = V_{in} + \varepsilon_{in}$$

- What is $\mathcal{C}_n$?
- What is ε_{in}?
- What is V_{in}?

The starting point is again the utility specification. And, in the context of this week, we will look at the three principal ingredients involved in this utility model. The first ingredient is the choice set that we denote by C_n. The second ingredient is the epsilon_in, the error term, and the third is the V_in, the deterministic part of the utility function. Well, the idea is that, in this week, we spend most of the time on the V_in: how to specify the deterministic part. This is where the modeling is the most involved. But let's first look at the choice set, the C_n. This will be the topic of this video.

Notes

Summary

0m 23s

Choice set

Individual's choice set

- No driver license
- No auto available
- Awareness of transit services
- Transit services unreachable
- Walking not an option for long distance

Mode choice

- ~~driving alone~~
- sharing a ride
- taxi
- motorcycle
- bicycle
- ~~walking~~
- ~~transit bus~~
- rail rapid transit

The choice set is the list of alternatives. And what we need to do is to distinguish between two types of choice set. The first type of choice set, I call it the universal choice set. This universal choice set is actually the point of view of the analyst. It contains all the potential alternatives that should be considered by the entire population. Well, it's actually restricted to the alternatives that are relevant for the analyst. Let's take an example where I'm interested to model the choice of transportation mode for commuters in the city. As an analyst, I'm interested in all possible transportation modes that are available in that city. People can drive alone to work. People can share a ride. They can take a taxi, they can use the motorcycle, or bicycle, they can walk, they can use transit, or the rail rapid transit. This is, in this example, my universal choice set. But I need also to take into account the point of view of the individual, of the traveler. And because not everybody has access to all these transportation modes which are available in the city. So the second choice set, the second concept of choice set, is called individual choice set, and it takes into account the fact that, for example, a traveler may have no driver's license.

Notes

Summary

1m 12s

Choice set

Individual's choice set

- No driver license
- No auto available
- Awareness of transit services
- Transit services unreachable
- Walking not an option for long distance

Mode choice

- ~~driving alone~~
- sharing a ride
- taxi
- motorcycle
- bicycle
- ~~walking~~
- ~~transit bus~~
- rail rapid transit

Therefore, driving alone is not an option. There may be no car available in the household. Again driving alone will not be an option either. Maybe people are not aware of transit services. They just moved in the city, they don't know that there are some buses available in their neighborhood. Or these transit services are not reachable, they are too far away. Then transit bus will be out. And the last example is that, if the distance to walk is too long, people will not walk. They will not consider it as an option. So for this specific individual who has no driver license, who is not aware of transit services and does not want to walk long distances, the choice set is reduced to sharing a ride, taxi, motorcycle, bicycle, and rail rapid transit. So we need to distinguish between the universal choice set and the individual choice set.

Notes

Summary

2m 39s

Choice set

Choice set generation is tricky

- How to model "awareness"?
- What does "long distance" exactly mean?
- What does "unreachable" exactly mean?

We assume here deterministic rules

- Car is available if *n* has a driver license and a car is available in the household
- Walking is available if trip length is shorter than 4km.

Actually identifying the choice set for each individual is a very difficult process. How do you model awareness? How do you know that people are aware of the existence of a bus service or not? How do you decide what does long distance mean? Would people walk two kilometers, five kilometers, ten, how do you know? How do you define this for each individual? And when you say that a bus stop is not reachable, what does it mean exactly? So these questions are very difficult to deal with. In the context of our models, because we want to derive operational models, we will assume that you can actually identify the choice set for each individual using deterministic rules. For example, a first rule is that the car is available if individual n has a driver's license and a car that is available in the household. And this information should be available as a data in the data file. Another rule could be: walking is available if the trip length is shorter than four kilometers. OK. Again maybe this "four kilometers" is an arbitrary decision, but that's what we decide here. These rules are deterministic.

Notes

Summary

3m 36s

Availability conditions

$$\delta_{in} = \begin{cases} 1 & \text{if } i \in \mathcal{C}_n, \\ 0 & \text{otherwise.} \end{cases}$$

Choice model

$$P_n(i|\mathcal{C}_n) = P_n(i|\delta_n, \mathcal{C}) = \Pr(U_{in} + \ln \delta_{in} \geq U_{jn} + \ln \delta_{jn})$$

The way to represent the availability conditions is by using an indicator. So for each individual n and each alternative i in the choice set, we will denote by delta_in the indicator that tells us if the alternative belongs to the choice set. So, formally, delta_in would be 1 if alternative i in in the choice set of individual n and 0 otherwise. So then we can write the choice model using delta_in as a variable actually. So instead of writing the choice model P of i given C_n, so the choice of choosing alternative i given the choice set C_n of individual n, we can actually write the same model but now conditional to the universal choice set C and the set of variables delta. And one way to do it is to say that, using utility theory, the probability to choose alternative i is equal to the probability that U_in plus the log of delta_in is the largest among all alternatives in the choice set. If delta_in is equal to 1, log of delta_in will be 0. And we are back to the original utility. If delta_in happens to be 0, the log of 0 is actually minus infinity. Therefore, it cannot be the largest possible. So this way of coding the model, of presenting it, allows us to write the model either as a function of the universal choice set, or as a function of the individual choice set. And depending on the context, we will use one or the over representation.

Notes

Summary

4m 51s

The error term

For all $i \in C_n$

$$U_{in} = V_{in} + \varepsilon_{in}$$

- What is C_n?
- What is ε_{in}?
- What is V_{in}?

So know that we have looked at the choice set, that we have made these assumptions about universal choice set, individual-specific choice set, the fact that we use deterministic rules, and the fact that we can write the model either using one or the other choice set, now we will move to the next video where we will talk about the error term, the epsilon_in. And the assumptions that we will have to do for the epsilon_in.

Notes

Summary

6m 35s