Simulation-Based Estimation with Applications in S-PLUS to Probabilistic Discrete Choice Models and Continuous-Time Financial Models

Eric Zivot

Department of Economics, University of Washington
and Insightful Corporation

Research funded by NSF SBIR Phase II
DMI-0132076

Note: Original PI: Dr. Jiahui Wang, Insightful Corporation

Slides and scripts available at

`http://faculty.washington.edu/ezivot`

1 Overview of Simulation-Based Econometric Estimation

Problem: Many estimation problems have objective functions that involve high-dimensional integrals that cannot be numerically evaluated using conventional methods

Solution 1: Approximate high dimensional integrals using simulation techniques

- Maximum simulated likelihood (MSL) estimation
- Method of simulated moments (MSM) estimation

Solution 2: Simulate from structural model and match moments to fitted auxiliary model

- Indirect inference
- Efficient method of moments (EMM) estimation

Goal of Grant: Create user-friendly software components for implementing simulation-based estimation techniques

References:

1. K. Train (2003). *Discrete Choice Models with Simulation*, Cambridge University Press
2. K. Train (2001). "A Distance Learning Course on Discrete Choice Models with Simulation," available at

   `http://elsa.berkeley.edu/~train/videos/vidindex.html`

3. Gourieroux, C. and A. Monfort (1997). *Simulation Based Econometric Methods*, Cambridge University Press.

4. Gallant, R. and G. Tauchen (1995). "Which Moments to Match," *Econometric Theory*.

5. Gallant, R. and G. Tauchen (2001). "SNP: A Program of Nonparametric Time Series Analysis, Version 8.9, User's Guide," available at

`ftp.econ.econ.duke.edu`

6. Gallant, R. and G. Tauchen (2002). "EMM: A Program for Efficient Method of Moments, Version 1.6, User's Guide," available at

`ftp.econ.econ.duke.edu`

2 Probabilistic Discrete Choice Models

Choice Set: $j = 1, \ldots, J$ discrete alternatives

- Mutually exclusive alternatives
- Choices are exhaustive
- Finite number of alternatives
- Order and ranking of choices may matter

Examples

- Transportation mode choice

- Political party candidate choice

- Bond rating choice

2.1 McFadden's Random Utility Model

- Decision maker n faces J alternatives from choice set

- $U_{nj} =$ unobserved utility for person for choice j

- Person chooses alternative with highest utility

$$\text{Choose alternative } i \text{ iff } U_{ni} > U_{nj} \ \forall \ j \neq i$$

Model Construction

- $\mathbf{x}_{nj}$ = observed attributes of alternatives faced by person n

- $\mathbf{s}_n$ = observed attributes of decision maker n

- $\boldsymbol{\varepsilon}_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nJ})'$ = unobserved attributes of alternatives

$$\boldsymbol{\varepsilon}_n \text{ has pdf } f(\varepsilon_n|\boldsymbol{\theta})$$

- $V_{nj} = V(\mathbf{x}_{nj}, \mathbf{s}_n) = \mathbf{x}'_{nj}\boldsymbol{\beta} + \mathbf{s}'_n\boldsymbol{\delta}_j$ = observed utility

- $U_{nj} = V_{nj} + \varepsilon_{nj}$ = unobserved random utility

Choice Probability

$$\begin{aligned}
P_{ni} &= \Pr(U_{ni} > U_{nj} \ \forall \ j \neq i) \\
&= \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \ \forall \ j \neq i) \\
&= \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \ \forall \ j \neq i) \\
&= \int \cdots \int I_{[\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \ \forall \ j \neq i]} f(\boldsymbol{\varepsilon}_n|\boldsymbol{\theta}) d\boldsymbol{\varepsilon}_n
\end{aligned}$$

Remarks:

1. Choice probabilities require evaluation of multi-dimensional integral

2. Only differences in utility determine choice probability: utility must be normalized for level and scale

3. Only effects that differ across alternatives can be identified

$$
\begin{aligned}
V_{ni} - V_{nj} &= (\mathbf{x}_{ni} - \mathbf{x}_{nj})'\boldsymbol{\beta} + \mathbf{s}_n'(\boldsymbol{\delta}_i - \boldsymbol{\delta}_j) \\
\text{set } \boldsymbol{\delta}_1 &= \mathbf{0} \text{ for identification}
\end{aligned}
$$

4. Distribution of ε_n determines choice model

Logit: $\varepsilon_{nj} \sim iid$ extreme value.

- Unobserved factors are uncorrelated and homoskedastic across alternatives.

- P_{ni} has closed form solution

GEV: $\varepsilon_n \sim$ generalized extreme value.

- Allows limited correlation across alternatives.

- Logit is special case. Nested logit is common model.

- P_{ni} has closed form solution.

Probit: $\varepsilon_n \sim N(0, \Omega)$.

- Allows general correlation and heteroskedasticity across alternatives.

- Easily modified to handle panel data.

- P_{ni} does not have closed form solution.

- Restrictions on $\boldsymbol{\Omega}$ are required to normalize utility for level and scale. Identification can be tricky.

Mixed-logit: $\varepsilon_{nj} = \boldsymbol{\eta}_n' \mathbf{z}_{nj} + \nu_{nj}$, $\boldsymbol{\eta}_n \sim f(\boldsymbol{\eta}_n | \boldsymbol{\theta})$ and $v_{nj} \sim iid$ extreme value.

- Allows for general correlation and heteroskedasticity across alternatives.

- Can approximate any random utility discrete choice model.

- $f(\boldsymbol{\eta}_n | \boldsymbol{\theta})$ can be any distribution.

- Easily modified to handle panel data.

- Identification can be tricky.

3 Choice Probabilities and Integration

$$
\begin{aligned}
y &= \text{decision outcome from choice situation} \\
x &= \text{observed factors} \\
\varepsilon &= \text{unobserved random factors with pdf } f(\varepsilon) \\
y &= h(x, \varepsilon) = \text{ behavioral model} \\
\Pr(y|x) &= \Pr(\varepsilon \text{ s.t. } y = h(x, \varepsilon)) = \text{ choice probability}
\end{aligned}
$$

Let $I_{[A]} = 1$ when event A occurs and 0 otherwise. Then

$$
\begin{aligned}
\Pr(y|x) &= \Pr(I_{[y=h(x,\varepsilon)]} = 1) \\
&= \int I_{[y=h(x,\varepsilon)]} f(\varepsilon) d\varepsilon
\end{aligned}
$$

3.1 Complete Close-Form Expression

Logit binary decision model:

$$\begin{aligned} y &= 1 \text{ if action is taken} \\ &= 0, \text{ otherwise} \end{aligned}$$

Net utility

$$\begin{aligned} U &= U_1 - U_0 = \mathbf{x}'\boldsymbol{\beta} + \varepsilon \\ f(\varepsilon) &= e^{-\varepsilon}/(1+e^{-\varepsilon})^2 \\ F(\varepsilon) &= 1/(1+e^{-\varepsilon}) \end{aligned}$$

Choice probability

$$\begin{aligned} P_1 &= \Pr(y=1|\mathbf{x}) = \Pr(U > 0) \\ &= \Pr(\varepsilon > -\mathbf{x}'\boldsymbol{\beta}) \\ &= \int I_{[\varepsilon > -\mathbf{x}'\boldsymbol{\beta}]} f(\varepsilon) d\varepsilon \\ &= \int_{-\mathbf{x}'\boldsymbol{\beta}}^{\infty} f(\varepsilon) d\varepsilon \\ &= \frac{e^{\mathbf{x}'\boldsymbol{\beta}}}{1+e^{\mathbf{x}'\boldsymbol{\beta}}} \end{aligned}$$

3.2 Complete Simulation

Choice probability is an expectation:

$$\begin{aligned}\text{Pr}(y|x) &= \text{Pr}(I_{[y=h(x,\varepsilon)]} = 1) \\ &= \int I_{[y=h(x,\varepsilon)]} f(\varepsilon) d\varepsilon \\ &= E[I_{[y=h(x,\varepsilon)]}]\end{aligned}$$

Approximate $E[I_{[y=h(x,\varepsilon)]}]$ using Monte Carlo Integration

- Draw ε^r from $f(\varepsilon)$

- If $y = h(x, \varepsilon^r) = 1$, set $I^r_{[y=h(x,\varepsilon^r)]} = 1$; otherwise set set $I^r_{[y=h(x,\varepsilon^r)]} = 1$

- Repeat process R times and compute Monte Carlo average

$$\widehat{\text{Pr}}(y|x) = \frac{1}{R} \sum_{r=1}^{R} I^r_{[y=h(x,\varepsilon^r)]}$$

4 Logit Models

Random Utility Model

$$U_{nj} = V_{nj} + \varepsilon_{nj} = \mathbf{x}'_{nj}\boldsymbol{\beta} + \mathbf{s}'_n\boldsymbol{\delta}_j + \varepsilon_{nj}$$
$$\varepsilon_{nj} \sim iid \text{ extreme value}$$
$$f(\varepsilon_{nj}) = \exp(-\varepsilon_{nj})\exp(-\exp(-\varepsilon_{nj}))$$
$$F(\varepsilon_{nj}) = \exp(-\exp(-\varepsilon_{nj}))$$

Properties:

- Form of density normalizes utility for level and scale
- Choice probability has closed form solution

$$\begin{aligned} P_{ni} &= \frac{\exp(V_{ni})}{\sum_{j=1}^{J}\exp(V_{nj})},\ i = 1, \ldots, J \\ P_{ni} &\geq 0,\ \sum_{i=1}^{J} P_{ni} = 1. \end{aligned}$$

- Marginal effects

$$\frac{\partial V_{nj}}{\partial \mathbf{x}_{nj}} = \boldsymbol{\beta}, \;\; \frac{\partial V_{nj}}{\partial \mathbf{s}_n} = \boldsymbol{\delta}_j$$

  Let z_{ni} be an element of $\mathbf{x}_{ni}$. Then

$$\begin{aligned} \frac{\partial P_{ni}}{\partial z_{ni}} &= \beta_z P_{ni}(1 - P_{ni}) \\ \frac{\partial P_{ni}}{\partial z_{nj}} &= -\beta_z P_{ni} P_{nj} \end{aligned}$$

- Elasticities

$$\begin{aligned} E_{i,z_{ni}} &= \frac{\partial P_{ni}}{\partial z_{ni}} \cdot \frac{z_{ni}}{P_{ni}} = \beta_z z_{ni}(1 - P_{ni}) \\ E_{i,z_{nj}} &= \frac{\partial P_{ni}}{\partial z_{nj}} \cdot \frac{z_{nj}}{P_{ni}} = -\beta_z z_{nj} P_{nj} \end{aligned}$$

  Note: cross elasticity only depends on $j \Rightarrow$ proportionate substitution

- Odds ratios/IIA property

$$\frac{P_{ni}}{P_{nk}} = \frac{\exp(V_{ni})}{\exp(V_{nk})} = \exp(V_{ni} - V_{nk})$$

Note: ratio only depends on alternatives i and $k \Rightarrow$ relative odds of choosing alternative i over alternative k is the same no matter what other alternatives are available. Hence, the ratio is independent from alternatives irrelevant to i and k.

- IIA: Red bus/Blue bus example

4.1 Maximum Likelihood Estimation

Define $y_{ni} = 1$ if individual n selects alternative i, and $y_{ni} = 0$ otherwise.

$$\begin{aligned}
&\Pr(\text{individual } n \text{ selects alternative that is observed}) \\
&= \prod_{i=1}^{J} (P_{ni})^{y_{ni}}
\end{aligned}$$

$$P_{ni} = \frac{e^{\mathbf{x}'_{ni}\boldsymbol{\beta}+\mathbf{s}'_n\boldsymbol{\delta}_i}}{\sum_j e^{\mathbf{x}'_{nj}\boldsymbol{\beta}+\mathbf{s}'_n\boldsymbol{\delta}_j}}$$

Given a random sample of size N, the joint probability that all individuals select the alternatives that were observed is

$$\prod_{n=1}^{N} \prod_{i=1}^{J} (P_{ni})^{y_{ni}}$$

which gives the log-likelihood function

$$\begin{aligned}
\ln L(\boldsymbol{\theta}) &= \sum_{i=1}^{N} \sum_{i=1}^{J} y_{ni} \ln(P_{ni}) \\
\boldsymbol{\theta} &= (\boldsymbol{\beta}', \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_J)'
\end{aligned}$$

The maximum likelihood estimator is defined as

$$\hat{\boldsymbol{\theta}}_{mle} = \arg\max_{\theta} \; \ln L(\boldsymbol{\theta})$$

- Logit log-likelihood is globally concave

- $\sqrt{N}(\hat{\boldsymbol{\theta}}_{mle} - \boldsymbol{\theta}) \rightarrow N(0, \mathbf{I}(\boldsymbol{\theta})^{-1}), \mathbf{I}(\boldsymbol{\theta}) = -E[\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}]$

4.2 Example: Commuter Mode Choice

- Data on the mode choice of 453 commuters: (1) car alone, (2) carpool, (3) bus, and (4) rail.

- Observe cost and time on each mode and the chosen mode

- Random utility model with $J = 4$ alternatives and $N = 453$ commuters

$$\begin{aligned} U_{ni} &= V_{ni} + \varepsilon_{ni}, \\ V_{ni} &= \mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{s}'_{n}\boldsymbol{\delta}_{i} \\ P_{ni} &= \Pr(U_{ni} > U_{nj},\ \forall\ j \neq i) \end{aligned}$$

  where $\mathbf{x}_{ni}$ denotes the 2×1 vector of cost and time variables that vary over individuals and alternatives, and $\mathbf{s}_n$ represents a 3×1 vector of alternative specific constants.

5 Probit Models

Random utility model

$$
\begin{aligned}
U_{nj} &= V_{nj} + \varepsilon_{nj} = \mathbf{x}'_{nj}\boldsymbol{\beta} + \varepsilon_{nj} \\
\boldsymbol{\varepsilon}_n &= (\varepsilon_{n1}, \ldots, \varepsilon_{nJ}) \sim N(0, \boldsymbol{\Omega}) \\
f(\boldsymbol{\varepsilon}_n|\boldsymbol{\theta}) &= \phi(\boldsymbol{\varepsilon}_n) = (2\pi)^{-J/2}|\boldsymbol{\Omega}|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon}'_n\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_n\right\} \\
F(\boldsymbol{\varepsilon}_n) &= \Phi(\boldsymbol{\varepsilon}_n) = \int_{-\infty}^{\varepsilon_{n1}} \cdots \int_{-\infty}^{\varepsilon_{nJ}} \phi(\mathbf{z}) d\mathbf{z}
\end{aligned}
$$

In terms of $\boldsymbol{\varepsilon}_n$, choice Probability is $J-$dimensional integral

$$
\begin{aligned}
P_{ni} &= \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \ \forall\ j \neq i) \\
&= \int \cdots \int I_{[\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \ \forall\ j \neq i]} \phi(\boldsymbol{\varepsilon}_n) d\boldsymbol{\varepsilon}_n
\end{aligned}
$$

Result: P_{ni} depends on distribution of $\varepsilon_{nj} - \varepsilon_{ni}\ \forall\ j \neq i \Rightarrow$ Dimension may be reduced to $J-1$. Define

$$
\begin{aligned}
\tilde{U}_{nji} &= U_{nj} - U_{ni} \\
\tilde{V}_{nji} &= V_{nj} - V_{ni} \\
\tilde{\varepsilon}_{nji} &= \varepsilon_{nj} - \varepsilon_{ni} \\
\tilde{\boldsymbol{\varepsilon}}_{ni} &= (\tilde{\varepsilon}_{n1i}, \ldots, \tilde{\varepsilon}_{nJi})' \text{ excluding } \tilde{\varepsilon}_{nii}
\end{aligned}
$$

Then

$$
\begin{aligned}
\tilde{\boldsymbol{\varepsilon}}_{ni} &\sim N(\mathbf{0}, \tilde{\boldsymbol{\Omega}}_i) \\
\tilde{\boldsymbol{\Omega}}_i &= \mathbf{M}_i \boldsymbol{\Omega} \mathbf{M}_i'
\end{aligned}
$$

where $\mathbf{M}_i$ is the identity matrix of dimension $J-1$ with a $(J-1)\times 1$ column of -1 values inserted at the ith column.

In terms of $\tilde{\boldsymbol{\varepsilon}}_{ni}$, P_{ni} is $J-1$ dimensional integral

$$P_{ni} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} I(\tilde{\varepsilon}_{nji} < -\tilde{V}_{nji}, \forall\ j \neq i)\phi(\tilde{\boldsymbol{\varepsilon}}_{ni})d\tilde{\boldsymbol{\varepsilon}}_{ni}$$

Remarks:

- For $J-1 > 3$ cannot accurately evaluate integral using numerical integration
- Integral may be accurately approximated using simulation
- Not all parameters of model are identified without further restrictions

5.1 Identification

1. Take utility differences with respect to the first alternative and compute

$$\tilde{\mathbf{\Omega}}_1 = \mathbf{M}_1 \mathbf{\Omega} \mathbf{M}_1'$$

2. Normalize the level and scale of utility by setting the $(1,1)$ element of $\tilde{\mathbf{\Omega}}_1$ to unity. This produces the $(J-1)\times(J-1)$ restricted error difference covariance matrix

$$\tilde{\mathbf{\Omega}}_1^* = \begin{pmatrix} 1 & \omega_{12}^* & \cdots & \omega_{1,J-1}^* \\ & \omega_{22}^* & \cdots & \omega_{2,J-1}^* \\ & & \ddots & \vdots \\ & & & \omega_{J-1,J-1}^* \end{pmatrix}$$

which has $J(J-1)/2-1$ unique elements.

3. Construct the normalized $J \times J$ covariance matrix

for the errors using

$$\begin{aligned}
\boldsymbol{\Omega}^* &= \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \omega^*_{12} & \cdots & \omega^*_{1,J-1} \\ 0 & \omega^*_{12} & \omega^*_{22} & \cdots & \omega^*_{2,J-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \omega^*_{1,J-1} & \omega^*_{2,J-1} & \cdots & \omega^*_{J-1,J-1} \end{pmatrix} \\
&= \begin{pmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \tilde{\boldsymbol{\Omega}}^*_1 \end{pmatrix}
\end{aligned}$$

4. Compute Choleski factorization

$$\tilde{\boldsymbol{\Omega}}^*_1 = \mathbf{C}\mathbf{C}'$$

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ c_{21} & c_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ c_{J-1,1} & c_{J-1,2} & \cdots & c_{J-1,J-1} \end{pmatrix}$$

where $\mathbf{C}$ has $J(J-1)/2-1$ elements

5. $\boldsymbol{\theta} = (\boldsymbol{\beta}', c_{21}, \ldots, c_{J-1,J-1})'$ represents the $K + J(J-1)2-1$ parameters to be estimated in the normalized probit model.

5.2 Accept-Reject Simulation of P_{ni}

$$\begin{aligned} P_{ni} &= \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \ \forall \ j \neq i) \\ &= \int \cdots \int I_{[\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \ \forall \ j \neq i]} \phi(\boldsymbol{\varepsilon}_n) d\boldsymbol{\varepsilon}_n \end{aligned}$$

- Draw $\boldsymbol{\varepsilon}_n^r$ from $\phi(\boldsymbol{\varepsilon}_n)$
- Calculate $U_{ni}^r = \mathbf{x}_{ni}'\boldsymbol{\beta} + \varepsilon_{ni}^r$ for $i = 1, \ldots, J$
- If $U_{ni}^r > U_{nj}^r$ for all $j \neq i$ set $I^r = 1$; otherwise set $I^r = 0$
- Repeat process R time and compute Monte Carlo average

$$\check{P}_{ni} = \frac{1}{R} \sum_{r=1}^{R} I^r$$

Problems

- $\breve{P}_{ni}$ may be zero

- $\breve{P}_{ni}$ is not a smooth function of $\boldsymbol{\beta}$

5.3 Truncated Normal Distribution

Let $z \sim N(0,1)$, let $\phi(z)$ denote the pdf and $\Phi(z)$ denote the CDF. Then for any upper truncation point v,

$$\begin{aligned} \tau &= z|z \leq v \sim TN(v) \\ f(z) &= \phi(z)/\Phi(v) \text{ for } z \leq v \\ &= 0, \text{ otherwise} \end{aligned}$$

5.3.1 Simulating from a Truncated Normal Density

Goal: Generate random draw from $TN(v)$

- Draw $u \sim U[0,1]$
- Compute $\bar{u} = \Phi(v)u$
- Compute $\tau = \Phi^{-1}(\bar{u})$

5.4 Smooth Simulation of P_{ni}

Let $J - 1 = M$. Then P_{ni} is the $M-$dimensional negative orthant probability:

$$
\begin{aligned}
P_{ni} &= \Pr(\mathbf{x} < \mathbf{v}) \\
& \quad \mathbf{x} = \tilde{\boldsymbol{\varepsilon}}_{ni} \sim N(0, \tilde{\boldsymbol{\Omega}}_i^*) \\
v &= (-\tilde{V}_{n1i}, -\tilde{V}_{n2i}, \ldots, -\tilde{V}_{nJi})' \\
&= (v_1, v_2, \ldots, v_M)'
\end{aligned}
$$

Define

$$
\begin{aligned}
\mathbf{x} &= \mathbf{Cz}, \ \tilde{\boldsymbol{\Omega}}_i^* = \mathbf{CC}' \\
\mathbf{C} &= \begin{pmatrix} c_{11} & 0 & \cdots & 0 \\ c_{21} & c_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ c_{M1} & c_{M2} & \cdots & c_{MM} \end{pmatrix} \\
& \quad \mathbf{z} \sim N(0, \mathbf{I}_M)
\end{aligned}
$$

Then

$$
\begin{gathered}
\Pr(\mathbf{x} \leq \mathbf{v}) = \Pr(\mathbf{Cz} \leq \mathbf{v}) \\
= \Pr(c_{11}z_1 \leq v_1, c_{21}z_1 + c_{22}z_2 \leq v_2, \ldots, \\
c_{31}z_1 + c_{32}z_2 + \cdots + c_{MM}z_M \leq v_M)
\end{gathered}
$$

Define

$$
\begin{aligned}
D &= \{z \in \mathbb{R} : z_1 \leq \frac{v_1}{c_{11}}, z_2 \leq \frac{v_2 - c_{21} z_1}{c_{22}}, \ldots, \\
z_M &\leq \frac{v_M - c_{M1} z_1 - \cdots - c_{M-1,M-1} z_{M-1}}{c_{MM}}\}
\end{aligned}
$$

Then

$$P_{ni} = \int_D \phi(z_1) \cdots \phi(z_M) dz_1 \cdots dz_M$$

Over D, $z_i \sim TN(\frac{v_i - c_{i1} z_1 - \cdots c_{i,i-1} z_{i-1}}{c_{ii}})$ and may be simulated using the recursive algorithm:

1. Draw $u_1, \ldots, u_{M-1}$ iid $U[0,1]$

2. Compute

$$\tau_1 = \Phi^{-1}\left(u_1 \Phi\left(\frac{v_1}{c_{11}}\right)\right)$$

$$\tau_2 = \Phi^{-1}\left(u_2 \Phi\left(\frac{v_2 - c_{21}\tau_1}{c_{22}}\right)\right)$$

$$\vdots$$

$$\tau_{M-1} = \Phi^{-1}(u_{M-1}$$

$$\times \Phi\left(\frac{v_{M-1} - c_{M-1,1}\tau_1 - \cdots - c_{M-1,M-1}\tau_{M-1}}{c_{MM}}\right))$$

Result:

$$P_{ni} = \Phi\left(\frac{v_1}{c_{11}}\right) \int_{[0,1]^{M-1}} \Phi\left(\frac{v_2 - c_{21}\tau_1}{c_{11}}\right) \cdots$$

$$\cdots \Phi\left(\frac{v_M - c_{M,1}\tau_1 - \cdots - c_{M,M-1}\tau_{M-1}}{c_{11}}\right) du_1 \cdots du_{M-1}$$

may be estimated using Monte Carlo integration.

5.4.1 GHK Algorithm

References:

1. Genz, A. (1992). "Numerical Computation of the Multivariate Normal Probabilities," *Journal of Computational and Graphical Statistics*.

2. Geweke, J. (1989). "Efficient Simulation from the Multivariate Normal Distribution Subject to Linear Inequality Constraints and the Evaluation of Constraint Probabilities," mimeo, Duke University.

3. Hajivassiliou, V. (1990). "Smooth Simulation Estimation of Panel Data LDV Models," mimeo, Yale University.

4. Keane, M. (1990). “A Computationally Practical Simulation Estimator for Panel Data, with Applications to Estimating Temporal Dependence in Employment and Wages,” mimeo, University of Minnesota.

Algorithm

1. Compute Pr $\left(z_1 < \frac{v_1}{c_{11}}\right) = \Phi(\frac{v_1}{c_{11}})$

2. Generate $\tau_1^r, \ldots, \tau_{M-1}^r$ recursively as described above

3. Compute

$$\Phi\left(\frac{v_2 - c_{21}\tau_1^r}{c_{22}}\right), \ldots,$$
$$\Phi\left(\frac{v_M - c_{M1}\tau_1^r - \cdots - c_{M,M-1}\tau_{M-1}^r}{c_{33}}\right)$$

4. Repeat 2-3 R times and compute MC estimate

$$\breve{P}_{ni} = \Phi\left(\frac{v_1}{c_{11}}\right)\frac{1}{R}\sum_{r=1}^{R}\Phi\left(\frac{v_2 - c_{21}\tau_1^r}{c_{22}}\right)$$
$$\times\Phi\left(\frac{v_3 - c_{31}\tau_1^r - c_{32}\tau_2^r}{c_{33}}\right)$$
$$\cdots \times \Phi\left(\frac{v_M - c_{M,1}\tau_1 - \cdots - c_{M,M-1}\tau_{M-1}}{c_{MM}}\right)$$

Remarks

- $\breve{P}_{ni}$ is an unbiased simulator: $E\left[\breve{P}_{ni}\right] = P_{ni}$
- $\breve{P}_{ni}$ is continuous and differential function of $\boldsymbol{\theta}$
- By law of large numbers, as $R \to \infty$

$$\breve{P}_{ni} \xrightarrow{p} P_{ni}$$
$$SD(\breve{P}_{ni}) = O(\sqrt{R})$$

5.4.2 Accuracy of GHK Simulator

Evaluate $p = \Pr(\mathbf{x} < \mathbf{v})$, $\mathbf{x} \sim N(0, \boldsymbol{\Omega})$, 100 times using $R = 100$ random draws and $R = 50$ draws using anti-thetic variates.

$$
\begin{aligned}
\mathbf{v}_1 &= \begin{pmatrix} -1.0 \\ -0.75 \\ -0.5 \\ -0.2 \end{pmatrix}, \boldsymbol{\Omega}_1 = \begin{pmatrix} 1.0 & & & \\ 0.2 & 1.0 & & \\ 0.3 & 0.4 & 1.0 & \\ 0.1 & 0.3 & 0.5 & 1.0 \end{pmatrix} \\
\mathbf{v}_2 &= \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 0.0 \end{pmatrix}, \boldsymbol{\Omega}_2 = \begin{pmatrix} 1.0 & & & \\ 0.2 & 1.0 & & \\ 0.2 & 0.4 & 1.0 & \\ 0.2 & 0.4 & 0.6 & 1.0 \end{pmatrix} \\
\mathbf{v}_2 &= \begin{pmatrix} 1.0 \\ 1.0 \\ 1.0 \\ 1.0 \end{pmatrix}, \boldsymbol{\Omega}_2 = \begin{pmatrix} 1.0 & & & \\ 0.9 & 1.0 & & \\ 0.0 & 0.0 & 1.0 & \\ 0.0 & 0.0 & 0.95 & 1.0 \end{pmatrix}
\end{aligned}
$$

Example 1	$R = 100$ No AV	$R = 50$ AV
p	0.0240	0.0240
Mean	0.0240	0.0240
$\lvert$Bias$\rvert$	0.0001	0.0000
Std. dev.	0.0008	0.0003

Example 2	$R = 100$ No AV	$R = 50$ AV
p	0.1497	0.1497
Mean	0.1495	0.1498
$\lvert$Bias$\rvert$	0.0002	0.0001
Std. dev.	0.0048	0.0014

Example 3	$R = 100$ No AV	$R = 50$ AV
p	0.6473	0.1497
Mean	0.6471	0.1498
$\lvert$Bias$\rvert$	0.0002	0.0015
Std. dev.	0.0090	0.0069

5.5 Maximum Simulated Likelihood Estimation

Log-likelihood function

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{i=1}^{J} y_{ni} \ln(P_{ni}(\boldsymbol{\theta}))$$

$$\boldsymbol{\theta} = (\boldsymbol{\beta}', c_{21}, \ldots, c_{J-1,J-1})'$$

$$P_{ni} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} I(\tilde{\varepsilon}_{nji} < -\tilde{V}_{nji}, \forall\ j \neq i)\phi(\tilde{\boldsymbol{\varepsilon}}_{ni})d\tilde{\boldsymbol{\varepsilon}}_{ni}$$

$$\tilde{\boldsymbol{\varepsilon}}_{ni} \sim N(0, \tilde{\boldsymbol{\Omega}}_i^*), \tilde{\boldsymbol{\Omega}}_i^* = \mathbf{M}_i \boldsymbol{\Omega}^* \mathbf{M}_i'$$

- $\ln L(\boldsymbol{\theta})$ cannot be analytically evaluated, $\hat{\boldsymbol{\theta}}_{mle} = \arg\max_\theta \ln L(\boldsymbol{\theta})$ is not feasible.

- Given data and fixed $\boldsymbol{\theta}$,for each n and chosen i, $P_{ni}(\boldsymbol{\theta})$ may be approximated using R replications of the GHK simulator giving $\check{P}_{ni}(\boldsymbol{\theta})$.
  - There are $R \cdot N \cdot (J-2)$ random numbers.

- Simulated log-likelihood function

$$\ln \breve{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{i=1}^{J} y_{ni} \ln(\breve{P}_{ni}(\boldsymbol{\theta}))$$

- $\breve{\boldsymbol{\theta}}_{msle} = \arg\max_{\theta} \ln \breve{L}(\boldsymbol{\theta})$ is called the maximum simulated likelihood estimator.

5.5.1 Properties of SMLE

- $\ln \breve{L}(\boldsymbol{\theta})$ is a biased estimate of $\ln L(\boldsymbol{\theta})$
  - For fixed R, $\breve{\boldsymbol{\theta}}_{msle}$ is not consistent

- If $R \to \infty$ slower than $\sqrt{N}$ then $\breve{\boldsymbol{\theta}}_{msle}$ is consistent but not asymptotically normal

- If $R \to \infty$ faster than $\sqrt{N}$ then $\breve{\boldsymbol{\theta}}_{msle}$ is consistent, asymptotically normal and equivalent to mle.

- In optimization, must use the same random numbers to compute $\breve{P}_{ni}(\boldsymbol{\theta})$ for all values of $\boldsymbol{\theta}$ to prevent chatter. This is a stochastic equicontinuity condition required for consistency.

5.6 Example

- Data on the mode choice of 453 commuters: (1) car alone, (2) carpool, (3) bus, and (4) rail.

- Observe cost and time on each mode and the chosen mode

- Random utility model with $J = 4$ alternatives and $N = 453$ commuters

$$
\begin{aligned}
U_{ni} &= V_{ni} + \varepsilon_{ni}, \\
V_{ni} &= \mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{s}'_{n}\boldsymbol{\delta}_i \\
P_{ni} &= \Pr(U_{ni} > U_{nj}, \ \forall\ j \neq i)
\end{aligned}
$$

  where $\mathbf{x}_{ni}$ denotes the 2×1 vector of cost and time variables that vary over individuals and alternatives, and $\mathbf{s}_n$ represents a 3×1 vector of alternative specific constants.

Normalized covariance for uncorrelated and heteroskedastic errors

Structural covariance

$$\mathbf{\Omega} = \mathbf{I}_4$$

Error difference covariance

$$\begin{aligned}\tilde{\mathbf{\Omega}}_1 &= \mathbf{M}_1 \mathbf{I}_4 \mathbf{M}_1' \\ &= \mathbf{M}_1 \mathbf{M}_1' \\ &= \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}\end{aligned}$$

Normalized error difference covariance

$$\tilde{\mathbf{\Omega}}_1^* = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}$$

6 Mixed Logit

The *mixed-logit model* is a random utility model

$$
\begin{aligned}
U_{ni} &= V_{ni}(\boldsymbol{\beta}) + \varepsilon_{ni},\ i = 1, \ldots, J; n = 1, \ldots, N \\
&\quad \varepsilon_{ni} \sim iid \text{ extreme value} \\
&\quad \boldsymbol{\beta} \text{ has pdf } f(\boldsymbol{\beta}|\boldsymbol{\theta})
\end{aligned}
$$

For fixed $\boldsymbol{\theta}$, P_{ni} is

$$P_{ni}(\boldsymbol{\theta}) = \int \cdots \int L_{ni}(\boldsymbol{\beta}) f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta}$$

where

$$L_{ni}(\boldsymbol{\beta}) = \frac{\exp(V_{ni}(\boldsymbol{\beta}))}{\sum_{j=1}^{J} \exp(V_{nj}(\boldsymbol{\beta}))}$$

- Mixed-logit is a mixture of logits with $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ used as the mixing distribution.

- If $f(\boldsymbol{\beta}|\boldsymbol{\theta}) = 1$ for $\boldsymbol{\beta} = \mathrm{b}$ (and zero otherwise) then the mixed logit choice probability reduces to the regular logit choice probability.

6.1 Random Coefficients Interpretation

Idea: "tastes" associated with attributes vary across individuals $\Rightarrow \beta$ varies with n :

$$U_{ni} = \mathbf{x}'_{ni}\boldsymbol{\beta}_n + \varepsilon_{ni}$$
$$\boldsymbol{\beta}_n \text{ has pdf } f(\boldsymbol{\beta}_n|\boldsymbol{\theta})$$
$$\varepsilon_{ni} \ iid \text{ extreme value}$$

If $\boldsymbol{\beta}_n$ were known, then the conditional choice probability is

$$P_{ni}|\boldsymbol{\beta}_n = \frac{\exp(\mathbf{x}'_{ni}\boldsymbol{\beta}_n)}{\sum_{j=1}^{J}\exp(\mathbf{x}'_{nj}\boldsymbol{\beta}_n)}$$

For unknown $\boldsymbol{\beta}_n$, the unconditional choice probability is obtained by integrating the conditional probability with respect to the marginal distribution for $\boldsymbol{\beta}_n$ giving the mixed logit probability.

Remarks:

- Random coefficients formulation produces correlation among utilities across alternatives:

$$\begin{aligned} cov(U_{ni}, U_{nj}) &= cov(\mathbf{x}'_{ni}\boldsymbol{\beta}_n + \varepsilon_{ni}, \mathbf{x}'_{nj}\boldsymbol{\beta}_n + \varepsilon_{nj}) \\ &= cov(\mathbf{x}'_{ni}\boldsymbol{\beta}_n, \mathbf{x}'_{nj}\boldsymbol{\beta}_n) = \mathbf{x}'_{ni} var(\boldsymbol{\beta}_n)\mathbf{x}_{nj} \end{aligned}$$

  which is, in general, non-zero even if $var(\boldsymbol{\beta}_n)$ is a diagonal matrix.

- Usually, the elements of $\boldsymbol{\beta}_n$ are assumed to be uncorrelated so that $\boldsymbol{\theta}$ may be partitioned into a vector of location parameters $\mathbf{b}$ and a vector of scale parameters $\mathbf{s}$. For example,

$$\boldsymbol{\beta}_n | \boldsymbol{\theta} \sim iid\ N(\mathbf{b}, \mathbf{S})$$

  where $\mathbf{S}$ is a diagonal matrix with s_{ii} along the diagonal. This implies that

$$\beta_{nk} \sim iid\ N(b_k, s_{kk})$$

- Usually, some elements of $\boldsymbol{\beta}_n$ are assumed to be fixed.

6.2 Random Effects Interpretation

Typical random-effects model formulation for utility

$$U_{ni} = \mathbf{x}_{ni}'\boldsymbol{\alpha} + \mathbf{z}_{ni}'\boldsymbol{\mu}_n + \varepsilon_{ni}$$

where

- $\mathbf{x}_{ni}$ are observed covariates with fixed coefficients $\boldsymbol{\alpha}$
- $\mathbf{z}_{ni}$ are observed covariates with random coefficients $\boldsymbol{\mu}_n$
- $\mathbf{x}_{ni}$ and $\mathbf{z}_{ni}$ may have common elements
- ε_{ni} is distributed iid extreme value.

Result: correlation in utility across alternatives is determined by z_{ni} and the distribution of $\boldsymbol{\mu}_n$

$$\begin{aligned} & cov(U_{ni}, U_{nj}) \\ = \; & cov(\mathbf{x}'_{ni}\boldsymbol{\alpha} + \mathbf{z}'_{ni}\boldsymbol{\mu}_n + \varepsilon_{ni}, \mathbf{x}'_{nj}\boldsymbol{\alpha} + \mathbf{z}'_{nj}\boldsymbol{\mu}_n + \varepsilon_{nj}) \\ = \; & cov(\mathbf{z}'_{ni}\boldsymbol{\mu}_n, \mathbf{z}'_{nj}\boldsymbol{\mu}_n) = \mathbf{z}_{ni} var(\boldsymbol{\mu}_n)\mathbf{z}_{nj} \end{aligned}$$

- The covariates z_{ni} are usually dummy variables constructed by the user in a particular way to produce a desired pattern of correlation across alternatives.

- Usually the elements of $\boldsymbol{\mu}_n$ are assumed uncorrelated with mean zero. For example,

  $$\boldsymbol{\mu}_n \sim iid\ N(0, \mathbf{S})$$

  where $\mathbf{S}$ is a diagonal matrix with s_{ii} along the diagonal. This implies that

  $$\mu_{nk} \sim iid\ N(0, s_{kk})$$

- Random coefficient model is a random effects model

$$\begin{aligned}
U_{ni} &= \mathbf{x}'_{ni}\boldsymbol{\beta}_n + \varepsilon_{ni} \\
&= \mathbf{x}'_{ni}\mathbf{b} + \mathbf{x}'_{ni}(\boldsymbol{\beta}_n - \mathbf{b}) + \varepsilon_{ni} \\
E[\boldsymbol{\beta}_n] &= \mathbf{b}
\end{aligned}$$

6.3 Mixing Distributions

Common univariate mixing distributions for β_{nk} are

- Normal
- Log normal - useful if β_{nk} is the same sign for all n
- Uniform - useful to bound range of β_{nk}
- Triangular - useful to bound range of β_{nk}

Above distributions may be characterized in terms of 2 parameters:

$$
\begin{aligned}
b_k &= \text{location} \\
s_k &= \text{scale}
\end{aligned}
$$

Simulation of β_{nk}:

$$\begin{aligned} \beta_k &= b_k + s_k \cdot F^{-1}(\varepsilon_k) \\ \varepsilon_k &\sim U[0,1] \\ F^{-1}() &= \text{ inverse standardized CDF} \end{aligned}$$

Remark: With the random-effects specification, usually $b_k = 0$.

6.4 Simulation of P_{ni}

- Draw $\boldsymbol{\beta}^r$ randomly from $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ R times

- Form the average of the logit kernels over the random draws:

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

- $\check{P}_{ni}(\boldsymbol{\theta})$ is an unbiased estimate of $P_{ni}(\boldsymbol{\theta})$ by construction.

- By the law of large numbers, as $R \rightarrow \infty$

$$\check{P}_{ni}(\boldsymbol{\theta}) \xrightarrow{p} P_{ni}(\boldsymbol{\theta})$$
$$SD(\check{P}_{ni}(\boldsymbol{\theta})) = O(\sqrt{R}).$$

6.5 Maximum Simulated Likelihood Estimation

Simulated Log-likelihood function

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{i=1}^{J} y_{ni} \ln(\check{P}_{ni}(\boldsymbol{\theta}))$$

$$\boldsymbol{\theta} = \text{ parameters of } f(\boldsymbol{\beta}|\boldsymbol{\theta})$$

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

6.6 Example

- Data on the mode choice of 453 commuters: (1) car alone, (2) carpool, (3) bus, and (4) rail.

- Observe cost and time on each mode and the chosen mode

- Random utility model with $J = 4$ alternatives and $N = 453$ commuters

$$\begin{aligned} U_{ni} &= V_{ni} + \varepsilon_{ni}, \\ V_{ni} &= \mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{s}'_n\boldsymbol{\delta}_i \\ P_{ni} &= \Pr(U_{ni} > U_{nj},\ \forall\ j \neq i) \end{aligned}$$

  where $\mathbf{x}_{ni}$ denotes the 2×1 vector of cost and time variables that vary over individuals and alternatives, and $\mathbf{s}_n$ represents a 3×1 vector of alternative specific constants.

- Mixed effects specification: alternative specific random effects

$$
\begin{aligned}
U_{ni} &= \beta_1 \mathsf{cost}_{ni} + \beta_2 \mathsf{time}_{ni} + \\
&\quad \beta_{3n} \mathsf{Carpool}_n + \beta_{4n} \mathsf{Bus}_n + \\
&\quad \beta_{5n} \mathsf{Rail}_n + \varepsilon_{ni} \\
&\quad \beta_{kn} \sim N(b_k, s_k^2),\ k = 3, 4, 5
\end{aligned}
$$

6.7 Extension to Panel Data

With panel data, utility may be expressed as

$$U_{njt} = \mathbf{x}'_{njt}\boldsymbol{\beta}_{nt} + \varepsilon_{njt}$$
$$\varepsilon_{njt} \sim iid \text{ extreme value } \forall\ n, j, t$$
$$\text{individuals: } n = 1, \ldots, N$$
$$\text{alternatives: } j = 1, \ldots, J$$
$$\text{time periods: } t = 1, \ldots, T$$

6.7.1 Time invariant random effects

- Assume tastes vary over individuals but remain constant over time: $\boldsymbol{\beta}_{nt} = \boldsymbol{\beta}_n$.

- Let $\mathbf{i}_n = (i_1, \ldots, i_{T_n})'$ denote a vector of alternatives for each time period $t = 1, \ldots, T_n$ faced by individual n (allows for unballanced panels).

- $\mathbf{x}'_{ni_tt}$ may contain lagged values of $\mathbf{x}_{ni_tt}$ and y_{ni_tt}

- Conditional on $\boldsymbol{\beta}_n$, the probability that individual n selects the sequence of alternatives specified by $\mathbf{i}_n$ is

$$P_{n\mathbf{i}_n}|\boldsymbol{\beta}_n = L_{n\mathbf{i}_n}(\boldsymbol{\beta}_n) = \prod_{t=1}^{T_n} \frac{\exp(\mathbf{x}'_{ni_tt}\boldsymbol{\beta}_n)}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{njt}\boldsymbol{\beta}_n)}$$

  For unknown $\boldsymbol{\beta}_n$, the unconditional choice probability is obtained by integrating the conditional probability with respect to the marginal distribution for $\boldsymbol{\beta}_n$ giving the mixed logit probability

$$P_{n\mathbf{i}_n}(\boldsymbol{\theta}) = \int \cdots \int L_{n\mathbf{i}_n}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n|\boldsymbol{\theta}) d\boldsymbol{\beta}_n$$

6.7.2 Time Dependent Random Effects

$$
\begin{gathered}
U_{njt} = \mathbf{x}'_{njt}\boldsymbol{\beta}_{nt} + \varepsilon_{njt} \\
\varepsilon_{njt} \sim iid \text{ extreme value } \forall\ n, j, t \\
\beta_{k,nt} = b_k + \tilde{\beta}_{k,nt} \\
\tilde{\beta}_{k,nt} = \rho\tilde{\beta}_{k,nt} + w_{k,nt} \\
w_{k,nt} \sim iid\ N(0, s^2)
\end{gathered}
$$

Simulation of P_{ni} (single β_{nt})

1. Draw w^r_{n1} for 1st period and calculate

$$
\begin{gathered}
\tilde{\beta}^r_{n1} = w^r_{n1} \\
\beta^r_{n1} = b + \tilde{\beta}^r_{n1} \\
L_{ni_1}(\beta^r_{n1})
\end{gathered}
$$

2. Draw w^r_{n2} for 2nd period and calculate

$$
\begin{gathered}
\tilde{\beta}^r_{n2} = \rho\beta^r_{n1} + w^r_{n2} \\
\beta^r_{n2} = b + \tilde{\beta}^r_{n2} \\
L_{ni_2}(\beta^r_{n2})
\end{gathered}
$$

3. Repeat 1 - 2 for $t = 3, \ldots, T$

4. Compute

$$\prod_{t=1}^{T} L_{ni_t}(\beta_{nt}^r)$$

5. Repeat 1 - 4 R times and compute

$$\check{P}_{n\mathbf{i}} = \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} L_{ni_t}(\beta_{nt}^r)$$

6.8 Example: Choice of Electricity Supplier

A sample of 361 residential electricity customers were asked a series of questions of choice experiments. In each experiment, 4 hypothetical electricity suppliers were described. The person was asked which of the 4 suppliers he/she would choose. As many as 12 experiments were presented to each person, and some people stopped before answering all questions. There are a total of 4308 experiments.

Characteristics of suppliers

- Price of supplier - (a) fixed price per kWh; (b) time-of-day rate; (c) seasonal rate
- Length of contract in years
- Local or well known company

Mixed logit model specification: time homogeneous panel

$$
\begin{aligned}
U_{nit} &= \beta_1 \mathsf{kWhPrice}_{nit} + \beta_{2n} \mathsf{length}_{nit} + \beta_{3n} \mathsf{local}_{nit} \\
&\quad + \beta_{4n} \mathsf{wellknown}_{nit} + \beta_5 \mathsf{TOD}_{nit} \\
&\quad + \beta_{6n} \mathsf{seasonal}_{nit} + \varepsilon_{nit} \\
&\quad \beta_{kn} \sim N(b_k, s_k^2)
\end{aligned}
$$

7 Low Discrepancy Sequences

A Monte Carlo estimate for the s-dimensional integral is given by:

$$I = \frac{1}{N} \sum_{i=0}^{N-1} f(\mathbf{x}_i), \mathbf{x}_i \in [0,1)^s$$

- If the $\mathbf{x}_i$ are pseudo-random uniform vectors the SD of the approximation error is $O(N^{1/2})$.
  - Rate is independent of the dimension of the problem.
  - Rate is slow; to obtain an extra digit of accuracy requires 100 times as many function evaluations.

- Since the early 1990s, more evenly-spaced sequences based on a mathematical concept called *discrepancy* have been widely used in physics, engineering, and financial applications, in so-called *quasi-Monte Carlo* experiments. In most cases low discrepancy sequences (LDS) improve the accuracy of the estimate, I, typically offering a convergence rate of $O(N^{-1}\log(N)^s)$, where s is the dimension.

  - As N increases for fixed s, rate behaves more like N^{-1} than $N^{-1/2}$, and an extra digit of accuracy requires little more than 10 times as many function evaluations.

  - The methods apply most naturally to applications that can be transformed into an integral over the s-dimension unit hypercube. Then a LDS in the sequel tries to choose points which are evenly distributed in the hypercube.

7.0.1 Discrepancy and Low Discrepancy Sequences

- $Q \subseteq [0,1]^s$ is an axially parallel $s-$dimensional rectangle

- $x_1, \ldots, x_N \in [0,1]^s$

- Intuitively, an evenly distributed point set satisfies

$$\frac{\#\text{ of } x_i \in Q}{\#\text{ of all points}} \approx \frac{vol(Q)}{vol([0,1]^s)}$$

  for as many rectangles Q as possible

Defn: The discrepancy of point set $\{x_1, \ldots, x_N\}$ is

$$D_N = \sup_Q \left| \frac{\#\text{ of } x_i \in Q}{N} - vol(Q) \right|$$

Defn: A sequence of points $x_1, x_2, \ldots, x_N$ is a low discrepancy sequence if there exists a constant C_s, independent of N, such that

$$D_N \leq C_s \frac{(\log N)^s}{N}$$

LDS implemented in S+lowDiscrepancy

- Halton sequences
  with scrambling by permutations or random shifts,

- Faure sequences
  with Owen and/or Faure-Tezuka scrambling,

- Sobol sequences
  with Owen and/or Faure-Tezuka scrambling,

- Niederreiter sequences
  with Owen and/or Faure-Tezuka scrambling, and

- Niederreiter-Xing sequences
  with Owen and/or Faure-Tezuka scrambling.

7.0.2 Problems with LDS

- If s is moderately large, the initial behavior of unscrambled versions of some LDSs can be bad.

- To overcome such poor behavior a number of *scrambling* methods have been proposed in the literature.

- With deterministic LDSs it is difficult to estimate the integration error rate of the approximation. However, with randomized LDSs it is possible to estimate the error rate using the usual statistical methods.

7.1 Improving Accuracy of GHK Simulator

References

- Sandor, Z. and P. Andras (2003). "Alternative Sampling Methods for Estimating Multivariate Normal Probabilities," mimeo, Econometric Institute, Erasmus University Rotterdam.
  - Studied the performance of the GHK simulator using a variety of LDSs.
  - The GHK simulator with LDSs outperformed the GHK simulator with random uniforms, sometimes by a huge margin
  - At dimension 10, for the percision of the GHK simulator with random draws to match the precision with Owen scrambled Niederreiter-Xing sequences the number of random draws need to be 100 times larger

– At dimension 50, random samples need to be about 4 times larger than LDS samples to achieve the same level of accuracy.

7.2 Improving Accuracy of Mixed Logit Probabilities

References

1. Train, K. (2000). "Halton Sequences for Mixed Logit," mimeo, Dept. of Economics, UC Berkeley.

- For $s \leq 5$, found simulation variance in estimated parameters from mixed logit models to be considerably smaller with 100 Halton draws than with 1000 random draws.

- Even coverage of Halton draws over the domain of integration reduces variability of $\check{P}_{ni}$ across n relative to random draws.

- With Halton sequences, the draws for one observation tend to fill in the spaces that were left empty by the previous observations causing $\check{P}_{ni}$ to be negatively correlated over observations. This reduces the variance of $\ln \check{L}(\theta)$.

2. Bhat, C. (2001). "Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model," *Transportation Research*.

   - For $s \leq 5$ and 125 Halton draws, found the simulation error to be half as large as with 1000 random draws and smaller than with 2000 random draws.

3. Sandor, Z. and K. Train (2002). "Quasi-random Simulation of Discrete Choice Models," mimeo, Dept. of Economics, UC Berkeley.

   - Analyzed the use of Halton, randomized Halton, and scrambled Neidereitter-Xing sequences for the evaluate on mixed logit probabilities.

- Found that 8 times as many random draws are required to reach the same level of performance as the best LDS studied (Neidereitter-Xing with Owen-type scrambling).

4. Bhat, C. (2003). "Simulation estimation of mixed discrete choice models using randomized and scrambled halton sequences," *Transportation Research*.

   - For high dimensional problems non-randomized Halton sequences do not perform as well as for low dimensional problems. For $s = 10$, 150 standard Halton draws are comparable to 500 random draws.

   - Randomized Halton sequences perform well for high dimensional problems. For $s = 10$, 100 scrambled Halton draws perform as well as 150 standard Halton draws. 150 scrambled Halton draws perform as well as 1000 random draws.