Conditional Logit Models

James J. Heckman
University of Chicago

Econ 312, Spring 2023

1 The Weibull Distribution

Suppose ε is i.i.d. Weibull. Then the CDF of ε is given as

$$\Pr(\varepsilon < c) = F(c) = \exp\left(-\exp\left(-\left(c + \alpha_i\right)\right)\right)$$

where α_i is a parameter of the Weibull CDF. Also, by the assumption of independence and identical distribution, we can write

$$F\left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\right) = \prod_{i=1}^{n} F\left(\varepsilon_i\right) = \prod_{i=1}^{n} \exp\left(-\exp\left(-\left(\varepsilon_i + \alpha_i\right)\right)\right)$$

The Weibull distribution has two useful features. First, the difference between two Weibulls is a logit. Second, Weibulls are closed under maximization, since (assuming independence)

$$\begin{aligned}
\Pr\left(\max_i\{\varepsilon_i\} \leq \varepsilon\right) &= \prod_{i=1}^{n} \Pr(\varepsilon_i \leq \varepsilon) \\
&= \prod_{i=1}^{n} \exp\left(-\exp\left(-\left(\varepsilon+\alpha_i\right)\right)\right) \\
&= \exp\left(-\left(\sum_{i=1}^{n} \exp\left(-\left(\varepsilon+\alpha_i\right)\right)\right)\right) \\
&= \exp\left(-\left(\exp\left(-\varepsilon\right)\right)\sum_{i=1}^{n} \exp(-\alpha_i)\right)
\end{aligned} \tag{1}$$

Consider $\sum_{i=1}^{n} \exp(-\alpha_i)$. We can solve for α in the following equation:

$$\sum_{i=1}^{n} \exp(-\alpha_i) = \exp(-\alpha)$$

which implies

$$-\alpha = \log\left(\sum_{i=1}^{n} \exp(-\alpha_i)\right).$$

We can then substitute this value of α into equation (1) to get

$$\begin{aligned}
\Pr\left(\max_i\{\varepsilon_i\} \leq \varepsilon\right) &= \exp\left(-\left(\exp\left(-\varepsilon\right)\right)\exp(-\alpha)\right) \\
&= \exp\left(-\exp\left(-\left(\varepsilon+\alpha\right)\right)\right)
\end{aligned}$$

which is indeed a Weibull.

2 Random Utility Model

An individual with characteristics s has a choice set x, where $x \subseteq B$, B is a feasible set. We write

$$\Pr\left(x \mid s, B\right)$$

as the probability that a person of characteristics s chooses x from the feasible set. We also suppose that

$$U\left(s, x\right) = v(s, x) + \varepsilon(s, x)$$

where ε is i.i.d. Weibull. From our information on Weibulls in section 1, we know that $\varepsilon_i + v_i$, (and thus U_i), has a Weibull distribution with parameter $\alpha_i - v_i$. To see this,

$$\begin{aligned} \Pr\left(\varepsilon_i + v_i < \varepsilon\right) &= \Pr(\varepsilon_i < \varepsilon - v_i) \\ &= \exp\left(-\exp\left(-\left(\varepsilon + \alpha_i - v_i\right)\right)\right) \end{aligned}$$

Let us now suppose that there are two goods and two corresponding utilities. Consumers govern their choices by the obvious decision rule: choose good one if $U_1 > U_2$. More generally, if there are n goods, then good j will be selected if $U_j \in \operatorname{argmax} \{U_i\}_{i=1}^{n}$. Specifically, in our two good case,

$$\Pr(1 \text{ is chosen}) = \Pr(U_1 > U_2) = \Pr(\varepsilon_1 + v_1 > \varepsilon_2 + v_2).$$

Imposing that ε is i.i.d. Weibull, we can be much more precise about this probability,

$$\begin{aligned}
& \Pr\left(\varepsilon_1 + v_1 > \varepsilon_2 + v_2\right) \\
= & \Pr\left(\varepsilon_1 + v_1 - v_2 > \varepsilon_2\right) \\
= & \int_{-\infty}^{\infty} f(\varepsilon_1)\left(\int_{-\infty}^{\varepsilon_1+v_1-v_2} f(\varepsilon_2)d\varepsilon_2\right)d\varepsilon_1 \\
= & \int_{-\infty}^{\infty} f(\varepsilon_1)\exp\left(-\exp-\left(\varepsilon_1 + v_1 - v_2 + \alpha_2\right)\right)d\varepsilon_1. \quad (2)
\end{aligned}$$

Observe that $F(\varepsilon_1) = \exp\left(-\exp-\left(\varepsilon_1+\alpha_1\right)\right)$, which implies

$$\begin{aligned}
f(\varepsilon_1) &= \frac{\partial F(\varepsilon_1)}{\partial \varepsilon_1} = \exp\left(\exp-\left(\varepsilon_1+\alpha_1\right)\right)\left(\exp-\left(\varepsilon_1+\alpha_1\right)\right) \\
&= \exp-\left(\varepsilon_1+\alpha_1\right)\left(\exp\left(-\exp-\left(\varepsilon_1+\alpha_1\right)\right)\right).
\end{aligned}$$

Substituting this into (2), gives us

$$
\begin{aligned}
& \Pr\left(1 \text{ is chosen}\right) \\
= \; & \int_{-\infty}^{\infty} \exp-\left(\varepsilon_1+\alpha_1\right)\left(\exp\left(-\exp-\left(\varepsilon_1+\alpha_1\right)\right)\right) \\
& \quad \times \exp\left(-\exp-\left(\varepsilon_1+v_1-v_2+\alpha_2\right)\right) d\varepsilon_1 \\
= \; & e^{-\alpha_1}\int_{-\infty}^{\infty}\left(e^{-\varepsilon_1}\right) e^{[-\exp(-\varepsilon_1)][\exp(-\alpha_1)-\exp-(v_1-v_2+\alpha_2)]} d\varepsilon_1 \\
= \; & \exp\left(-\alpha_1\right)\left[\frac{1}{\exp\left(-\alpha_1\right)+\exp-\left(v_1-v_2+\alpha_2\right)}\right] \\
& \quad \times \left[e^{[-\exp(-\varepsilon_1)][\exp(-\alpha_1)-\exp-(v_1-v_2+\alpha_2)]}\right]_{-\infty}^{\infty} \\
= \; & \frac{\exp\left(-\alpha_1\right)}{\exp\left(-\alpha_1\right)+\exp-\left(v_1-v_2+\alpha_2\right)} \\
= \; & \frac{\exp(v_1-\alpha_1)}{\exp(v_1-\alpha_1)+\exp(v_2-\alpha_2)}
\end{aligned}
$$

This result generalizes, because the max over $(n-1)$ choices is still a Weibull so we can make a two stage maximization argument:

$$\begin{aligned}
& \Pr\left(\varepsilon_1 + v_1 > \varepsilon_i + v_i, \quad i = 1, 2, \ldots, n\right) \\
= \; & \Pr\left(\varepsilon_1 + v_1 > \max_{i=2,\ldots,n}\left(\varepsilon_i + v_i\right)\right) \\
= \; & \frac{\exp(v_1 - \alpha_1)}{\exp(v_1 - \alpha_1) + \exp(v_2 - \alpha_2) + \cdots + \exp(v_n - \alpha_n)} \\
= \; & \frac{\exp(\tilde{v}_1)}{\sum_{i=1}^{n} \exp\left(\tilde{v}_i\right)}
\end{aligned}$$

where $\tilde{v}_j = v_j - \alpha_j$.

3 Derivation of Logit

We will now show how the multinomial logit can be derived from Luce Axioms presented below.

3.1 Luce Axioms

Axiom 1 *Independence of Irrelevant Alternatives.*

Suppose that $x, y \in B$, $s \in S$. Then,

$$\Pr(x \mid s, \{x, y\}) \Pr(y \mid s, B) = \Pr(y \mid s, \{x, y\}) \Pr(x \mid s, B)$$

or,

$$\frac{\Pr(x \mid s, \{x, y\})}{\Pr(y \mid s, \{x, y\})} = \frac{\Pr(x \mid s, B)}{\Pr(y \mid s, B)}.$$

The term on the left is the odds ratio; the ratio of probabilities of choosing x to y given characteristics s and $\{x, y\}$. This axiom has been named "Independence of Irrelevant Alternatives" for an obvious reason—the odds of our choice are not effected by adding additional alternatives.

Axiom 2 *Positivity.*

$$\Pr(y \mid s, B) > 0 \quad \forall\, y \in B$$

With the preceding assumptions, we can now proceed to our derivation of the logit. Define $P_{yx} = \Pr(y \mid s, \{x, y\})$. Then by Axiom 1 above, we know

$$\left(\frac{P_{yx}}{P_{xy}}\right) \Pr(x \mid s, B) = \Pr(y \mid s, B) \tag{3}$$

Summing over y,

$$\Pr(x \mid s, B) \sum_{y \in B} \left(\frac{P_{yx}}{P_{xy}}\right) = 1 \Rightarrow \Pr(x \mid s, B) = \frac{1}{\sum_{y \in B} \left(\frac{P_{yx}}{P_{xy}}\right)}. \tag{4}$$

Again using Axiom 1, for $z \in B$,

$$\begin{aligned}
\left(\frac{P_{yz}}{P_{zy}}\right) \Pr\left(z \mid s, B\right) &= \Pr\left(y \mid s, B\right), \quad \text{and} \\
\left(\frac{P_{xz}}{P_{zx}}\right) \Pr\left(z \mid s, B\right) &= \Pr\left(x \mid s, B\right).
\end{aligned} \tag{5}$$

Substituting this in equation (3),

$$\left(\frac{P_{yx}}{P_{xy}}\right) = \frac{\Pr\left(y \mid s, B\right)}{\Pr\left(x \mid s, B\right)} = \frac{\left(\frac{P_{yz}}{P_{zy}}\right) \Pr\left(z \mid s, B\right)}{\left(\frac{P_{xz}}{P_{zy}}\right) \Pr\left(z \mid s, B\right)} = \frac{\frac{P_{yz}}{P_{zy}}}{\frac{P_{xz}}{P_{zx}}} \tag{6}$$

Write $v(s,x,z) = \ln \frac{P_{xz}}{P_{zx}}$, which implies $\frac{P_{xz}}{P_{zx}} = \exp\left(v(s,x,z)\right)$. Define a comparable expression for $\frac{P_{yz}}{P_{zy}}$. Replacing this into equation (5) produces

$$\frac{P_{yx}}{P_{xy}} = \frac{\exp\left(v(s,y,z)\right)}{\exp\left(v(s,x,z)\right)}$$

Thus from (4),

$$
\begin{aligned}
\Pr(x \mid s, B) &= \frac{1}{\sum_{y \in B} \left(\frac{\exp(v(s,y,z))}{\exp(v(s,x,z))} \right)} \\
&= \frac{1}{\left(\frac{1}{\exp(v(s,x,z))} \right) \sum_{y \in B} (\exp(v(s,y,z)))} \\
&= \frac{\exp(v(s,x,z))}{\sum_{y \in B} (\exp(v(s,y,z)))}.
\end{aligned}
$$

Assuming A-3: additive separability, $v(s, x, z) = v(s, x) - v(s, z)$. (This is equivalent to assuming irrelevance of the benchmark). From this assumption,

$$\begin{aligned}
\Pr(x \mid s, B) &= \frac{\exp\left(v(s, x) - v(s, z)\right)}{\sum_{y \in B}\left(\exp\left(v(s, y) - v(s, z)\right)\right)} \\
&= \frac{\exp v(s, x) \exp\left(-v(s, z)\right)}{\exp\left(-v(s, z)\right)\left(\sum_{y \in B}\exp\left(v(s, y)\right)\right)} \\
&= \frac{\exp v(s, x)}{\sum_{y \in B}\exp\left(v(s, y)\right)} \qquad (7)
\end{aligned}$$

which gives the multinomial logit. McFadden (1974) shows that Luce Axioms and a condition on ε ("Translation Completeness") produce the Weibull.

3.2 Consequences of Independence

We just showed

$$P_i = \frac{\exp(v_i)}{\sum_i \exp(v_i)}$$

so that

$$\frac{P_i}{P_j} = \frac{\frac{\exp(v_i)}{\sum_i \exp(v_i)}}{\frac{\exp(v_j)}{\sum_i \exp(v_i)}} = \frac{\exp(v_i)}{\exp(z_j)} = \exp\left(v_i - v_j\right) \Rightarrow \ln\left(\frac{P_i}{P_j}\right) = v_i - v_j.$$

A common specification for v_i is $v_i = z_i\beta$. Thus,

$$\ln\left(\frac{P_i}{P_j}\right) = (z_i - z_j)\,\beta \Rightarrow \frac{\partial \ln\left(\frac{P_i}{P_j}\right)}{\partial z_j} = -\beta$$

or, changes in characteristics z_j have a common effect on the ratio of log probabilities. This allows for estimation of the probabilities of purchasing a new good. (One could obtain an estimate of β from the existing goods. This estimate can then be combined with the characteristics, z_{new}, of the new good to estimate the probability of selection, as in equation (6)).

Further, from equation (6),

$$\Pr\left(2 \mid \{1,2\}\right) = \frac{e^{v_2}}{e^{v_1} + e^{v_2}}$$

and

$$\Pr\left(2 \mid \{1,2,3\}\right) = \frac{e^{v_2}}{e^{v_1} + e^{v_2} + e^{v_3}} < \Pr\left(2 \mid \{1,2\}\right).$$

This leads us to a restrictive property of the multinomial logit model—we have assumed independence of the ε_i, when in fact, they may be correlated. This is illustrated by McFadden's famous red bus, blue bus problem: Suppose we are modeling transportation choice and our alternatives consist of {car, bus, train}. If the alternatives are replaced by {car, red bus, blue bus}, then we have violated our assumption of dissimilar alternatives; if $U_2 > U_1$, then the event $U_3 > U_1$ is more likely. One can see by the preceding equation that adding more bus colors continually decreases the probability that car travel is chosen. We can deal with the problem of similar alternatives by using the nested logit model (Section 5) or the random coefficient probit model.

4 Probit: Random Coefficients

Suppose

$$U_i = Z_i \beta_i + \eta_i$$

where $\eta_i \sim N(0, \sigma_i^2)$, $\eta_i \perp\!\!\!\perp Z_j, \beta_i, \forall i, j$. Moreover, β is a random variable, with $\beta \sim (\bar{\beta}, \Sigma_\beta)$, so that

$$U_i = Z_i \left(\bar{\beta} + \varepsilon_i\right) + \eta_i = Z_i \bar{\beta} + Z_i \left(\beta - \bar{\beta}\right) + \eta_i.$$

It follows that

$$\begin{aligned}
& U_1 - U_2 \geq 0 \\
& \qquad \Longleftrightarrow (Z_1 - Z_2)\,\bar{\beta} + (Z_1 - Z_2)\left(\beta - \bar{\beta}\right) + (\eta_1 - \eta_2) \geq 0 \\
& U_1 - U_3 \geq 0 \\
& \qquad \Longleftrightarrow (Z_1 - Z_3)\,\bar{\beta} + (Z_1 - Z_3)\left(\beta - \bar{\beta}\right) + (\eta_1 - \eta_3) \geq 0
\end{aligned}$$

Further,

$$
\begin{aligned}
& \operatorname{Var}\left(U_1-U_2\right) \\
= \; & E\left\{\left[\left(U_1-U_2\right)-E\left(U_1-U_2\right)\right]^{\prime}\left[\left(U_1-U_2\right)-E\left(U_1-U_2\right)\right]\right\} \\
= \; & E\left\{\begin{array}{c}
\left[\left(Z_1-Z_2\right)\left(\beta-\bar{\beta}\right)+\left(\eta_1-\eta_2\right)\right]^{\prime} \\
\times\left[\left(Z_1-Z_2\right)\left(\beta-\bar{\beta}\right)+\left(\eta_1-\eta_2\right)\right]
\end{array}\right\} \\
= \; & E\left(\begin{array}{c}
\left(Z_1-Z_2\right)\left(\beta-\bar{\beta}\right)\left(\beta-\bar{\beta}\right)^{\prime}\left(Z_1-Z_2\right)^{\prime} \\
+\left(\eta_1-\eta_2\right)\left(\eta_1-\eta_2\right)^{\prime}
\end{array}\right) \\
= \; & \left(Z_1-Z_2\right) \Sigma_\beta\left(Z_1-Z_2\right)^{\prime}+\sigma_1^2+\sigma_2^2
\end{aligned}
$$

and similarly,

$$
\operatorname{Var}\left(U_1-U_3\right)=\left(Z_1-Z_3\right) \Sigma_\beta\left(Z_1-Z_3\right)^{\prime}+\sigma_1^2+\sigma_3^2
$$

Thus,

$$Cov\left(U_1 - U_2, U_1 - U_3\right) = \left(Z_1 - Z_2\right)\Sigma_\beta\left(Z_1 - Z_3\right)' + \sigma_1^2$$

so

$$\rho = \operatorname{Corr}\left(U_1 - U_2, U_1 - U_3\right) = \frac{\left(Z_1 - Z_2\right)\Sigma_\beta\left(Z_1 - Z_3\right)' + \sigma_1^2}{\sqrt{\operatorname{Var}\left(U_1 - U_2\right)\operatorname{Var}\left(U_1 - U_3\right)}}$$

We now seek to derive the probability of choosing good 1 in a three good case,

$$\Pr\left(\text{choosing } 1\right) = \Pr\left(U_1 - U_2 \geq 0 \text{ and } U_1 - U_3 \geq 0\right).$$

From before, we know that

$$\begin{aligned} U_1 - U_2 &\sim N\left(\left(Z_1 - Z_2\right)\bar{\beta}, \operatorname{Var}\left(U_1 - U_2\right)\right) \\ U_1 - U_3 &\sim N\left(\left(Z_1 - Z_3\right)\bar{\beta}, \operatorname{Var}\left(U_1 - U_3\right)\right). \end{aligned}$$

Thus,

$$
\begin{aligned}
&\Pr\left(U_1 - U_2 \geq 0 \text{ and } U_1 - U_3 \geq 0\right) \\
&\quad = \Pr\Big[\sqrt{\operatorname{Var}\left(U_1 - U_2\right)}t_1 + \left(Z_1 - Z_2\right)\bar{\beta} \geq 0 \\
&\quad\quad \text{and } \sqrt{\operatorname{Var}\left(U_1 - U_3\right)}t_2 + \left(Z_1 - Z_3\right)\bar{\beta} \geq 0\Big]
\end{aligned}
$$

where t_1 and t_2 are standard normal. Thus, the above equation reduces to

$$
\begin{aligned}
&\Pr\left(t_1 \geq -\frac{\left(Z_1 - Z_2\right)\bar{\beta}}{\sqrt{\operatorname{Var}\left(U_1 - U_2\right)}} \text{ and } t_2 \geq -\frac{\left(Z_1 - Z_3\right)\bar{\beta}}{\sqrt{\operatorname{Var}\left(U_1 - U_3\right)}}\right) \\
&\quad = \Pr\left(t_1 \leq \frac{\left(Z_1 - Z_2\right)\bar{\beta}}{\sqrt{\operatorname{Var}\left(U_1 - U_2\right)}} \text{ and } t_2 \leq \frac{\left(Z_1 - Z_3\right)\bar{\beta}}{\sqrt{\operatorname{Var}\left(U_1 - U_3\right)}}\right)
\end{aligned}
$$

As t_1 and t_2 may be correlated, we integrate over the joint density to get the probability

$$\begin{aligned}
&\Pr\left(\text{choosing } 1\right) \\
&= \int_{-\infty}^{a}\left(\int_{-\infty}^{b}\left(\frac{1}{2\pi\sqrt{1-\rho^2}}\exp-\frac{1}{2}\left(\frac{t_1^2-2\rho t_1 t_2+t_2^2}{1-\rho^2}\right)\right)dt_2\right)dt_1
\end{aligned}$$

where

$$a=\frac{\left(Z_1-Z_2\right)\bar{\beta}}{\sqrt{\mathrm{Var}\left(U_1-U_2\right)}},\ \text{and } b=\frac{\left(Z_1-Z_3\right)\bar{\beta}}{\sqrt{\mathrm{Var}\left(U_1-U_3\right)}}.$$

Now consider adding a third good to the two good case. If the third good has identical characteristics as the first, then $Z_2 = Z_3$. If there is no stochastic component (no utility innovation), then $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0$. Therefore, in this case,

$$\begin{aligned}
\Pr(1 \text{ chosen}) &= \Pr(U_1 - U_2 \geq 0 \text{ and } U_1 - U_3 \geq 0) \\
&= \Pr(U_1 - U_2 \geq 0).
\end{aligned}$$

Thus, there is no change in the probability of choosing good 1 despite the addition of a third good.

Again focusing on the two good case, we observe

$$\Pr(1 \text{ chosen}) = \Pr(U_1 - U_2 \geq 0) = \Pr\left(z \leq \frac{(Z_1 - Z_2)\,\bar{\beta}}{\sqrt{\mathrm{Var}\,(U_1 - U_2)}}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{(Z_1 - Z_2)\bar{\beta}}{[(Z_1 - Z_2)\Sigma_\beta (Z_1 - Z_2)' + \sigma_1^2 + \sigma_2^2]^{1/2}}} \left(\exp\left(-\frac{t^2}{2}\right) dt\right),$$

which can be evaluated to derive the desired probability.

Finally, consider a McFadden-Luce type of set up, where one imposes $\Sigma_\beta = 0$. Defining $\sigma^* = \sqrt{\sigma_1^2 + \sigma_2^2}$, we observe that the probability of choosing good 1 in the two-good case is

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{(Z_1 - Z_2)\bar{\beta}}{\sigma^*}} \left(\exp\left(-\frac{t^2}{2}\right) dt\right).$$

Adding a third good to the scene with identical characteristics, $(Z_2 = Z_3)$, yields the probability for good 1 being purchased as

$$\int_{-\infty}^{\frac{(Z_1-Z_2)\bar{\beta}}{\sigma^*}} \int_{-\infty}^{\frac{(Z_1-Z_2)\bar{\beta}}{\sigma^*}} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\left(\frac{t_1^2 - 2\rho t_1 t_2 + t_2^2}{1-\rho^2}\right)\right] dt_1 dt_2.$$

One can show that, upon evaluation of these integrals, the probability derived from addition of the third good is less than the probability in the two good case. This leads us to a similar problem as the multinomial logit - adding alternatives decreases the probability of choice, despite the fact that the alternatives are quite similar.

5 Nested Logit

5.1 Generalized Extreme Value (GEV) Model

Consider a function $G(y_1, y_2, \ldots, y_J)$, where G satisfies:

i. Non-negativity:

$$G\left(y_1, y_2, \ldots, y_J\right) \geq 0 \quad \forall \left(y_1, y_2, \ldots, y_J\right) \geq 0.$$

ii. Homogeneous of degree 1:

$$G\left(\alpha y_1, \alpha y_2, \ldots, \alpha y_J\right) = \alpha G\left(y_1, y_2, \ldots, y_J\right).$$

iii.

$$\begin{aligned} \frac{\partial^k G}{\partial y_1 \partial y_2 \ldots \partial y_J} &\geq 0 \quad \text{if } k \text{ even} \\ &\leq 0 \quad \text{if } k \text{ odd.} \end{aligned}$$

If G satisfies these conditions, then we get the following probability:

$$P_i = \frac{y_i G_i\left(y_1, y_2, \ldots, y_J\right)}{G\left(y_1, y_2, \ldots, y_J\right)},$$

where P_i is a probability that can be derived from utility maximization. We can use the theorem above to derive a special case of the nested logit model.

Define

$$\begin{aligned} & G\left(\exp(v_1), \exp(v_2), \ldots, \exp(v_J)\right) \\ = \; & \exp(v_1) \\ & + \left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right) + \cdots + \exp\left(\frac{v_J}{1-\sigma}\right)\right]^{1-\sigma}. \end{aligned}$$

Observe that $\sigma = 0$ is the ordinary logit model. (With G defined in this way, we are assuming that ε_1 is uncorrelated with all of the other ε_j, while the remaining ε_i may be correlated).

This function obviously meets the conditions for the GEV model. For,

i. Non-negativity: obvious as $0 < \sigma < 1$

ii. Homogeneity:

$$
\begin{aligned}
& G\left(\alpha \exp(v_1), \alpha \exp(v_2), \ldots, \alpha \exp(v_J)\right) \\
= \;& \alpha \exp(v_1) + \left((\alpha \exp(v_2))^{\frac{1}{1-\sigma}} + \cdots + (\alpha \exp\left(v_J\right))^{\frac{1}{1-\sigma}} \right)^{1-\sigma} \\
= \;& \alpha \exp(v_1) + \left(\begin{array}{c} \left(\alpha^{\frac{1}{1-\sigma}}\right) (\exp(v_2))^{\frac{1}{1-\sigma}} \\ + \cdots + \left(\alpha^{\frac{1}{1-\sigma}}\right) \left(\exp\left(v_J\right)\right)^{\frac{1}{1-\sigma}} \end{array} \right)^{1-\sigma} \\
= \;& \alpha \exp(v_1) + \alpha \left(\begin{array}{c} (\exp(\frac{v_2}{1-\sigma})) \\ + \cdots + \left(\exp\left(\frac{v_J}{1-\sigma}\right)\right) \end{array} \right)^{1-\sigma} \\
= \;& \alpha \left(\exp(v_1) + \left[\begin{array}{c} \exp\left(\frac{v_2}{1-\sigma}\right) \\ + \cdots + \exp\left(\frac{v_J}{1-\sigma}\right) \end{array} \right]^{1-\sigma} \right) \\
= \;& \alpha \left(G\left(\exp(v_1), \exp(v_2), \ldots, \exp(v_J)\right)\right).
\end{aligned}
$$

iii. By inspection, one can see that this derivative property will hold. (It is obvious when differentiating with respect to $\exp(v_1)$. For other derivatives, the fact that $0 < \sigma < 1$ gives the needed alternation in sign).

Thus, we can now proceed to derive our probabilities. First, consider

$$\begin{aligned}
\Pr\left(1 \mid \{1,2\}\right) &= \frac{e^{v_1}}{e^{v_1} + \left(e^{\frac{v_2}{1-\sigma}}\right)^{1-\sigma}} \\
&= \frac{e^{v_1}}{e^{v_1} + e^{v2}},
\end{aligned}$$

which is simply our binomial logit model.

Also note that in the three good case,

$$
\begin{aligned}
G_2 &= (1-\sigma)\left(\exp(v_1) + \left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right)\right]^{-\sigma}\right) \\
&\qquad \times \frac{1}{1-\sigma}\exp\left(\frac{\sigma v_2}{1-\sigma}\right) \\
&= \exp\left(\frac{\sigma v_2}{1-\sigma}\right) \\
&\qquad \times \left(\exp(v_1) + \left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right)\right]^{-\sigma}\right).
\end{aligned}
$$

Now suppose that we eliminate choice 1 (by letting $v_1 \to -\infty$). Then,

$$
\begin{aligned}
& \Pr(2 \mid \{2,3\}) \\
= & \frac{\exp(v_2)\exp\left(\frac{\sigma v_2}{1-\sigma}\right)\left[\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right]^{-\sigma}}{\left[\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right]^{1-\sigma}} \\
= & \frac{\exp\left(\frac{v_2}{1-\sigma}\right)}{\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)}.
\end{aligned}
$$

Observe

$$\begin{aligned}
\Pr\left(1 \mid \{1,2,3\}\right) &= \frac{e^{v_1}}{e^{v_1} + \left(e^{\frac{v2}{1-\sigma}} + e^{\frac{v3}{1-\sigma}}\right)^{1-\sigma}} \\
&= \frac{e^{v_1}}{e^{v_1} + \left\{e^{\frac{v2}{1-\sigma}}\left(1 + e^{\frac{v_3 - v_2}{1-\sigma}}\right)\right\}^{1-\sigma}} \\
&= \frac{e^{v_1}}{e^{v_1} + e^{v_2}\left(1 + \left(\frac{e^{v_3}}{e^{v_2}}\right)^{\frac{1}{1-\sigma}}\right)^{1-\sigma}}. \qquad (8)
\end{aligned}$$

Letting $\sigma \to 1$, and supposing $e^{v_2} > e^{v_3}$, we get

$$\left(\frac{e^{v_3}}{e^{v_2}}\right) < 1 \Rightarrow \left(\frac{e^{v_3}}{e^{v_2}}\right)^{\frac{1}{1-\sigma}} \to 0,$$

and thus from equation (8),

$$\Pr\left(1 \mid \{1,2,3\}\right) \to \frac{e^{v_1}}{e^{v_1} + e^{v_2}}. \tag{9}$$

Conversely, if $e^{v_3} > e^{v_2}$, just reverse the roles of v_2 and v_3 so

$$\begin{aligned} \Pr\left(1 \mid \{1,2,3\}\right) &\to \frac{e^{v_1}}{e^{v_1} + e^{v_2}\left(\dfrac{e^{v_3}}{e^{v_2}}\right)} \\ &= \frac{e^{v_1}}{e^{v_1} + e^{v_3}}. \end{aligned} \tag{10}$$

Combining equations (9) and (10), we get, as $\sigma \to 1$

$$\Pr\left(1 \mid \{1,2,3\}\right) \to \frac{e^{v_1}}{e^{v_1} + \max\{e^{v_2}, e^{v_3}\}}. \tag{11}$$

Similarly, we find

$$
\begin{aligned}
& \Pr\left(2 \mid \{1,2,3\}\right) && (12)\\
= & \frac{e^{v_2}\left[(1-\sigma)\left\{\begin{array}{c}\exp\left(\dfrac{v_2}{1-\sigma}\right)\\ +\exp\left(\dfrac{v_3}{1-\sigma}\right)\end{array}\right\}^{-\sigma}\right]^{\frac{1}{1-\sigma}}\exp\left(\dfrac{\sigma v_2}{1-\sigma}\right)}{e^{v_1}+\left\{\exp\left(\dfrac{v_2}{1-\sigma}\right)+\exp\left(\dfrac{v_3}{1-\sigma}\right)\right\}^{1-\sigma}}\\
= & \frac{\exp\left(\frac{v_2}{1-\sigma}\right)\left\{\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right\}^{-\sigma}}{e^{v_1}+\left\{\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right\}^{1-\sigma}}\\
= & \frac{\exp\left(\frac{v_2}{1-\sigma}\right)}{\left(e^{v_1}+\left\{\begin{array}{c}\exp\left(\frac{v_2}{1-\sigma}\right)\\ +\exp\left(\frac{v_3}{1-\sigma}\right)\end{array}\right\}^{1-\sigma}\right)\left(\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right)^{\sigma}}.
\end{aligned}
$$

When $\sigma = 0$ we have ordinary conditional logit. Suppose $v_2 > v_3$ and $\sigma \to 1$. By appealing to the result derived in equation (11),

$$P(2 \mid \{1,2,3\}) = \left(\frac{e^{\frac{v_2}{1-\sigma}}}{e^{\frac{v_2}{1-\sigma}} + e^{\frac{v_3}{1-\sigma}}} \right)$$

$$\cdot \left[\frac{\left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right)\right]^{\frac{1}{1-\sigma}}}{\exp v_1 + \left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right)\right]^{\frac{1}{1-\sigma}}} \right]$$

for $v_2 > v_3$,

$$\left(\frac{e^{v_3}}{e^{v_2}}\right) < 1 \Rightarrow \left(\frac{e^{v_3}}{e^{v2}}\right)^{\frac{1}{1-\sigma}} \to 0,\ \ as\ \sigma \to 1$$

and thus, from equation (12),

$$\Pr\left(2 \mid \{1,2,3\}\right) \to \frac{\exp(v_2)}{\exp(v_1) + \exp(v_2)},\ \ as\ \sigma \to 1.$$

(One could derive a similar result be assuming that $v_3 > v_2$). Finally, suppose that $v_2 = v_3$. then,

$$
\begin{aligned}
G &= e^{v_1} + \left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_2}{1-\sigma}\right)\right]^{1-\sigma} \\
&= e^{v_1} + \left[2\exp\left(\frac{v_2}{1-\sigma}\right)\right]^{1-\sigma} \\
&= \exp(v_1) + 2^{1-\sigma}\exp\left(v_2\right).
\end{aligned}
$$

Thus,

$$\begin{aligned}
\Pr\left(2 \mid \{1,2,3\}\right) &= \frac{\exp\left(v_2\right) 2^{-\sigma}}{\exp(v_1) + 2^{1-\sigma} \exp\left(v_2\right)} \\
&= \frac{\exp v_2}{2^{\sigma} \exp v_1 + 2 \exp v_2} \\
\lim_{\sigma \to 1} \Pr\left(2 \mid \{1,2,3\}\right) &\to \frac{1}{2} \frac{\exp(v_2)}{\exp(v_1) + \exp(v_2)}.
\end{aligned}$$

This final equation tells us if the characteristics are identical in the nested logit model, then the probability in the three choice case reduces to the binomial logit—the probability of the two choice case.