Notes on *Discrete Choice Methods with Simulation* (Train, 2009)

Chapter 6. Mixed logit

- Mixed logit is a highly flexible model that can approximate any random utility model. It obviates the three limitations of the logit model (including GEV) and is not restricted to normal distributions like probit.

6.1. Choice probabilities

- The mixed logit model is defined on the basis of the functional form for its choice probabilities. Any behavioral specification whose derived choice probabilities are the integrals of standard logit probabilities over a density of parameters is called a mixed logit model. That is,

  $$P_{ni} = \int L_{ni}(\beta) f(\beta) \mathrm{d}\beta,$$

  where $L_{ni}(\beta)$ is the logit probabilities evaluated at parameter β:

  $$L_{ni}(\beta) = \frac{\mathrm{e}^{V_{ni}(\beta)}}{\sum_{j=1}^{J} \mathrm{e}^{V_{nj}(\beta)}},$$

  or

  $$P_{ni} = \int \frac{\mathrm{e}^{\beta' x_{ni}}}{\sum_{j=1}^{J} \mathrm{e}^{\beta' x_{nj}}} f(\beta) \mathrm{d}\beta$$

  if utility is linear in β.
- The mixed logit probability is a weighted average of the logit formula evaluated at different values of β, with the weights given by the *mixing distribution* $f(\beta)$.
  - Standard logit is a special case where the mixing distribution is degenerate at fixed parameter b.
  - The mixed distribution can also be a discrete distribution. In this case, the mixed logit becomes the *latent class model*, whose choice probability for any alternative i is a weighted average of different choice probabilities for alternative i of M segments of decision makers:

    $$P_{ni} = \sum_{m=1}^{M} s_m \left(\frac{\mathrm{e}^{b'_m x_{ni}}}{\sum_{j=1}^{J} \mathrm{e}^{b'_m x_{nj}}} \right).$$

- In most applications that have actually been called mixed logit, $f(\beta)$ is specified to be continuous. For example, the density of β can be specified to be normal with mean b and covariance W. Denote the parameters that describe the density of β as θ (e.g., $\theta = \{b, W\}$ in the normal distribution case), the mixed logit probabilities are

  $$P_{ni} = \int L_{ni}(\beta) f(\beta \mid \theta) \mathrm{d}\beta.$$

  Note that these choice probabilities are functions of θ, the parameters to be estimated.
- The ratio of mixed logit probabilities, P_{ni}/P_{nj}, depends on all the data, including attributes of alternatives other than i or j. The percentage change in the probability for alternative i given a

percentage change in the m-th attribute of another alternative j is

$$E_{nix_{nj}^m} = -x_{nj}^m \int \beta^m L_{nj}(\beta) \left[\frac{L_{ni}(\beta)}{P_{ni}}\right] f(\beta)\mathrm{d}\beta,$$

where β^m is the m-th element of β.

6.2. Random coefficients

- The mixed logit probability can be derived from utility-maximizing behavior in several ways that are formally equivalent but provide different interpretations. The most straightforward derivation, and most widely used in recent applications, is based on random coefficients.
- Suppose that the utility of person n from alternative j is specified as

  $$U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj},$$

  where β_n is a vector of coefficients for person n representing that person's tastes and ε_{nj} is a random term that is i.i.d. extreme value. The coefficients vary over decision makers in the population with density $f(\beta \mid \theta)$.
- The choice probability conditional on β_n is

  $$L_{ni}(\beta) = \frac{\mathrm{e}^{\beta_n' x_{ni}}}{\sum_j \mathrm{e}^{\beta_n' x_{nj}}},$$

  therefore, the unconditional choice probability is therefore the integral of $L_{ni}(\beta)$ over all possible variables of β_n:

  $$P_{ni} = \int \frac{\mathrm{e}^{\beta' x_{ni}}}{\sum_j \mathrm{e}^{\beta' x_{nj}}} f(\beta)\mathrm{d}\beta,$$

  which is the mixed logit probability.
- Researchers can impose different assumptions on the mixing distribution:
  - Normal or lognormal distribution, with parameters b and W that are estimated.
  - Triangular or uniform distribution between $b - s$ and $b + s$, where the mean b and spread s are estimated. Such distribution can be used to limit the possible values of coefficients. For example, letting $s = b$ ensures that the coefficients have the same sign for all decision makers.
  - ...
- Some variations in tastes that are related to observed attributes of the decision maker can are captured through specification of the explanatory variables (see 3.3 in Chapter 3). Observed attributes of the decision maker can also enter $f(\beta)$, so that higher-order moments of taste variation can also depend on attributes of the decision maker. For example, $f(\beta)$ can be log-normal with mean and variance depending on decision maker's characteristics.

6.3. Error components

- A mixed logit model can be used without a random-coefficients interpretation, as simply representing error components that create correlations among the utilities for different alternatives. Utility is specified as

  $$U_{nj} = \alpha' x_{nj} + \mu_n' z_{nj} + \varepsilon_{nj},$$

  where x_{nj} and z_{nj} are observed variables, α is a vector of fixed coefficients, μ_n is a vector of random terms with zero mean, and ε_{nj} is i.i.d. extreme value. Therefore, the unobserved portion of the utility is $\eta_{nj} = \mu_n' z_{nj} + \varepsilon_{nj}$.

- Utility is correlated over alternatives:

$$\operatorname{Cov}(\eta_{nj}, \eta_{nj}) = z_{ni}' W z_{nj},$$

  where W is the covariance matrix of μ_n. Various correlation patterns can be obtained by an appropriate choice of variables to enter as error components:
  - An analog to nested logit is obtained by specifying a dummy variable for each nest that equals one for each alternative in the nest and zero for alternatives outside the nest; that is, $\mu_n' z_{nj} = \sum_{k=1}^{K} \mu_{nk} d_{jk}$, where $\mu_{nk} \sim_{\text{i.i.d.}} N(0, \sigma_k)$ and $d_{jk} = 1$ if alternative j is in nest k and 0 otherwise.
  - An analog to overlapping nests is captured with dummies that identify overlapping sets of alternatives.
  - Walker et al. (2007) provide guidance on how to specify these variables appropriately.
- Error-component and random coefficient specifications are formally equivalent, since decomposing the random coefficients into their mean α and deviations μ_n produces the error components. However, the way a researcher thinks about the model affects the specification of the mixed logit:
  - When researchers are interested in the pattern of tastes, a random coefficient model is naturally to be used.
  - When the primary goal is to represent substitution patterns appropriately through the use of error components, the emphasis is placed on specifying variables that can induce correlations over alternatives.

6.4. Approximation to any random utility model

- McFadden & Train (2000) show that any random utility model (RUM) can be approximated to any degree of accuracy by a mixed logit with appropriate choice of variables and mixing distribution.

6.5. Simulation

- The choice probabilities are approximated through simulation for any given value of θ: (1) Draw a value of β from $f(\beta \mid \theta)$, and label it β^r with the superscript $r = 1$ referring to the first draw; (2) Calculate the logit formula $L_{ni}(\beta^r)$ with this draw; (3) Repeat steps 1 and 2 many times, and average the results. This average is the simulated probability:

$$\check{P}_{ni} = \frac{1}{R} \sum_{r=1}^{R} L_{ni}(\beta^r),$$

  where R is the number of draws. $\check{P}_{ni}$ is an unbiased estimator of P_{ni} by construction.
- The simulated probabilities are inserted into the log-likelihood function to give a simulated log likelihood:

$$\mathrm{SLL} = \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \ln \check{P}_{nj},$$

  where $d_{nj} = 1$ if n chose j and zero otherwise. The maximum simulated likelihood estimator (MSLE) is the value of θ that maximizes SLL.
- The mixed logit simulator can be seen as a logit-smoothed AR simulator of any RUM: draws of the random terms are taken, utility is calculated for these draws, the calculated utilities are inserted into the logit formula, and the results are averaged.

6.6. Panel data

- Utility from alternative j in choice situation t by person n is $U_{njt} = \beta_n x_{njt} + \varepsilon_{njt}$ with ε_{njt} being i.i.d. extreme value over time, people, and alternatives.
- To begin with, consider the simplest specification that treats the coefficients that enter utility as varying over people but being constant over choice situations for each person. Define a sequence of alternatives, one for each time period, $\mathbf{i} = \{i_1, \cdots, i_T\}$. Conditional on β, the probability that the person makes this sequence of choices is the product of logit formulas:

  $$\mathbf{L}_{n\mathbf{i}}\,(\beta) = \prod_{t=1}^{T} \left[\frac{\mathrm{e}^{\beta_n' x_{ni_t t}}}{\sum_j \mathrm{e}^{\beta_n' x_{njt}}} \right].$$

  The unconditional probability is

  $$P_{n\mathbf{i}} = \int \mathbf{L}_{n\mathbf{i}}\,(\beta) f\,(\beta) \mathrm{d}\beta.$$

  Accordingly, the simulation procedure is the same as that in the cross-sectional setting: (1) Draw a value of β from $f(\beta \mid \theta)$, and label it β^r with the superscript $r = 1$ referring to the first draw; (2) Calculate the product of logit formulas $\mathbf{L}_{n\mathbf{i}}(\beta^r)$ with this draw; (3) Repeat steps 1 and 2 many times, and average the results.
- Past and future exogenous variables can be added to the utility in a given period to represent lagged response and anticipatory behavior. The inclusion of these variables won't change the estimation procedure because conditional on β_n, the remaining random terms are independent over time. In this regard, mixed logit is more convenient than probit for representing state dependence.
- The coefficients associated with each person can be specified to vary over time in a variety of ways. For example, each person's tastes might be serially correlated over choice situations:

  $$\begin{aligned} U_{njt} &= \beta_{nt} x_{njt} + \varepsilon_{njt}, \\ \beta_{nt} &= b + \tilde{\beta}_{nt}, \\ \tilde{\beta}_{nt} &= \rho \tilde{\beta}_{n,t-1} + \mu_{nt}, \end{aligned}$$

  where b is fixed and μ_{nt} is i.i.d. over n and t (see the textbook for an example and its simulation procedure).