Choice with multiple alternatives – 5.1
Derivation of the logit model

Michel Bierlaire

To derive the logit model, we consider the following ingredients:

- a choice set for each individual n: $\mathcal{C}_n = \{1, \ldots, J_n\}$, and
- a utility function for each individual and each alternative: $U_{in} = V_{in} + \varepsilon_{in}$.

We assume that the error terms ε_{in} are

- independent, both across alternatives and individuals, and
- Extreme Value distributed, with the same parameters for each individual and each alternative:

$$\varepsilon_{in} \sim \mathrm{EV}(0, \mu). \tag{1}$$

These assumptions are summarized by the statement "i.i.d. Extreme Value", where "i.i.d." stands for *independent and identically distributed.*

The choice model is

$$P(i|\mathcal{C}_n) = \Pr(V_{in} + \varepsilon_{in} \geq \max_{j=1,\ldots,J_n} V_{jn} + \varepsilon_{jn}). \tag{2}$$

The idea of the derivation is to consider this model as a binary logit model, as we have already derived its specification. In order to be chosen, alternative i must have a utility larger than all other alternatives. Now, consider within the set $\mathcal{C}_n \setminus \{i\}$ composed of the other alternatives, the one associated with the highest utility. We do not know which specific alternative achieves this, but we know that its utility is

$$U_n^* = \max_{j \in \mathcal{C}_n \setminus \{i\}} U_{in} = \max_{j \in \mathcal{C}_n \setminus \{i\}} (V_{jn} + \varepsilon_{jn}). \tag{3}$$

Therefore, the choice model can be written:

$$P(i|\mathcal{C}_n) = \Pr(U_{in} \geq U_n^*), \tag{4}$$

that involves only two alternatives. In order to derive the choice model, we need to know the distribution of U_n^*. From a property of the extreme value distribution (see property 6 in the appendix below), and the fact that all error terms are i.i.d., we have that

$$U_n^* \sim \mathrm{EV}\left(\frac{1}{\mu}\ln \sum_{j\in\mathcal{C}_n\setminus\{i\}} e^{\mu V_{jn}}, \mu\right). \tag{5}$$

Equivalently (see property 4 in the appendix), we can write

$$U_n^* = V_n^* + \varepsilon_n^* \tag{6}$$

where

$$V_n^* = \frac{1}{\mu}\ln\sum_{j=2}^{J_n} e^{\mu V_{jn}} \tag{7}$$

and

$$\varepsilon_n^* \sim \mathrm{EV}(0, \mu). \tag{8}$$

Consequently, (4) is a binary logit model and

$$P(i|\mathcal{C}_n) = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_n^*}} \tag{9}$$

where

$$V_n^* = \frac{1}{\mu}\ln \sum_{j\in\mathcal{C}_n\setminus\{i\}} e^{\mu V_{jn}}. \tag{10}$$

We have

$$e^{\mu V_n^*} = e^{\ln \sum_{j\in\mathcal{C}_n\setminus\{i\}} e^{\mu V_{jn}}} = \sum_{j\in\mathcal{C}_n\setminus\{i\}} e^{\mu V_{jn}}, \tag{11}$$

and (9) can be written

$$P(i|\mathcal{C}_n) = \frac{e^{\mu V_{in}}}{e^{\mu V_{1n}} + \sum_{j\in\mathcal{C}_n\setminus\{i\}} e^{\mu V_{jn}}}, \tag{12}$$

to finally obtain

$$\frac{e^{\mu V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{\mu V_{jn}}}. \tag{13}$$

This is the logit model. Interestingly, it is a straightforward extension of the binary logit model, where the sum at the denominator involves now all alternatives in the choice set.

Properties of the extreme value distribution

The extreme value distribution with location parameter η and scale parameter μ has the following properties:

1. The mode is η.
2. The mean is $\eta + \frac{\gamma}{\mu}$, where

$$\gamma = -\int_0^{+\infty} e^{-x} \ln x dx \approx 0.5772 \tag{14}$$

is Euler's constant.

3. The variance is $\frac{\pi^2}{6\mu^2}$.
4. If $\varepsilon \sim \mathrm{EV}(\eta,\mu)$, then

$$a\varepsilon + b \sim \mathrm{EV}(a\eta + b, \frac{\mu}{a}),$$

where $a, b \in \mathbb{R}$, $a > 0$.

5. If $\varepsilon_a \sim \mathrm{EV}(\eta_a, \mu)$ and $\varepsilon_b \sim \mathrm{EV}(\eta_b, \mu)$ are independent with the same scale parameter μ, then

$$\varepsilon = \varepsilon_a - \varepsilon_b \sim \mathrm{Logistic}(\eta_a - \eta_b, \mu),$$

namely

$$f_\varepsilon(\xi) = \frac{\mu e^{-\mu(\xi - \eta_a + \eta_b)}}{(1 + e^{-\mu(\xi - \eta_a + \eta_b)})^2}, \tag{15}$$

$$F_\varepsilon(\xi) = \frac{1}{1 + e^{-\mu(\xi - \eta_a + \eta_b)}}, \ \mu > 0, -\infty < \xi < \infty. \tag{16}$$

(17)

6. If $\varepsilon_i \sim \text{EV}(\eta_i, \mu)$, for $i = 1, \ldots, J$, and ε_i are independent with the same scale parameter μ, then

$$\varepsilon = \max_{i=1,\ldots,J} \varepsilon_i \sim \text{EV}(\eta, \mu) \tag{18}$$

where

$$\eta = \frac{1}{\mu} \ln \sum_{i=1}^{J} e^{\mu\eta_i}. \tag{19}$$

It is important to note that this property holds only if all ε_i have the same scale parameter μ. As ε follows an extreme value distribution, its expected value is

$$E[\varepsilon] = \eta + \frac{\gamma}{\mu}.$$

Equivalently,

$$\eta = E[\varepsilon] - \frac{\gamma}{\mu}.$$

Therefore, (19) provides the expected value of the maximum, up to a constant.