Problem Set 1 Due: in class Monday Sept 9th

John Rust, Professor of Economics Georgetown University Fall 2019

I. This problem set leads you through the steps to derive the multinomial logit model (MNL) as a random utility model with additive random utilities that have independent Type 1 extreme value distributions. Recall from the notes that the MNL model was initially derived by the mathematical psychologist Duncan Luce based on the axiom of *independence from irrelevant alternatives* (IIA), leading to this formula for the conditional choice probability

$$P(d|x) = \frac{\exp\{u(x,d)\}}{\sum_{d' \in D(x)} \exp\{u(x,d')\}} \tag{1}$$

where $u(x,d)$ is some function of x and the discrete choice d, and for each x, $D(x)$ is a finite choice set. McFadden derived the MNL model as a *random utility model* (RUM). That is, Mcfadden (following a lead of Thurstone who worked on similar models in the 1930s) assumed that the choice of an alternative $d \in D(x)$ is not only affected by x but also by a *random utility component* $\varepsilon(d)$ that reflects other factors and "states" of the individual making the choice that the econometrician does not observe. So the individual's choice is governed by a *decision rule* $d(x,\varepsilon)$ that depends on a vector of variables that the econometrician can observe (both states of the individual and potentially characteristics or "attributes" of the items the individual is choosing between) given by

$$d(x,\varepsilon) = \underset{d \in D(x)}{argmax}[u(x,d) + \varepsilon(d)] \tag{2}$$

Note that McFadden's formulation of the random utility model invokes an implicit *additive separability* (AS) assumption: the utility, which might be written in general as $u(x,\varepsilon,d)$ takes the specific additively separable form $u(x,d) + \varepsilon(d)$ where we make an assumption about the *probability distribution* of the unobserved components of the utility function, $\varepsilon \equiv \{\varepsilon(d)|d \in D(x)\}$. If ε is a multivariate continuous random vector with full support over $R^{|D(x)|}$ (where $|D(x)|$ is the number of elements in the choice set), with CDF $F(\varepsilon|x)$ then it is not hard to show that the implied *conditional choice probability* $P(d|x)$ given by

$$P(d|x) = \int_{\varepsilon} I\{d(x,\varepsilon) = d\}\, dF(\varepsilon|x) \tag{3}$$

satisfies $P(d|x) > 0$ for each $d \in D(x)$. That is, there is positive probability of observing the individual choosing any alternative $d \in D(x)$. McFadden showed that if ε has a multivariate Type 1 extreme value distribution (also called a Gumbel distribution), with CDF given by

$$F(\varepsilon) = \prod_{d \in D(x)} \exp\{-\exp\{-(\varepsilon(d) - \mu(d))/\sigma\}\} \tag{4}$$

where $\mu(d) \in (-\infty,\infty)$ is the *location parameter* of the continuous random variable $\varepsilon(d)$ and σ is the *scale parameter.*

A. Show that $F(\varepsilon)$ is a valid multivariate CDF and shows its support is all of $R^{|D(x)|}$. Are the random variables $\{\varepsilon(d)|d \in D(x)\}$ *IID*? Are they independently distributed?

B. Show that the collection of extreme value distriubuted random variables $\{\varepsilon(d)|d \in D(x)\}$ is *max-stable* i.e. show that $\eta = \max\{\varepsilon(d)|d \in D(x)\}$ is also a Type 1 extreme value distribution and calculate its location parameter μ and scale parameter σ.

C. Using the result in part B and the fact that if η is a Type 1 extreme value distribution with location parameter μ and scale parameter σ its mean is given by

$$E\{\eta\} = \mu + \gamma\sigma \tag{5}$$

where $\gamma \simeq 0.577\ldots$ is *Euler's constant* and

$$\mathrm{var}(\eta) = \sigma^2 \frac{\pi^2}{6} \tag{6}$$

write a formula for the *expected maximum utility*

$$E\left\{\max_{d\in D(x)} [u(x,d) + \varepsilon(d)]\right\} \tag{7}$$

which McFadden called the *social surplus function.* Why would he call it that? Is there any analogy you can draw between the social surplus function and the *indirect utility function* of consumer theory?

D. Now forget about the Type 1 extreme value distribution for a moment, and suppose the unobserved additively separable components of utility $\varepsilon = \{\varepsilon(d) | d \in D(x)\}$ could be any continuous multivariate distribution with CDF $F(\varepsilon|x)$ such as a multivariate normal distribution. Show under as much generality as you can that the *Williams-Daly-Zachary Theorem* holds:

$$P(d|x) = \frac{\partial}{\partial u(x,d)} E\left\{\max_{d\in D(x)} [u(x,d) + \varepsilon(d)]\right\}. \tag{8}$$

HINT: Use the *Lebesgue Dominated Convergence Theorem* to show you can interchange the partial derivative and expectation operators in equation (8) to get equation (3). Can you draw a further analogy between equation (8) and Roy's Identity?

E. Now using the result you derived in part C where you (hopefully) were able to derive a closed form expression for the social surplus function, use the Williams-Daly-Zachary Theorem (8) to show McFadden's result, namely that Type 1 extreme value distributed preference shocks (random utilities) results in the classic multinomial logit model formula (1), except that the utilities $u(x,d)$ have to be divided by the scale parameter σ.

F. Show that the MNL model is the same as what people in the machine learning literature refer to as the *soft-max function* and prove this

$$\lim_{\sigma\downarrow 0} P(d|x) = \begin{cases} 1/n & \text{if } u(x,d) \geq u(x,d') \quad \forall d' \in D(x) \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where n is the number of alternatives in $D(x()$ that achieve the maximal utility. Similarly, we might call the social surplus function for the Type 1 extreme value distribution the *smoothed max function* since you should also prove that

$$\lim_{\sigma\downarrow 0} E\left\{max_{d\in D(x)}[u(x,d) + \varepsilon(d)]\right\} = \max_{d\in D(x)} [u(x,d)]. \tag{10}$$

G. Write down the Axiom of Independence from Irrelevant Alternatives and show that the MNL model satisfies this axiom. Provide an example of a random utility model that does not satisfy the IIA axiom. Is the IIA axiom "reasonable" and consistent with "rational choice" or does it imply unrealistic restrictions on choice probabilities?