Random Sampling of Alternatives for Route Choice Modeling

Emma Frejinger and Michel Bierlaire

Transport and Mobility Laboratory, EPFL, transp-or.epfl.ch
Outline

• Introduction
• Stochastic path enumeration approach
• Sampling of alternatives
• Numerical results
• Conclusions
Introduction

- Route choice problem
  Given a transportation network composed of nodes, links, origin and destinations. For a given transportation mode and origin-destination pair, which is the chosen route?

- Discrete choice modeling framework

- Issue
  Universal choice set very large, individual specific choice set unknown
Introduction

Set of all paths $\mathcal{U}$ from $o$ to $d$

Path generation
- Deterministic
  - $\mathcal{M} \subseteq \mathcal{U}$
- Stochastic
  - $\mathcal{M}_n \subseteq \mathcal{U}$

Route choice Choice set formation
- Deterministic
- Probabilistic

Route choice model

$P(i|C_n) P(i) = \sum_{C_n \in G_n} P(i|C_n) P(C_n)$
Introduction

- Underlying assumption in existing approaches: the actual choice set is generated
- Empirical results suggest that this is not always true
- Our approach:
  - True choice set = universal set $\mathcal{U}$
  - Too large
  - Sampling of alternatives
Sampling of Alternatives

- Multinomial logit model (e.g. Ben-Akiva and Lerman, 1985):

\[
P(i|C_n) = \frac{q(C_n|i)P(i)}{\sum_{j\in C_n} q(C_n|j)P(j)} = \frac{e^{V_{in} + \ln q(C_n|i)}}{\sum_{j\in C_n} e^{V_{jn} + \ln q(C_n|j)}}
\]

- \(C_n\): set of sampled alternatives
- \(q(C_n|j)\): probability of sampling \(C_n\) given that \(j\) is the chosen alternative
Importance Sampling of Alternatives

- Attractive paths have higher probability of being sampled than unattractive paths.
- Path utilities must be corrected in order to obtain unbiased estimation results.
MNL Route Choice Models

- Path Size Logit (Ben-Akiva and Ramming, 1998 and Ben-Akiva and Bierlaire, 1999) and C-Logit (Cascetta et al. 1996)
- Additional attribute in the deterministic utilities capturing correlation among alternatives
- These attributes should reflect the true correlation structure
- Hypothesis: attributes should be computed based on all paths (or as many as possible)
Stochastic Path Enumeration

- Flexible approach that can be combined with various algorithms, here a biased random walk approach
- The probability of a link $\ell$ with source node $v$ and sink node $w$ is modeled in a stochastic way based on its distance to the shortest path
- Kumaraswamy distribution, cumulative distribution function $F(x_\ell|a, b) = 1 - (1 - x_\ell^a)^b$ for $x_\ell \in [0, 1]$.

$$x_\ell = \frac{SP(v, d)}{C(\ell) + SP(w, d)}$$
Stochastic Path Enumeration

\[ F(x_{\ell} | a, b) \]

\[ a = 1, 2, 5, 10, 30 \]
\[ b = 1 \]
Stochastic Path Enumeration

- Probability for path $j$ to be sampled

$$q(j) = \prod_{\ell=(v,w) \in \Gamma_j} q((v, w)|E_v)$$

- $\Gamma_j$: ordered set of all links in $j$
- $v$: source node of $j$
- $E_v$: set of all outgoing links from $v$
- In theory, the set of all paths $\mathcal{U}$ may be unbounded. We treat it as bounded with size $J$. 
Sampling of Alternatives

- Following Ben-Akiva (1993)
- Sampling protocol
  1. A set $\tilde{C}_n$ is generated by drawing $R$ paths with replacement from the universal set of paths $\mathcal{U}$
  2. Add chosen path to $\tilde{C}_n$
- Outcome of sampling: $(\tilde{k}_1, \tilde{k}_2, \ldots, \tilde{k}_J)$ and $\sum_{j=1}^J \tilde{k}_j = R$

$$P(\tilde{k}_1, \tilde{k}_2, \ldots, \tilde{k}_J) = \frac{R!}{\prod_{j \in \mathcal{U}} \tilde{k}_j！} \prod_{j \in \mathcal{U}} q(j)^{\tilde{k}_j}$$

- Alternative $j$ appears $k_j = \tilde{k}_j + \delta_{cj}$ in $\tilde{C}_n$
Sampling of Alternatives

- Let \( C_n = \{ j \in U \mid k_j > 0 \} \)

\[
q(C_n|i) = q(\tilde{C}_n|i) = \frac{R!}{(k_i - 1)!} \prod_{j \in C_n, j \neq i} k_j! \prod_{j \in C_n} q(j)^{k_j} = K_{C_n} \frac{k_i}{q(i)}
\]

\[
K_{C_n} = \frac{R!}{\prod_{j \in C_n} k_j!} \prod_{j \in C_n} q(j)^{k_j}
\]

\[
P(i|C_n) = \frac{e^{V_{in} + \ln\left(\frac{k_i}{q(i)}\right)}}{\sum_{j \in C_n} e^{V_{jn} + \ln\left(\frac{k_j}{q(j)}\right)}}
\]
Numerical Results

- Estimation of models based on synthetic data generated with a postulated model
- Evaluation of
  - Sampling correction
  - Path Size attribute
  - Biased random walk algorithm parameters
Numerical Results
Numerical Results

- True model: Path Size Logit

\[ U_j = \beta_{PS} PS^U_j + \beta_L \text{Length}_j + \beta_{SB} \text{SpeedBumps}_j + \varepsilon_j \]

\[ \beta_{PS} = 1, \beta_L = -0.3, \beta_{SB} = -0.1 \]

\( \varepsilon_j \) distributed Extreme Value with scale 1 and location 0

\[ PS^U_j = \sum_{\ell \in \Gamma_j} \frac{L_\ell}{L_j} \frac{1}{\sum_{p \in \mathcal{U}} \delta_{lp}} \]

- 3000 observations
Numerical Results

- Four model specifications

<table>
<thead>
<tr>
<th>Path Size</th>
<th>Without</th>
<th>With</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>$M_{PS(C)}^{NoCorr}$</td>
<td>$M_{PS(C)}^{Corr}$</td>
</tr>
<tr>
<td>U</td>
<td>$M_{PS(U)}^{NoCorr}$</td>
<td>$M_{PS(U)}^{Corr}$</td>
</tr>
</tbody>
</table>

$$PS_i^U = \sum_{\ell \in \Gamma_i} \left( \frac{L_{\ell}}{L_i} \frac{1}{\sum_{j \in U} \delta_{\ell j}} \right)$$

$$PS_{in}^C = \sum_{\ell \in \Gamma_i} \left( \frac{L_{\ell}}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{\ell j}} \right)$$
Numerical Results

- **Model** $M_{PS(C)}^{NoCorr}$:
  \[ V_{in} = \mu \left( \beta_{PS} PS_{in}^C - 0.3Length_i + \beta_{SB} SpeedBumps_i \right) \]

- **Model** $M_{PS(C)}^{Corr}$:
  \[ V_{in} = \mu \left( \beta_{PS} PS_{in}^C - 0.3Length_i + \beta_{SB} SpeedBumps_i + \ln \left( \frac{k_i}{q(i)} \right) \right) \]

- **Model** $M_{PS(U)}^{NoCorr}$:
  \[ V_{in} = \mu \left( \beta_{PS} PS_{in}^U - 0.3Length_i + \beta_{SB} SpeedBumps_i \right) \]

- **Model** $M_{PS(U)}^{Corr}$:
  \[ V_{in} = \mu \left( \beta_{PS} PS_{in}^U - 0.3Length_i + \beta_{SB} SpeedBumps_i + \ln \left( \frac{k_i}{q(i)} \right) \right) \]
## Numerical Results

<table>
<thead>
<tr>
<th></th>
<th>True PSL</th>
<th>$M_{PS(c)}^{NoCorr}$ PSL</th>
<th>$M_{PS(c)}^{Corr}$ PSL</th>
<th>$M_{PS(u)}^{NoCorr}$ PSL</th>
<th>$M_{PS(u)}^{Corr}$ PSL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_L$ fixed</td>
<td>-0.3</td>
<td>-0.3</td>
<td>-0.3</td>
<td>-0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>1</td>
<td>0.182</td>
<td>0.724</td>
<td>0.141</td>
<td>0.994</td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
<td>0.0277</td>
<td>0.0226</td>
<td>0.0263</td>
<td>0.0286</td>
</tr>
<tr>
<td>t-test w.r.t. 1</td>
<td></td>
<td>-29.54</td>
<td>-12.21</td>
<td>-32.64</td>
<td>-0.2</td>
</tr>
<tr>
<td>$\hat{\beta}_{PS}$</td>
<td>1</td>
<td>1.94</td>
<td>0.411</td>
<td>-1.02</td>
<td>1.04</td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
<td>0.428</td>
<td>0.104</td>
<td>0.383</td>
<td>0.0474</td>
</tr>
<tr>
<td>t-test w.r.t. 1</td>
<td></td>
<td>2.20</td>
<td>-5.66</td>
<td>-5.27</td>
<td>0.84</td>
</tr>
<tr>
<td>$\hat{\beta}_{SB}$</td>
<td>-0.1</td>
<td>-1.91</td>
<td>-0.226</td>
<td>-2.82</td>
<td>-0.0867</td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
<td>0.25</td>
<td>0.0355</td>
<td>-6.58</td>
<td>0.0238</td>
</tr>
<tr>
<td>t-test w.r.t. -0.1</td>
<td></td>
<td>-7.24</td>
<td>-3.55</td>
<td>0.41</td>
<td>0.56</td>
</tr>
</tbody>
</table>
# Numerical Results

<table>
<thead>
<tr>
<th>True PSL</th>
<th>$M_{PS(C)}^{NoCorr}$ PSL</th>
<th>$M_{PS(C)}^{Corr}$ PSL</th>
<th>$M_{PS(U)}^{NoCorr}$ PSL</th>
<th>$M_{PS(U)}^{Corr}$ PSL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Log-likelihood</td>
<td>-6660.45</td>
<td>-6082.53</td>
<td>-6666.82</td>
<td>-5933.98</td>
</tr>
<tr>
<td>Adj. Rho-square</td>
<td>0.018</td>
<td>0.103</td>
<td>0.017</td>
<td>0.125</td>
</tr>
</tbody>
</table>

Null Log-likelihood: -6784.96, 3000 observations

Algorithm parameters: 10 draws, $a = 5$, $b = 1$, $C(\ell) = L_\ell$

Average size of sampled choice sets: 9.66

BIOGEME (Bierlaire, 2007 and Bierlaire, 2003) has been used for all model estimations
Numerical Results

- Speed Bump Coefficient
- Scale Parameter
- Path Size Coefficient

Importance sampling of alternatives for route choice models – p.21/23
Numerical Results

![3D graph showing the average number of paths in choice sets as a function of the number of draws and the Kumaraswamy parameter a. The graph is color-coded, with darker colors indicating higher values.]
Conclusions

- New point of view on choice set generation and route choice modeling
- Path generation is considered an importance sampling approach
- We present a path generation algorithm and derive the corresponding sampling correction
- Path Size should be computed based on true correlation structure
- Numerical results are very promising