The Journal of Choice Modelling 19 (2016) 40–53

Contents lists available at ScienceDirect

The Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Mixed logit with a flexible mixing distribution ☆

Kenneth Train

Department of Economics, University of California, Berkeley, United States

ARTICLE INFO

Article history:
Received 28 February 2016
Received in revised form
19 July 2016
Accepted 21 July 2016
Available online 3 August 2016

Keywords:
Mixed logit
Mixing distribution
Nonparametric

ABSTRACT

This paper presents a flexible procedure for representing the distribution of random parameters in mixed logit models. A logit formula is specified for the mixing distribution, in addition to its use for the choice probabilities. The properties of logit assure positivity and provide the normalizing constant for the mixing distribution. Any mixing distribution can be approximated to any degree of accuracy by this specification. The researcher defines variables to describe the shape of the mixing distribution, using flexible forms such as polynomials, splines, and step functions. The gradient of the log-likelihood is easy to calculate, which facilitates estimation. The procedure is illustrated with data on consumers' choice among video streaming services.

1. Introduction

A mixed logit model (e.g., Revelt and Train, 1998) contains two parts: a logit specification of a person's probability of choosing a given alternative, which depends on parameters that enter the person's utility function; and a specification of the distribution – often called the mixing distribution – of these utility-parameters over people. McFadden and Train (2000) have shown that a mixed logit model can, under benign conditions, approximate any choice model to any degree of accuracy. However, this theoretical generality is constrained in practice by the difficulty of specifying and estimating parameter distributions that are sufficiently flexible and yet feasible from a computational perspective. The vast majority of studies have used normal and lognormal distributions, with a few using Johnson's S_b, gamma, and triangular distributions. However, these distributions are limiting, and most researchers will probably agree that: whatever parametric distribution the researcher specifies, he/she quickly becomes dissatisfied with its properties.

The current paper introduces a new way of specifying the distribution of random parameters that is relatively simple numerically and yet allows a high degree of flexibility. For a finite parameter space, the probability of each parameter value is given by a logit function with terms that are defined by the researcher to describe the shape of the distribution. The researcher can specify polynomials, splines, steps, and other functions that have been developed for general approximation. The exponential in the logit numerator assures that the probability is positive, and the summation in the denominator assures that the probabilities sum to one. Sampling from the parameter space is facilitated by the fact that the logit probability takes the same form on subsets as the full set. When used in a mixed logit, the model contains two logits: one for the decision-maker's choice among alternatives, and another for the "selection" of parameters for the decision-maker. The procedure is easy to program and fast computationally.

Procedures for flexible mixing distributions have been previously proposed by Bajari et al. (2007), Fosgerau and Bierlaire

☆ Matlab codes to implement the procedures in this paper are available at http://eml.berkeley.edu/~train/software.html.

E-mail addresses: kenneth.train@nera.com, train@econ.berkeley.edu

http://dx.doi.org/10.1016/j.jocm.2016.07.004
1755-5345/

(2007), Train (2008), Fox et al. (2011), Burda et al. (2008) and Fosgerau and Mabit (2013). The current method is an approximate generalization of the first four of these papers. Burda et al. (2008) obtain flexibility through a convolution of a normal kernel with a skewing function. Fosgerau and Mabit (2013) suggest an approach that approximates a different function than the density of the utility parameters. In particular, random terms from a standardized distribution (such as uniform or standard normal) are transformed as they enter utility, and this transformation is approximated by, e.g., a polynomial.

This movement toward greater flexibility, which the current paper augments, shifts the emphasis for future research from overcoming distributional constraints to developing richer datasets that can more clearly distinguish among the variety of shapes that the parameter distribution might take.

Sections 2 and 3 describe the model form and its estimation. Section 4 gives examples of how to specify variables to describe the mixing distribution. Section 5 extends the procedure in two fairly obvious ways. Section 6 provides an application, and Section 7 concludes.

2. A Logit-mixed logit (LML) model

Consider a situation in which each decision-maker makes only one choice, since the double-use of logits is most apparent in this situation; generalization to multiple choices by each decision-maker is described in the next section. Let the utility that person n obtains from alternative j in choice set J be denoted in the usual way as $U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$ where x_{nj} is a vector of observed attributes, β_n is a corresponding vector of utility coefficients that vary randomly over people, and ε_{nj} is a random term that represents the unobserved component of utility. The unobserved term ε_{nj} is assumed to be distributed iid extreme value. Under this assumption, the probability that person n chooses alternative i, conditional on β_n, is the logit formula:

$$Q_{ni}(\beta_n) = \frac{e^{\beta_n' x_{ni}}}{\sum_{j \in J} e^{\beta_n' x_{nj}}} \tag{1}$$

The researcher does not observe the utility coefficients of each person and knows that the coefficients vary over people. The cumulative distribution function of β_n in the population is $F(\beta)$ which is called the mixing distribution. Let F be discrete with finite support set S. This specification is not restrictive since a continuous distribution can be approximated to any degree of accuracy by a discrete distribution with a sufficiently large and dense S. The probability mass at any $\beta_r \in S$ is expressed as a logit formula:

$$\text{Prob}(\beta_n = \beta_r) \equiv W(\beta_r | \alpha) = \frac{e^{\alpha' z(\beta_r)}}{\sum_{s \in S} e^{\alpha' z(\beta_s)}} \tag{2}$$

where $z(\beta_r)$ is a vector-valued function of β_r and α is a corresponding vector of coefficients. The z variables are chosen to capture the shape of the probability mass function; their specification is discussed in Section 4 below.

The unconditional choice probability is then:

$$\text{Prob}(n \text{ chooses } i) = \sum_{r \in S} W(\beta_r | \alpha) \cdot Q_{ni}(\beta_r) = \sum_{r \in S} \left(\frac{e^{\alpha' z(\beta_r)}}{\sum_{s \in S} e^{\alpha' z(\beta_s)}} \right) \cdot \left(\frac{e^{\beta_r' x_{ni}}}{\sum_{j \in J} e^{\beta_r' x_{nj}}} \right) \tag{3}$$

The model contains a logit formula for the probability that the decision-maker chooses alternative i *and* a logit formula for the probability that the decision-makers has utility coefficients β_r. The researcher's task is to specify the x variables that describe the probability of each alternative and the z variables that describe the probability of each β_r.

The advantage of using a logit model for the mixing distribution is that it allows for easy and flexible specification of relative probabilities. The researcher specifies z variables that describe the shape of the distribution, without needing to be concerned about assuring positivity or summation to one: the exponential in the logit numerator guarantees that the probability at each point is positive, and the sum in the denominator guarantees that probabilities sum to one over points.

The specification is entirely general in the sense that any choice model with any mixing distribution can be approximated to any degree of accuracy by a model of the form of Eq. (3). McFadden's (1975) "mother logit" theorem shows that any model that describes the choice among alternatives can be represented by a logit formula of the form in Eq. (1). An analogous derivation applies for representing the mixing distribution as a logit formula.

Result: For any mixing distribution, there exists a sequence of probability distributions in the form of Eq. (2) that converges weakly (i.e., in distribution) to that mixing distribution. Stated more intuitively, any mixing distribution can be approximated to any degree of accuracy by a logit model of the form given in Eq. (2). Proof: For any distribution F^*, there exists a sequence of discrete distributions, labeled F_m, $m = 1, 2, \ldots$, that converges weakly to F^* (e.g. Chamberlain, 1987).[1] Let the probability mass function associated with distribution F_m be denoted $k_m f_m(\beta)$ at each $\beta \in S_m$, where k_m is the normalizing constant, $f_m(\beta)$ is the kernel, and S_m is the support set. Define $g_m(\beta) = \ln(f_m(\beta))$, which exists for each $\beta \in S_n$; that is, function

[1] Dan McFadden has suggested (personal communication) a simple demonstration: Take m draws from F^* and define $F_m(\beta) = \sum_{i=1}^{m} I(\beta_i \leq \beta)/m$. By the strong law of large numbers, this sequence converges weakly to F^*.