Discrete Choice Model - Eurobarometer Survey

Derivation of a Multichoice Logit model

Thomas Monk

The agent is presented with a normalised set of J options. She is instructed to pick the j_1 most important options from that set. The following assumes $j_1 = 2$, however this could be extended arbitrarilaly.

This derivation adapts Ophem, Stam et al. (1999), and Ben-Akiva, Lerman et al. (1985, pp 104-107).

1 Additive random utility model

Consider a random utility model, following McFadden (1974), where the utility generated by alternative j, $j \in \{1, 2, ..., M\}$. Individual subscripts are dropped for simplicity.

$$U_j = V_j + \epsilon_j \tag{1}$$

where V_j is the deterministic component, depending on individual specific and alternative specific factors, and ϵ_j is stochastic component. Assume ϵ_j is independently distributed across both alternatives and observations, and is has a Gumbel extreme value distribution:

$$F(\epsilon_j) = H_j(U_j) = \exp(-\exp(-\epsilon_j)) \tag{2}$$

substituting in the model,

$$F(\epsilon_j) = \exp(-\exp(-(U_j - V_j))) \tag{3}$$

Define $\mathcal{J} = 1, 2, ..., J$, and $k = 1, ..., j_1$. Further, define $\hat{U}$ as the maximum utility from l alternatives, where $l = M \setminus k$.

We are therefore interested in estimating the probability that the utilities generated by options 1 and 2 exceed all the utilities of the options in l:

$$P(\{U_1, U_2\} > \max_{j=k+1,...,J} \{U_j\}) \tag{4}$$

Following Ophem, Stam et al. (1999, p. 119), equation 4 can be written as:

$$1 - P(\max_{j=3,...,J} U_j > \max\{U_1, U_2\}) - P(\max_{j=3,...,J} U_j > U_1) + P(\max_{j=3,...,J} U_j > U_2) \tag{5}$$

1.1 Single choice made

In the case of None or Don't Know, only one choice is made for all observations. In this case, we need to calculate the simpler density:

$$P(U_1 > \max_{j=2,...,15} U_j) \tag{6}$$

1.2 Properties of the Gumbel distribution (Ben-Akiva, Lerman et al., 1985, p 104)

If the random variable ϵ is distributed under the Gumbel distribution, then:

$$F(\epsilon) = \exp(-\exp(-\mu(\epsilon - \eta))) \tag{7}$$

where η is a location parameter, and μ is a positive scale parameter.
The distribution has the following properties:

1. The mode is η.
2. The mean is $\eta + \gamma/\mu$, where γ is Euler's constant.
3. The variance is $\pi^2/6\mu^2$.
4. If ϵ is Gumbel distributed with parameters (η, μ), and V and α are scalar constants, then $\alpha\epsilon + V$ is Gumbel distributed with parameters $(\alpha\eta + V, \mu/\alpha)$.
5. If ϵ_1 and ϵ_2 are independent Gumbel distributed variables with parameters (η_1, μ) and (η_2, μ) respectively, then $\epsilon^\star = \epsilon_1 - \epsilon_2$ is logistically distributed:

$$F(\epsilon^\star) = \frac{1}{1 + \exp(\mu(\eta_2 - \eta_1 - \epsilon^\star))} \tag{8}$$

6. If $(\epsilon_1, \epsilon_2, ..., \epsilon_J)$ are J independent Gumbel distributed random variables with parameters (η_j, μ) respectively, then $\max(\epsilon_1, \epsilon_2, ..., \epsilon_J)$ is Gumbel distributed with parameters:

$$\left(\frac{1}{\mu} \ln \sum_{j=1}^{J} \exp\left(\mu \cdot \eta_j\right), \mu\right) \tag{9}$$

The mean of ϵ_j is not identified if V_j contains an intercept. We can then, without loss of generality impose that $\eta = 0, \forall j$.

More generally than the above, the overall scale of utility is not identified. Therefore, only $J - 1$ scale parameters may be identified, and a natural choice of normalisation is to impose that one of the μ_j is equal to 1. McFadden (1974) further imposes the hypothesis that $\mu_j = 1, \ \forall j$.

Therefore, equations 8 simplifies to:

$$F(\epsilon^\star) = \frac{1}{1 + \exp(-\epsilon^\star)} \tag{8a}$$

1.3 Deriving the multichoice logit model

Our equation of interest is the following:

$$1 - P(\max_{j=3,\dots,J} U_j > U_1) - P(\max_{j=3,\dots,J} U_j > U_2) + P(\max_{j=3,\dots,J} U_j > \max\{U_1, U_2\}) \tag{5}$$

Which is equivalent to:

$$1 - P(U_1 - \max_{j=3,\dots,J} U_j \leq 0) - P(U_2 - \max_{j=3,\dots,J} U_j \leq 0) + P(\max\{U_1, U_2\} - \max_{j=3,\dots,J} U_j \leq 0) \tag{5}$$

Given property 4, and our assumptions, the affine transformation of the random variable ϵ, as in the case of the random utility model, $U_j = V_j + e_j$, is distributed with parameters $(\eta, \mu) = (V_j, 1)$.

Therefore, given equation 9, $U^* \equiv \max_{j=3,\dots,J} U_j$ is also Gumbel distributed, with the following parameters (η, μ):

$$\max_{j=3,\dots,J} U_j \equiv U^* \sim G\Big(\ln \sum_{j=3}^{J} \exp(V_j), 1\Big) \tag{10}$$

Similarly, as a special case, $U^*_{1,2} \equiv \max(U_1, U_2)$ is also Gumbel distributed with the following parameters (η, μ):

$$\max\{U_1, U_2\} \sim G\Big(\ln[\exp(V_1) + \exp(V_2)], 1\Big) \tag{11}$$

Therefore, $U^*_1 \equiv U^* - U_1$ is logistically distributed, given property 5:

$$F(U^*_1) = \frac{1}{1 + \exp(V_1 - \ln \sum_{j=3}^{J} \exp(V_j) - U^*_1)} \tag{12}$$

Therefore, using standard properties of exponentials:

$$\begin{aligned} P(U^*_1 \leq 0) = F(0) &= \frac{1}{1 + \exp(V_1 - \ln \sum_{j=3}^{J} e^{V_j})} \\ &= \frac{1}{1 + e^{V_1} / e^{\ln \sum_{j=3}^{J} e^{V_j}}} \\ &= \frac{1}{1 + e^{V_1} / \sum_{j=3}^{J} e^{V_j}} \\ &= \frac{\sum_{j=3}^{J} e^{V_j}}{e^{V_1} + \sum_{j=3}^{J} e^{V_j}} \end{aligned} \tag{13}$$

Similarly, defining $U^*_2 \equiv U^* - U_2$:

$$F(U^*_2) = \frac{\sum_{j=3}^{J} e^{V_j}}{e^{V_2} + \sum_{j=3}^{J} e^{V_j}} \tag{14}$$

Finally, $U^{**}_{1,2} = U^* - U^*_{1,2}$ is also logistically distributed:

$$\begin{aligned} F(U^{**}_{1,2}) &= \frac{1}{1 + e^{\ln[e^{V_1} + e^{V_2}] - \ln \sum_{j=3}^{J} e^{V_j}} - U^*_2} \\ F(0) &= \frac{\sum_{j=3}^{J} e^{V_j}}{e^{V1} + e^{V_2} + \sum_{j=3}^{J} e^{V_j}} \end{aligned} \tag{15}$$

This means that our object of interest becomes:

$$P(\{U_1, U_2\} \geq \max_{j=3,\dots,J} U_j) = P(U^*_1 \leq 0) + P(U^*_2 \leq 0) - P(U^{**}_{1,2} \leq 0) \tag{16}$$

which can be represented in closed form as:

$$= \frac{\sum_{j=3}^{J} e^{V_j}}{e^{V_1} + \sum_{j=3}^{J} e^{V_j}} + \frac{\sum_{j=3}^{J} e^{V_j}}{e^{V_2} + \sum_{j=3}^{J} e^{V_j}} - \frac{\sum_{j=3}^{J} e^{V_j}}{e^{V1} + e^{V_2} + \sum_{j=3}^{J} e^{V_j}} \tag{17}$$

This is equivalent[1] to Ophem, Stam et al. (1999, pg.120) derivation.

$$P_{q,k,s} = 1 - \frac{\sum_{j=3}^{J} e^{V_j}}{e^{V_1} + \sum_{j=3}^{J} e^{V_j}} - \frac{\sum_{j=3}^{J} e^{V_j}}{e^{V_2} + \sum_{j=3}^{J} e^{V_j}} + \frac{\sum_{j=3}^{J} e^{V_j}}{e^{V1} + e^{V_2} + \sum_{j=3}^{J} e^{V_j}} \tag{18}$$

[1] 1 minus, I need to understand why

1.3.1 Single choice made

In this case, our equation of interest also becomes logistically distributed:

$$\begin{aligned} P(\max_{j=2,\ldots,15} U_j \leq U_1) &= \frac{1}{1+\exp(V_1 - \ln\sum_{j=2}^{J} e^{V_j})} \\ &= \frac{\sum_{j=2}^{J} e^{V_j}}{e^{V_1} + \sum_{j=2}^{J} e^{V_j}} \end{aligned} \tag{6}$$

1.4 The likelihood function

Assume the deterministic part of utility V_j is a linear function of the parameters to be estimated $V_j = X'\beta_j$, where X is a k vector of attributes, and β_j is a k vector of parameters.

To write down the likelihood function, we need to construct dummy variables, $d_{i,j}$ of $S = \binom{M-2}{2}$ combinations of responses: the set is $M-2$ as Don't Know and None are single choice options. Indicators s, t map to the specific combination in S.

The likelihood function is therefore:

$$
\begin{aligned}
L(y_i|x_i;\beta) &= \prod^{N} f(y_i|X;\beta) \\
&= \prod^{N}\Bigg[\prod_{s,t}^{S}\left(1-\frac{\sum_{j\neq s,t} e^{V_j}}{e^{V_s}+\sum_{j\neq s,t} e^{V_j}}-\frac{\sum_{j\neq s,t} e^{V_j}}{e^{V_t}+\sum_{j\neq s,t} e^{V_j}}+\frac{\sum_{j\neq s,t} e^{V_j}}{e^{V_s}+e^{V_t}+\sum_{j\neq s,t} e^{V_j}}\right)^{d_{s,t}} \\
&\quad\cdot\prod_{s=M-1}^{M}\left(\frac{\sum_{j\neq s} e^{V_j}}{e^{V_s}+\sum_{j\neq s} e^{V_j}}\right)^{d_s}\Bigg] \\
&= \prod^{N}\Bigg[\prod_{s,t}^{S}\left(1-\frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_s'\beta_j}+\sum_{j\neq s,t} e^{X_j'\beta_j}}-\frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_t'\beta_j}+\sum_{j\neq s,t} e^{X_j'\beta_j}}+\frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_s'\beta_j}+e^{X_t'\beta_j}+\sum_{j\neq s,t} e^{X_j'\beta_j}}\right)^{d_{s,t}} \\
&\quad\cdot\prod_{s=M-1}^{M}\left(\frac{\sum_{j\neq s} e^{X_j'\beta_j}}{e^{X_s'\beta_j}+\sum_{j\neq s} e^{X_j'\beta_j}}\right)^{d_s}\Bigg]
\end{aligned}
\tag{19}
$$

1.4.1 Using dual choices only

In this case, we use only the subset of the data in which each agent chooses two options. Call the number of observations in this subset N_d, and the number of optons D. Therefore, $S_d = \binom{D}{2}$.

$$
L(y_i|x_i;\beta) = \prod^{N_d}\Bigg[\prod_{s,t}^{S_d}\left(1-\frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_s'\beta_j}+\sum_{j\neq s,t} e^{X_j'\beta_j}}-\frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_t'\beta_j}+\sum_{j\neq s,t} e^{X_j'\beta_j}}+\frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_s'\beta_j}+e^{X_t'\beta_j}+\sum_{j\neq s,t} e^{X_j'\beta_j}}\right)^{d_{s,t}}\Bigg]
\tag{20}
$$

1.5 Maximum Likelihood Estimation

Using the dual choice model (equation ??):

$$\begin{aligned}
\hat{\beta} &= \arg\max_{\beta} L(y_i|x_i;\beta) \\
&= \arg\max_{\beta} \ln L(y_i|x_i;\beta) \\
&= \ln \prod^{N_d} \left[\prod_{s,t}^{S_d} \left(1 - \frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_s'\beta_j} + \sum_{j\neq s,t} e^{X_j'\beta_j}} - \frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_t'\beta_j} + \sum_{j\neq s,t} e^{X_j'\beta_j}} + \frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_s'\beta_j} + e^{X_t'\beta_j} + \sum_{j\neq s,t} e^{X_j'\beta_j}} \right)^{d_{s,t}} \right] \\
&= \sum^{N_d} \sum_{s,t}^{S_d} d_{s,t} \cdot \ln \left[1 - \frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_s'\beta_j} + \sum_{j\neq s,t} e^{X_j'\beta_j}} - \frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_t'\beta_j} + \sum_{j\neq s,t} e^{X_j'\beta_j}} + \frac{\sum_{j\neq s,t} e^{X_j'\beta_j}}{e^{X_s'\beta_j} + e^{X_t'\beta_j} + \sum_{j\neq s,t} e^{X_j'\beta_j}} \right]
\end{aligned} \tag{21}$$

1.6 Jacobian and Hessian

$$\sum^{N_d}\sum^{S_d}_{s,t} d_{s,t}\cdot\ln\left[1-\frac{\sum_{j\neq s,t}e^{X_j'\beta_j}}{e^{X_s'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}-\frac{\sum_{j\neq s,t}e^{X_j'\beta_j}}{e^{X_t'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}+\frac{\sum_{j\neq s,t}e^{X_j'\beta_j}}{e^{X_s'\beta_j}+e^{X_t'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}\right]$$
$$=\sum^{N_d}\sum^{S_d}_{s,t} d_{s,t}\cdot\ln\left[1-\sum_{j\neq s,t}e^{X_j'\beta_j}\cdot\left[\frac{1}{e^{X_s'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}-\frac{1}{e^{X_t'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}+\frac{1}{e^{X_s'\beta_j}+e^{X_t'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}\right]\right] \tag{22}$$

The interior summation can just be seen as an indicator function, $I(s,t\in\text{choice})$ - it only turns on the expression when the agent n picks the selection s,t - i.e. it is only true once for each agent. This means it can essentially be ignored in practice.

1.6.1 Jacobian

The derivative with respect to a specific *scalar* β_s and X_s is as follows. We only need the derivative from the respect of a β_s as only in the case that the β_i is β_s or β_t does this contribute any value to the Jacobian.

I need to check this result with a vector B_s

For example, consider the derivative of $\beta_i, i\neq s$. $d_{s,t}$ in this case is 0, so the whole contribution to the summation can be ignored.

$$\frac{\partial\mathcal{L}}{\partial\beta_s'}=\sum^{N_d}\sum^{S_d}_{s,t} d_{s,t}\cdot -\frac{X_s'\cdot e^{X_s'\beta_j}\cdot\sum_{j\neq s,t}e^{X_j'\beta_j}\cdot\left(\frac{1}{(e^{X_s'\beta_j}+e^{X_t'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j})^2}-\frac{1}{(e^{X_s'\beta_j}+e^{X_t'\beta_j})^2}\right)}{1-\sum_{j\neq s,t}e^{X_j'\beta_j}\cdot\left(\frac{1}{e^{X_s'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}-\frac{1}{e^{X_t'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}+\frac{1}{e^{X_s'\beta_j}+e^{X_t'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}\right)}$$
$$=\sum^{N_d} I(\cdot)-\frac{X_s'\cdot e^{X_s'\beta_j}\cdot\sum_{j\neq s,t}e^{X_j'\beta_j}\cdot\left(\frac{1}{(e^{X_s'\beta_j}+e^{X_t'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j})^2}-\frac{1}{(e^{X_s'\beta_j}+e^{X_t'\beta_j})^2}\right)}{1-\sum_{j\neq s,t}e^{X_j'\beta_j}\cdot\left(\frac{1}{e^{X_s'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}-\frac{1}{e^{X_t'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}+\frac{1}{e^{X_s'\beta_j}+e^{X_t'\beta_j}+\sum_{j\neq s,t}e^{X_j'\beta_j}}\right)} \tag{23}$$

References

Ben-Akiva, M. E., Lerman, S. R. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, volume 9. MIT Press.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

McFadden, D. (1974). *The Measurement of Urban Travel Demand*. *Journal of Public Economics*, 3(4), 303–328.

Ophem, H. V., Stam, P. and Praag, B. V. (1999). *Multichoice Logit: Modeling Incomplete Preference Rankings of Classical Concerts*. *Journal of Business & Economic Statistics*, 17(1), 117–128.