GEV

Gabriel Petrini

September 10th, 2020

Contents


```
(require 'org-attach-screenshot)
```

1 Pre-class reading: Train (2009, Ch 4.1–4.2)

1.1 Reading Quiz

1. Explain why mixed logit is such a popular model, according to Train. What are its strong points vis-à-vis the IIA problem?

2. What are the computational drawbacks to mixed logit?

3. What are two names for a mixed logit model that has a discrete mixing distribution?

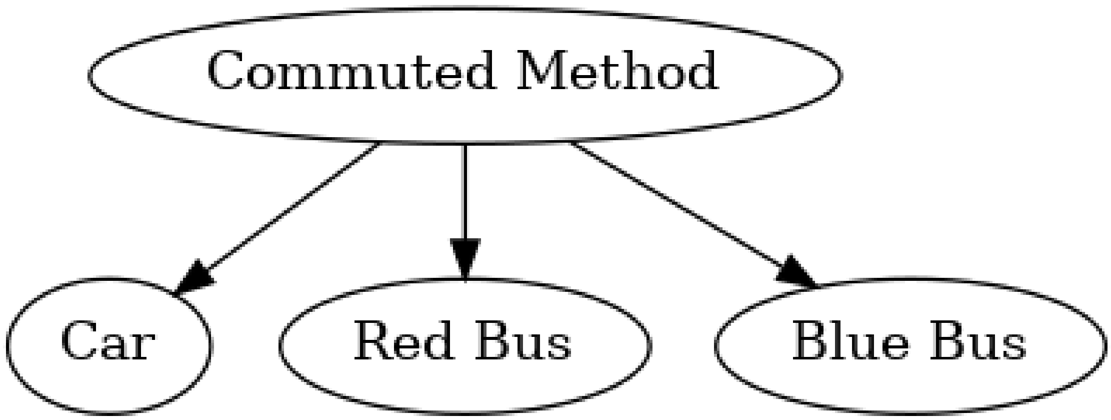
4. Explain why the EM algorithm is a potentially powerful tool for estimating discrete choice models.

5. How does the EM algorithm work? Why does it work?

2 Review

2.1 Red bus / blue bus choice set

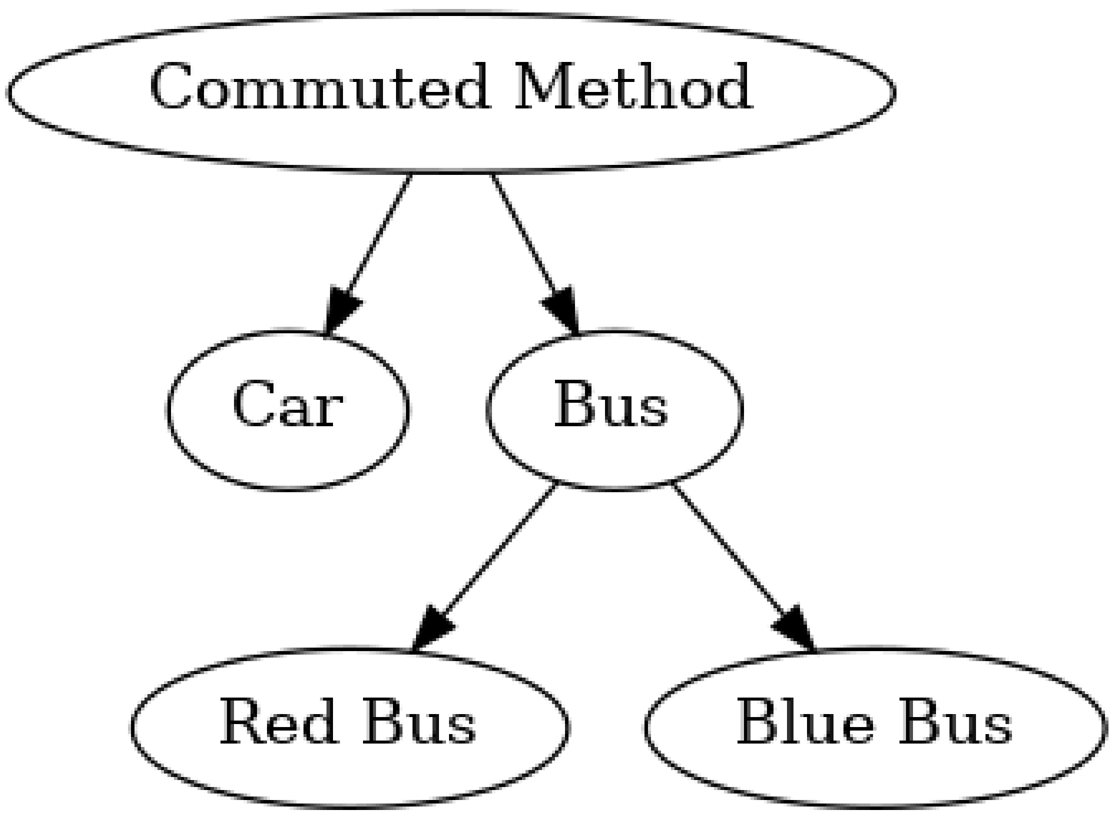
- As we discussed last time, adding another bus results in odd substitution patterns
- This is because of the IIA property of multinomial logit models



2.2 Nesting the choice set

One way to get around IIA is to **nest** the choice set

- Nesting explicitly introduces correlation across alternatives within the same nest



2.3 Other cases where nesting is useful

Elections

- Suppose we have two candidates, A and B
- If we introduce C, whose platform resembles B, what will new vote shares be?
- (Primary elections are a kind of nesting)

Product markets

- Nesting "branded" and "generic" products (e.g. branded vs. micro-brewed beer)
- But some consumers won't purchase either type of the product
- Ignoring non-buyers, will give misleading price elasticities of demand

3 Nested Logit

Coming back to the red bus/blue bus problem, we would like some way for the errors for the red bus to be correlated with the errors for the blue bus

- The **nested logit** allows a nest-specific error:

$$
\begin{aligned}
U_{i,RedBus} &= u_{i,RedBus} &+ \nu_{i,Bus} + \lambda\epsilon_{i,RedBus} \\
U_{i,BlueBus} &= u_{i,BlueBus} &+ \nu_{i,Bus} + \lambda\epsilon_{i,BlueBus} \\
U_{i,Car} &= u_{i,Car} &+ \nu_{i,Car} + \lambda\epsilon_{i,Car}
\end{aligned}
$$

where:

1. $\nu_{ik} + \lambda\epsilon_{ij}$ is distributed Type I Extreme Value
2. The ν's and the ϵ's are independent
3. ϵ_{ij} is distributed Type I extreme value
4. Distribution of ν_k's is derived in Theorem 2.1 of **cardell1997**

Composite error term for car is independent from either the red bus error or the blue bus error

- If we added a yellow bus, all errors in the bus nest would be independent conditional on choosing to take a bus (i.e. **IIA within nest**)
- But the bus nest errors are correlated from the viewpoint of the top level (i.e. before conditioning on nest choice)
- Note: adding two extreme value errors does .hi[not] give back an extreme value error
  - But the difference between two T1EV errors is distributed logistic

More important than the exact error distribution is the choice probabilities:

$$P_{iC} = \frac{\exp(u_{iC})}{\left[\exp\left(\frac{u_{iRB}}{\lambda}\right) + \exp\left(\frac{u_{iBB}}{\lambda}\right)\right]^{\lambda} + \exp(u_{iC})}$$

$$P_{iRB} = \frac{\exp\left(\frac{u_{iRB}}{\lambda}\right)\left[\exp\left(\frac{u_{iRB}}{\lambda}\right) + \exp\left(\frac{u_{iBB}}{\lambda}\right)\right]^{\lambda-1}}{\left[\exp\left(\frac{u_{iRB}}{\lambda}\right) + \exp\left(\frac{u_{iBB}}{\lambda}\right)\right]^{\lambda} + \exp(u_{iC})}$$

Not particularly intuitive, but can break it down into parts $P(B)P(RB|B)$:

$$P_{iRB} = \left(\frac{\left[\exp\left(\frac{u_{iRB}}{\lambda}\right) + \exp\left(\frac{u_{iBB}}{\lambda}\right)\right]^{\lambda}}{\left[\exp\left(\frac{u_{iRB}}{\lambda}\right) + \exp\left(\frac{u_{iBB}}{\lambda}\right)\right]^{\lambda} + \exp(u_{iC})}\right) \times \left(\frac{\exp\left(\frac{u_{iRB}}{\lambda}\right)}{\exp\left(\frac{u_{iRB}}{\lambda}\right) + \exp\left(\frac{u_{iBB}}{\lambda}\right)}\right) \tag{1}$$

4 Nested Logit Estimation

The log likelihood can then be written as:

$$\begin{aligned}
\ell &= \sum_{i=1}^{N} \sum_{j \in J} (d_{ij} = 1) \ln(P_{ij}) \\
&= \sum_{i=1}^{N} \left[(d_{iC} = 1) \ln(P_{iC}) + \sum_{j \in J_B} (d_{ij} = 1) \ln(P_{iB} P_{ij|B}) \right] \\
&= \sum_{i=1}^{N} \left[(d_{iC} = 1) \ln(P_{iC}) + \sum_{j \in J_B} (d_{ij} = 1)(\ln(P_{iB}) + \ln(P_{ij|B})) \right] \\
&= \sum_{i=1}^{N} \Bigg[(d_{iC} = 1) \ln(P_{iC}) + (d_{iBB} = 1 + d_{iRB} = 1) \ln(P_{iB}) \\
&\qquad + \sum_{j \in J_B} (d_{ij} = 1) \ln(P_{ij|B}) \Bigg]
\end{aligned}$$

Could estimate a nested logit by straight maximum likelihood. An alternative follows from decomposing the nests into the product of two probabilities: $P(RB|B)P(B)$

- In order to do this, however, first decompose u_{RB} into two parts:

$$u_{iRB} = u_{iB} + u_{iRB|B}$$

- We also need to choose normalizations:
  - $u_{iC} = 0$
  - $u_{iBB|B} = 0$
- So we will estimate $(\beta_B, \beta_{RB}, \gamma, \lambda)$ where γ corresponds to the Z's (alt-specific)

Note that our normalizations imply the following observable components of utility

$$\begin{aligned}
u_{iC} &= 0 \\
u_{iBB} &= \beta_B X_i + \gamma(Z_{BB} - Z_C) \\
u_{iRB} &= (\beta_B + \beta_{RB})X_i + \gamma(Z_{RB} - Z_C)
\end{aligned}$$

- Now estimate β_{RB} and γ in a 1st stage using only observations that chose bus, N_B:

$$\ell_1 = \sum_{i=1}^{N_B}(d_{iRB} = 1)(u_{iRB|B}/\lambda) + \ln\Big(1 + \exp(u_{iRB|B}/\lambda)\Big)$$

- The 1 in the ln() operator corresponds to $\exp(u_{iBB|B}/\lambda)$ since $u_{iBB|B} = 0$

Now consider the term in the numerator of $P(B)$ in (1). We can rewrite this as:

$$\begin{aligned}
\left[\exp\left(\frac{u_{iRB}}{\lambda}\right) + \exp\left(\frac{u_{iBB}}{\lambda}\right)\right]^{\lambda} &= \exp(u_{iBB})\left[\exp\left(\frac{u_{iRB|B}}{\lambda}\right) + 1\right]^{\lambda} \\
&= \exp(u_{iBB} + \lambda I_{iB})
\end{aligned}$$

where I_{iB} is called the .hi[inclusive value] and is given by:

$$I_{iB} = \ln\left(\exp\left(\frac{u_{iRB|B}}{\lambda}\right) + 1\right)$$

Note: looks like E (utility) associated with a particular nest (minus Euler's constant)

Taking the estimates of $u_{iRB|B}$ as given and calculating the inclusive value, we now estimate a second logit to get β_B:

$$\ell_2 = \sum_i (d_{iB} = 1)(u_{iBB} + \lambda I_{iB} - u_{iC}) + \ln(1 + \exp(u_{iBB} + \lambda I_{iB} - u_{iC}))$$

- Could do all this because log of the probabilities was additively separable. Consider the log likelihood

contribution of someone who chose red bus:

$$\ln(P_{iB}(\beta_B, \beta_{RB}, \gamma, \lambda)) + \ln(P_{iRB|B}(\beta_{RB}, \gamma))$$

- We get estimates of β_{RB} and γ only from the second part of log likelihood
- Then we take these as given when estimating β_B and λ

5 The Nested Logit as a Dynamic Discrete Choice Model

Instead of having individuals know their full error, consider the case where the error is revealed in stages

- First individuals choose whether or not to ride the bus and there is an extreme value error associated with both the bus and the car option
- Individuals take into account that if they choose the bus option they will get to make a choice about which bus in the next period (option value)
- With the errors in the second choice also distributed Type I extreme value, independent from each other, and independent from the errors in the first period, the expectation on the value of the second period decision is λI_{iB} plus Euler's constant.

5.1 Proposition 1 (mcfadden1978)

Let $Y_j = e^{u_j}$. Suppose we have a function $G(Y_1, ..., Y_J)$ that maps from R^J into R^1

If G satisfies:

1. $G \geq 0$
2. G is homogeneous of some degree k
3. $G \to \infty$ as $Y_j \to \infty$ for any j
4. Cross partial derivatives weakly alternate in sign, beginning with $G_i \geq 0$

then:

$$F(u_1, ..., u_{\mathcal{J}}) = \exp\left[-G(Y_1, ..., Y_J)\right]$$

is the cumulative distribution of a multivariate extreme value function and:

$$P_i = \frac{Y_i G_i}{G}$$

where G_i denotes the derivative of G with respect to Y_i

6 Logit from GEV

Another way of thinking about the last statement is that:

$$P_i = \frac{\partial \ln(G)}{\partial u_i}$$

- For the multinomial logit case, the G function is:

$$G = \sum_{j=1}^{J} \exp(u_j)$$

with the derivative of the log of this giving multinomial logit probabilities

- But $\ln(G)$ (plus Euler's constant) is .hi[also] expected utility
- In fact, for all GEV models $\ln(G)$ is expected utility!

Suppose a nested logit model with two nests (F, NF) and a no-purchase option N

- The G function is then:

$$G = \left(\sum_{j \in F} \exp(u_j/\lambda_F)\right)^{\lambda_F} + \left(\sum_{j \in NF} \exp(u_j/\lambda_{NF})\right)^{\lambda_{NF}} + \exp(u_N)$$

- Differentiating $\ln(G)$ (the expected utility function) with respect to u_j where $k \in F$ yields the probability k is chosen:

$$P_k = \frac{\exp(u_k/\lambda_F)\left(\sum_{j \in F} \exp(u_j/\lambda_F)\right)^{\lambda_F - 1}}{\left(\sum_{j \in F} \exp(u_j/\lambda_F)\right)^{\lambda_F} + \left(\sum_{j \in NF} \exp(u_j/\lambda_{NF})\right)^{\lambda_{NF}} + \exp(u_N)}$$

6.1 Overlapping nests (Bresnahan et al., 1997)

We can also come up with more general nesting structures

- **bst1997** model 4 overlapping nests for computers:
- Branded but not Frontier $\{B, NF\}$
- Generic but Frontier $\{NB, F\}$
- Branded and Frontier $\{B, F\}$
- Generic but not Frontier $\{NB, NF\}$
- Use the model to understand market power in PC sector in late 1980s
- Overlapping nests explain coexistence of imitative entry and innovative investment

References