Derivation of the Logit Choice Probabilities

Philip A. Viton

October 30, 2013

1 Introduction

This optional note presents an indirect approach to deriving the logit choice probabilities, via Bayes’ Theorem. I do it this way because the approach is widely useful, and is important, for example, in the derivation of the Mixed-Logit model, which is the most recent extension of logit-type mode-choice models to appear in the literature. A more direct derivation appears in Domencich and McFadden, Urban Travel Demand (p. 63), which is in the packet of readings for the course.

2 Bayes’ Theorem

The version of the theorem we shall be using involves the recovery of unconditional probabilities from conditional ones. Consider any event $A$ and let $B_1, B_2, \ldots, B_n$ be a collection of mutually exclusive and collectively exhaustive discrete events (ie, one of the $B_i$ must happen, and only one can happen). Then

\[
\Pr(A) = \Pr[\cup_i (A \& B_i)] = \sum_i \Pr(A \& B_i) = \sum_i \Pr(A \mid B_i) \cdot \Pr(B_i)
\]

Optional means that you will not be expected to reproduce this in an examination.
from the definition of conditional probability. In the continuous case (the one we’ll be using) this takes the form

\[
\Pr(A) = \int_{-\infty}^{\infty} f_{A|B} \cdot f_B
\]

where \( f_{A|B} \) is the conditional pdf of \( A \) given \( B \), and \( f_B \) is the marginal pdf of \( B \).

This formula says that if you know the conditional density of \( A \) given \( B \), and if you also know the marginal density of \( B \), then you can recover the density of \( A \) (by integration). In our case, the problem is to come up with an expression for the mode choice probabilities for the logit model. We’ll use Bayes’ Theorem here as follows:

- We argue that if the value of a certain random variable (\( B \)) were known, we could easily compute the choice probability \( \Pr(A) \). This will give us the conditional choice probability density \( f_{A|B} \).
- But the value of this random variable is not known. However, we do know its marginal density, \( f_B \). (By assumption, under the logit model this density will be T1EV, see below).
- Hence, using Bayes’ Theorem, we can compute the unconditional probability by integrating the product \( f_{A|B} \cdot f_B \).
- The final step is to carry out the integration. The result will be the logit formula for the choice probabilities.

3 The logit model

Consider an individual \( i \) making a discrete choice among \( J \) alternatives. The utility she obtains if alternative \( j \) is selected is

\[
\begin{align*}
    u_{ij} &= x_{ij} \beta + \varepsilon_{ij} \\
    &= v_{ij} + \varepsilon_{ij}
\end{align*}
\]

where \( v_{ij} \) is the systematic part of utility and, under the logit assumptions, the \( \varepsilon_{ij} \) are iid Type-1-Extreme-Value (T1EV) random variates; that is:

\[
\begin{align*}
    f_{\varepsilon_{ij}} &= e^{-e^{-\varepsilon_{ij}}} e^{-\varepsilon_{ij}} \\
    F_{\varepsilon_{ij}}(\alpha) &= \Pr[\varepsilon_{ij} \leq \alpha] = e^{-e^{-\alpha}}
\end{align*}
\]
Individual $i$ will choose alternative $j$ if this choice maximizes her utility. The probability of this event is:

$$P_{ij} = \Pr[u_{ij} > u_{i1} \& u_{ij} > u_{i2} \& \ldots \& u_{ij} > u_{ik} \& \ldots]$$

$$= \Pr[u_{ij} > u_{ik}, \text{ for all } k \neq j]$$

Now use the definition of $u_{ij}$ as $v_{ij} + \epsilon_{ij}$ to write this as:

$$P_{ij} = \Pr[v_{ij} + \epsilon_{ij} > v_{i1} + \epsilon_{i1} \& v_{ij} + \epsilon_{ij} > v_{i2} + \epsilon_{i2} \& \ldots]$$

$$= \Pr[\epsilon_{i1} < v_{ij} - v_{i1} \& \epsilon_{i2} < v_{ij} - v_{i2} \& \ldots]$$

$$= \Pr[\epsilon_{ik} < v_{ij} - v_{ik}, \text{ for all } k \neq j].$$

Note that in our work the values of the $v$’s will always be assumed to be known: these are the observable (systematic) determinants of choice.

### 3.1 The conditional choice probabilities

The result is that we can write:

$$P_{ij} = \Pr[\epsilon_{ik} < \epsilon_{ij} + v_{ij} - v_{ik}, \text{ for all } k \neq j].$$

Suppose we knew the value of the random variable $\epsilon_{ij}$ (part of the right-hand side of each of the inequalities) in the display. Then since all the $v$’s are also known, the entire right-hand side of each of the inequalities would be known, and therefore we could compute $P_{ij}$, since we are assuming that the $\epsilon_{ik}$ have independent T1EV densities. That is, we know the choice probability $P_{ij}$ conditional on knowing $\epsilon_{ij}$. If we write $B_k = \epsilon_{ij} + v_{ij} - v_{ik}$ and $P_{ij}|\epsilon_{ij}$ for the conditional probability, we have:

$$P_{ij}|\epsilon_{ij} = \Pr[\epsilon_{i1} < B_1 \& \epsilon_{i2} < B_2 \& \ldots]$$

$$= \Pr[\epsilon_{ik} < B_k, \text{ for all } k \neq j]$$

But the $\epsilon_{ik}$ are iid T1EV random variates, so this is (using independence) just the product of the individual densities, which (from equation (1)) is:

$$P_{ij}|\epsilon_{ij} = \Pr[\epsilon_{i1} < B_1] \cdot \Pr[\epsilon_{i2} < B_2] \cdot \ldots \cdot \Pr[\epsilon_{ik} < B_k] \cdot \ldots \cdot \Pr[\epsilon_{iJ} < B_J, \text{ for all } i \neq j]$$

$$= e^{-e^{-B_1}} \cdot e^{-e^{-B_2}} \cdot \ldots \cdot e^{-e^{-B_k}} \cdot \ldots \cdot e^{-e^{-B_J}}$$

$$= \prod_{k \neq j} e^{-e^{-B_k}}$$
3.2 The unconditional choice probabilities

The problem is, we do not know the value of \( \varepsilon_{ij} \). However, we do know its density: it is T1EV. So, using Bayes’ Theorem as stated above and the functional form for the T1EV density, the unconditional choice probability is:

\[
P_{ij} = \int_{-\infty}^{\infty} P_{ij|\varepsilon_{ij}} \cdot f_{\varepsilon_{ij}} \, d\varepsilon_{ij} = \int_{-\infty}^{\infty} \left( \prod_{k \neq j} e^{-e^{-B_k}} \right) \cdot e^{-e^{-\varepsilon_{ij}}} e^{-\varepsilon_{ij}} \, d\varepsilon_{ij}
\]

3.3 Evaluation of the integral

We shall now evaluate this integral.

First, let’s get rid of the restriction \( k \neq j \) in the product term. When \( k = j \) the term \(-B_j \) is just \(-\varepsilon_{ij} \). So if we include the \( k \) term \( \exp[-e^{-\varepsilon_{ij}}] \) in the product, and then multiply by \( \exp[e^{-\varepsilon_{ij}}] \) we preserve the equation. This yields:

\[
P_{ij} = \int_{-\infty}^{\infty} \left( \prod_{k=1}^{j} e^{-e^{-B_k}} \right) \cdot e^{-e^{-\varepsilon_{ij}}} e^{-\varepsilon_{ij}} \, d\varepsilon_{ij}
\]

and note that the first and third terms after the product cancel, leaving:

\[
P_{ij} = \int_{-\infty}^{\infty} \left( \prod_{k=1}^{j} e^{-e^{-B_k}} \right) \cdot e^{-\varepsilon_{ij}} \, d\varepsilon_{ij}.
\]

Second, look at the product term itself. This is:

\[
\prod_{k=1}^{j} e^{-e^{-B_k}} = e^{-e^{-B_1}} \cdot e^{-e^{-B_2}} \cdot \ldots \cdot e^{-e^{-B_j}}
\]

\[
= \exp \left[ - \sum_{k=1}^{j} e^{-B_k} \right] = \exp \left[ - \sum_{k=1}^{j} e^{\varepsilon_{ij} + b_{ij} - b_{ik}} \right] = \exp \left[ - e^{-\varepsilon_{ij}} \sum_{k=1}^{j} e^{(b_{ij} - b_{ik})} \right]
\]
so our choice probability may be written:

\[ P_{ij} = \int_{-\infty}^{\infty} \exp\left[ e^{-\varepsilon_{ij}} Q \right] e^{\varepsilon_{ij}} \, d\varepsilon_{ij} \]

where \( Q = \sum_{k=1}^{J} e^{-(v_{ij} - v_{ik})} \), independent of the variable of integration \( (\varepsilon_{ij}) \).

We shall now make the change of variable \( y = e^{-\varepsilon_{ij}} \). This transformation maps \([-\infty, \infty]\) onto \([0, \infty]\). The inverse transformation is \( \varepsilon_{ij} = -\ln y \), the Jacobian of the inverse transform is \( J = d\varepsilon_{ij} / dy = -1/y \), and since \( y > 0 \), the absolute value of the Jacobian is \( |J| = 1/y \). Then, under the change of variable,

\[ P_{ij} = \int_{0}^{\infty} e^{-Qy} \cdot y \cdot |J| \, dy = \int_{0}^{\infty} e^{-Qy} \, dy \]

But this is easy to evaluate by inspection. We get:

\[ P_{ij} = \left| \frac{1}{Q} \right| e^{-Qy} \bigg|_{0}^{\infty} = \frac{1}{Q} \]

so that, inserting the definition of \( Q \), our (unconditional) choice probability is:

\[ P_{ij} = \frac{1}{\sum_{k=1}^{J} e^{-(v_{ij} - v_{ik})}} = \frac{1}{e^{-v_{ij}} \sum_{k=1}^{J} e^{v_{ik}}} = \frac{1}{e^{v_{ij}} \sum_{k=1}^{J} e^{v_{ik}}} \]

which is the logit model’s formula for the choice probabilities.