A Discrete Choice Model with Residual Uncertainty

1 Framework

Assume some universal partitioning of the unit interval into W disjoint intervals with cutoffs denoted by the $(W+1)$-dimensional vector $\boldsymbol{\pi} \in \Delta^{W-1}$. Let $\boldsymbol{\alpha}$ denote the partial sums of $\boldsymbol{\pi}$, so that

$$\boldsymbol{\alpha} = \{0, \quad \pi_1, \quad \pi_1 + \pi_2, \quad \dots, \quad \underbrace{\textstyle\sum_{i=1}^{w} \pi_i}_{\alpha_w}, \quad \dots, \quad 1\} \tag{1}$$

$$\mathcal{W}_w = [\alpha_{w-1}, \alpha_w) \tag{2}$$

$$\bigcup_{w=1}^{W} \mathcal{W}_i = [0, 1) \tag{3}$$

Consumer i's utility for good j is given by

$$U_{ij} = X_j{}' \beta_i + \eta_{ij}, \quad \eta_{ij} \sim \text{Gumbel}(0, 1) \tag{4}$$

where X is the design matrix of good characteristics and X^j denotes the j^{th} row of X. The "zero"-th good is special and is associated with a zero vector of good characteristics. Consumers observe X and η for all goods $j \in \mathcal{J}$, but they *do not* observe η_{i0}.[1]

First, consumers report their preferred good in $\mathcal{J} = \{1, 2, \dots, J\}$, which is given by

$$t_i = e_{j_i^*} \text{ s.t. } j_i^* = \arg\max_{j \in \mathcal{J}} U_{ij}. \tag{5}$$

Let V_i denote the utility of good j_i^* for consumer i.

Second, consumers report the probability that they prefer good j_i^* to the outside good 0. Consumers know V_i, but do not know their draw of η_{i0}, and thus this probability is given by

$$p_i = \mathbb{P}(\eta_{i0} < V_i). \tag{6}$$

Each consumer reports the interval $\mathcal{W}_w$ into which p_i falls

$$w_i = w \text{ s.t. } p_i \in [\alpha_{w-1}, \alpha_w). \tag{7}$$

2 Deriving the Likelihood

As is well known,[2] V_i follows a Gumbel distribution with location parameter $\overline{\mu}_i$ and scale parameter 1, where

$$\overline{\mu}_i = \ln\left(\sum_{j \in \mathcal{J}} \exp\left(X_j{}' \beta_i\right)\right). \tag{8}$$

Under these assumptions, we have

$$p_i = F_{\text{Gumbel}(0,1)}(V_i) \tag{9}$$

The consumer's report of w_i is therefore equivalent to reporting that

$$V_i \in \left[F^{-1}_{\text{Gumbel}(0)}\left(\alpha_{w(i)-1}\right), F^{-1}_{\text{Gumbel}(0)}\left(\alpha_{w(i)}\right)\right), \tag{10}$$

1. This can be motivated by a framework in which $U_0 = 0$ and $U_{ij} = X_j{}'\beta_i + \eta_{ij} - \eta_0$ for $j \in \mathcal{J}$. Here, η_0 captures the consumer's uncertainty about their future tastes. Given that utilities are ordinal and η_0 is a common shock, it plays no role in the choice among the most preferred good $j^* \in \mathcal{J}$.

2. For example, see McFadden (1981) and Cardell (1997).

where we omit the common scale parameter for brevity.

This occurs with probability

$$\mathbb{P}\Big(V_i \in \mathcal{W}_{w(i)}\Big) = F_{\text{Gumbel}(\overline{\mu})}\Big(F^{-1}_{\text{Gumbel}(0)}\Big(\alpha_{w(i)}\Big)\Big) - F_{\text{Gumbel}(\overline{\mu})}\Big(F^{-1}_{\text{Gumbel}(0)}\Big(\alpha_{w(i)-1}\Big)\Big) \tag{11}$$

$$= \Big(\alpha_{w(i)}\Big)^{\exp(\overline{\mu})} - \Big(\alpha_{w(i)-1}\Big)^{\exp(\overline{\mu})}. \tag{12}$$

This is $p(w_i \mid \alpha, \beta)$, the conditional likelihood of w_i.[3]

3 Sketch of Estimation

- Specify hyper-parameters governing the prior distribution of $(\boldsymbol{\pi}, \boldsymbol{\beta})$. There is a relatively large amount of flexibility in the prior distribution over β. $\boldsymbol{\pi} \sim$ Dirichlet deterministically maps to $\boldsymbol{\alpha}$.
- (First Branch) Given draws of (α, β), the probability that consumer i reports j_i^* is given by a softmax

$$p(j_i^* \mid \alpha, \beta) = p(j_i^* \mid \beta) = \frac{\exp\left({X_j}'\beta_i\right)}{\sum_{j' \in \mathcal{J}} \exp\left({X_{j'}}'\beta_i\right)} \tag{13}$$

- (Second Branch) Given (α, β), the probability that consumer i reports w_i is given by

$$p(w_i \mid \alpha, \beta, j_i^*) = p(w_i \mid \alpha, \beta) = \Big(\alpha_{w(i)}\Big)^{\exp(\overline{\mu})} - \Big(\alpha_{w(i)-1}\Big)^{\exp(\overline{\mu})} \tag{14}$$

  where $\overline{\mu}_i(\beta_i)$ is a function of consumer i's tastes β_i and the design matrix X provided in Equation 8.[4]

Thus the overall likelihood (conditional on some draw of parameters) looks like

$$p((j_i^*, w_i) \mid \alpha, \beta) = p\big(w_i \mid \alpha, \overline{\mu}_i(\beta)\big) \times p(j^* \mid \beta). \tag{15}$$

Comment: Relation to the Brazell et al. (2006) Dual Response Model

Another common framework when soliciting consumer preferences is to directly ask consumers if $V_i \geq 0$. This implicitly assumes that $\eta_{i0} = 0$, so that consumers deterministically know whether or not the "inside" good j_i^* is preferred to the outside good. As we have already noted, our framework can nest that standard case by assuming that, rather than following a standard Gumbel distribution, η_0 is instead a degenerate distribution. In such a case, all but 2 of the W partitions are empty, and the non-empty partitions are $\mathcal{W}_0 = \{0\}$ and $\mathcal{W}_W = \{1\}$, as $p_i \in \{0, 1\}$.

We acknowledge that this requires a slight abuse of notation, as $\mathcal{W}_w$ was defined above using left-closed intervals. These have the advantage of being invertible under the inverse-CDF mapping. In the degenerate case, we instead have $p(w_i = W \mid \alpha, \beta) = \mathbb{P}(\eta_{i0} < V_i) = \mathbb{P}(V_i > 0) = 1 - \exp\left(-\exp\left(\overline{\mu}\right)\right)$.

3. While the parametrization of $\eta_0 \sim \text{Gumbel}(0, 1)$ preserves symmetry among the $J + 1$ goods and is thus a natural choice, the framework can easily accommodate an alternative distribution for η_0. For example, one could use an affine function of individual characteristics to accommodate individual-level variation in the propensity to prefer the outside good.

4. Note that the observed choice j_i^* is irrelevant for this likelihood through "Gumbel distribution magic". Specifically, it is the "memory-less" property of the Gumbel that yields this property.

References

Brazell, Jeff D., Christopher G. Diener, Ekaterina Karniouchina, William L. Moore, Válerie Séverin, and Pierre-Francois Uldry. 2006. "The no-choice option and dual response choice designs." *Marketing Letters* 17, no. 4 (December): 255–268. ISSN: 1573-059X. https://doi.org/10.1007/s11002-006-7943-8. http://dx.doi.org/10.1007/s11002-006-7943-8. (Cited on page 2).

Cardell, N. Scott. 1997. "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity." *Econometric Theory* 13, no. 2 (April): 185–213. ISSN: 1469-4360. https://doi.org/10.1017/s0266466600005727. (Cited on page 1).

McFadden, Daniel L. 1981. "Structural Discrete Probability Models Derived from Theories of Choice" [in English]. In *Structural Analysis of Discrete Data with Econometric Applications,* 198–272. The MIT Press. ISBN: 9780262131599. (Cited on page 1).