Mathematical modeling of behavior — Discussions in class — Session 1

Michel Bierlaire

Scale of logit

Why is the scale of the logit model not identified?

1. Because it is normalized to 1. No. It is the consequence, not the cause.
2. Because it does not matter. No. The scale does matter.
3. **Because only the order of utility matters, not the value.** Indeed. The choice probability is given by
$$P_n(i|\mathcal{C}_n) = \Pr(U_{in} \geq U_{jn} \forall j \in \mathcal{C}_n),$$
and is not affected by a change of scale:
$$P_n(i|\mathcal{C}_n) = \Pr(\mu U_{in} \geq \mu U_{jn} \forall j \in \mathcal{C}_n),$$
for any $\mu > 0$. Note that it is true not only for logit, but for any random utility model.
4. I don't know.

Generic or alternative specific?

We consider a mode choice model with two variables: travel cost and travel time. For each of these variables, should the coefficient be generic or alternative specific, based on behavioral considerations? More specifically, which one of the following models is the most behaviorally meaningful?

1. Model 1: both parameters are generic:
$$\begin{aligned} U_{in} &= \beta_c \mathrm{cost}_{in} + \beta_t \mathrm{time}_{in}, \\ U_{jn} &= \beta_c \mathrm{cost}_{jn} + \beta_t \mathrm{time}_{jn}, \end{aligned}$$
that involves two unknown parameters.

2. Model 2: the cost parameter is generic and the time parameter is alternative specific:

$$U_{in} = \beta_c \text{cost}_{in} + \beta_{ti} \text{time}_{in},$$
$$U_{jn} = \beta_c \text{cost}_{jn} + \beta_{tj} \text{time}_{jn},$$

that involves three unknown parameters.

3. Model 3: the time parameter is generic and the cost parameter is alternative specific:

$$U_{in} = \beta_{ci} \text{cost}_{in} + \beta_t \text{time}_{in},$$
$$U_{jn} = \beta_{cj} \text{cost}_{jn} + \beta_t \text{time}_{jn},$$

that involves three unknown parameters.

4. Model 4: both parameters are alternative specific:

$$U_{in} = \beta_{ci} \text{cost}_{in} + \beta_{ti} \text{time}_{in},$$
$$U_{jn} = \beta_{cj} \text{cost}_{jn} + \beta_{tj} \text{time}_{jn},$$

that involves four unknown parameters.

5. **All models are equally behaviorally meaningful.**

6. None of them is behaviorally meaningful.

All are behaviorally meaningful. If time (cost, resp.) is considered as a resource, one minute is one minute, irrespectively how you spend it. Therefore, the time (cost, resp.) coefficient should be generic. But a behavioral argument based on a different perception of time in one alternative (the car, say), or the other (the bus, say) may also support an alternative specific specification. In that case, we ackowledge the fact that one minute spent in the car may not be perceived in the same way as one minute spent in the bus.

Therefore, each potential specification is equally meaningful from a behavioral point of view. Hypothesis testing with real data will help selecting the most appropriate specification.

Segmentation

Consider a model involving only one variable (travel time, say). And there is a time coefficient for males and one for females. We have a sample of 200 males and 200 females. The estimates are $\beta_m = -0.123$ and $\beta_f = -0.096$. We collect more data from another 100 females and re-estimate the same model with the sample of 500 individuals. Will the parameters have the exact same value or not? Which one of the following cases are you expecting to happen?

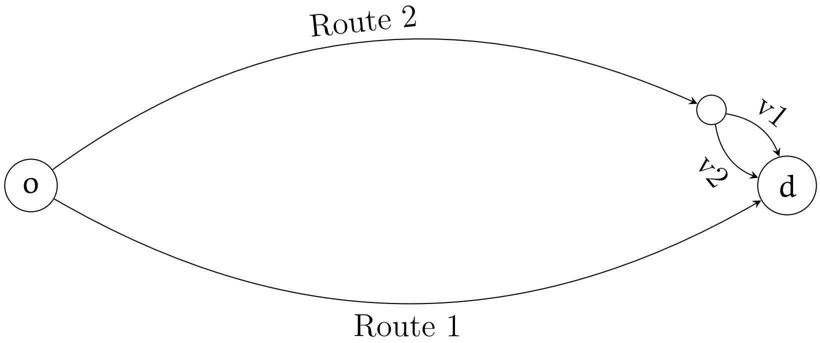
1. β_m same value (-0.123), β_f same value (-0.096),
2. β_m same value (-0.123), β_f different value, **(this is the correct answer)**
3. β_m different value, β_f same value (-0.096),
4. β_m different value, β_f different value.

We can consider that we have two populations: males and females. And the two parameters are estimated independently in the two populations. During the second wave of estimation, the data available to estimate the coefficient for males are exactly the same as during the first wave. Therefore, the estimator will be exactly the same. On the contrary, we have more data for females during the second wave than during the first. It is therefore extremely unlikely that the value of the estimate will be exactly the same. Indeed, we expect the precision of the estimator to increase with the number of data.

Overlapping paths

We consider two routes linking an origin and a destination. The two routes have exactly the same travel time T. The second route includes two variants for a small portion of the itinerary. We consider a logit model with three alternatives (route 1, route 2 variant 1, and route 2 variant 2) where travel time is the only explanatory variable. What is the probability predicted by the model for a given individual to choose route 1?

1. $\approx 1/2$,
2. $\approx \mathbf{1/3}$,
3. $\approx 1/4$,
4. ≈ 0.



Intuitively, we expect each of the two routes to have roughly 50% probability to be chosen. If we apply the logit model, we obtain

- utility of route 1: βT,
- utility of route 2, variant 1: βT,
- utility of route 2, variant 2: βT.

Therefore,

$$
\begin{aligned}
P_n(1) &= \frac{e^{\beta T}}{e^{\beta T} + e^{\beta T} + e^{\beta T}} \\
&= \frac{1}{3}.
\end{aligned}
$$

The reason why we do not obtain the intuitive value of 1/2 is due to the assumption of independence of the error terms associated with the derivation of the logit model. Indeed, the two variants of route 2 share all the unobserved variables associated with route 2. Therefore, the error terms are certainly not independent, in this example.

Captivity

Consider a binary model choice model between car (i) and train (j) for commute: $P_n(i|\{i,j\})$. Individuals without a driving license are said to be captive, as they have no choice. They have to take the train. The analyst does not have information about the possession of a driving license. But she knows the age of the respondents, and she knows that, if an individual is under 24, the probability to have a driving license is 45%. Using the model, how can the analyst calculate the probability for such an individual to use the car?

We have to decompose the model into each possible scenarios:

$$\begin{aligned}
\Pr(\text{car}) = &\Pr(\text{car}|\text{license}) \Pr(\text{license}) \\
&+ \Pr(\text{car}|\text{no license}) \Pr(\text{no license}) \\
= &\Pr(\text{car}|\text{license}) 0.45 + 0 \cdot (1 - 0.45) \\
= &0.45 \Pr(\text{car}|\text{license}).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\Pr(\text{bus}) = &\Pr(\text{bus}|\text{license}) \Pr(\text{license}) \\
&+ \Pr(\text{bus}|\text{no license}) \Pr(\text{no license}) \\
= &\Pr(\text{bus}|\text{license}) 0.45 + 1 \cdot (1 - 0.45) \\
= &0.45 \Pr(\text{bus}|\text{license}) + 0.55.
\end{aligned}$$

Such a model is sometimes called a latent class model, as the class of the individuals is not observed.