Stochastic Path Generation Algorithm for Route Choice Models

Emma Frejinger and Michel Bierlaire

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

- Introduction
- Stochastic path enumeration approach
- Sampling of alternatives
- Preliminary 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

- Choice sets need to be defined prior to the route choice modeling.
- Path enumeration algorithms are used for this purpose, many heuristics have been proposed, for example:
  - Deterministic approaches: link elimination (Azevedo et al., 1993), labeled paths (Ben-Akiva et al., 1984)
  - Stochastic approaches: simulation (Ramming, 2001) and doubly stochastic (Bovy and Fiorenzo-Catalano, 2006)
Introduction

- Underlying assumption: the actual choice set is generated
- Empirical results suggest that this is not always true
- Our approach:
  - True choice set = universal set
  - 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
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 \( v \)
- Issue: in theory, the set of all paths \( \mathcal{U} \) is 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 $C_n = \{ j \in U \mid k_j > 0 \}$

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

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

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

- Estimation of models based on synthetic data generated with postulated models
  - Non-correlated paths
    Postulated model same as estimated model (multinomial logit)
  - Correlated paths in a “grid-like” network
    Postulated model is probit and estimated models are multinomial logit and path size logit
- True parameter values are compared to estimates
Preliminary Numerical Results

\[ SB_1 = 3, \quad L_1 = 2 \]
\[ SB_2 = 4, \quad L_2 = 4 \]
\[ SB_3 = 1, \quad L_3 = 6 \]
\[ SB_{39} = 0, \quad L_{39} = 78 \]
\[ SB_{40} = 0, \quad L_{40} = 80 \]
Preliminary Numerical Results

- True model: multinomial logit
  \[ U_j = \beta_L \text{length}_j + \beta_{SB} \text{nbspeedbumps}_j + \varepsilon_j \]
  \[ \beta_L = -0.6 \text{ and } \beta_{SB} = -0.3 \]
  \[ \varepsilon_j \text{ is distributed Extreme Value with location parameter 0 and scale 1} \]

- 500 observations, therefore 500 choice sets are sampled

- Biased random walk using 40 draws with \( a = 2 \) and \( b = 1 \)

  Generated choice sets include at least 7, maximum 18 and on average 11.9 paths
## Preliminary Numerical Results

<table>
<thead>
<tr>
<th>Sampling correction</th>
<th>MNL without</th>
<th>MNL with</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_L (-0.6)$</td>
<td>-0.203</td>
<td>-0.286</td>
</tr>
<tr>
<td>Scaled estimate</td>
<td>-0.600</td>
<td>-0.600</td>
</tr>
<tr>
<td>Robust std.</td>
<td>0.0193</td>
<td>0.019</td>
</tr>
<tr>
<td>Robust t-test</td>
<td>-10.53</td>
<td>-15.01</td>
</tr>
<tr>
<td>$\hat{\beta}_{SB} (-0.3)$</td>
<td>-0.0194</td>
<td>-0.143</td>
</tr>
<tr>
<td>Scaled estimate</td>
<td>-0.0573</td>
<td>-0.300</td>
</tr>
<tr>
<td>Robust std.</td>
<td>0.0662</td>
<td>0.0661</td>
</tr>
<tr>
<td>Robust t-test</td>
<td>-0.29</td>
<td>-2.17</td>
</tr>
<tr>
<td>Null log-likelihood</td>
<td>-1069.453</td>
<td>-1633.501</td>
</tr>
<tr>
<td>Final log-likelihood</td>
<td>-788.42</td>
<td>-759.848</td>
</tr>
<tr>
<td>Adjusted $\hat{\rho}^2$</td>
<td>0.261</td>
<td>0.288</td>
</tr>
</tbody>
</table>

BIOGEME has been used for all model estimations.
Preliminary Numerical Results
Preliminary Numerical Results

- True model: probit (Burrell, 1968)
  \[ U_\ell = \beta_L \text{length}_\ell + \beta_{SB} \text{nbspeedbumps}_\ell + \sigma \sqrt{L_\ell} \nu_\ell \]
  \( \beta_L = -0.6 \) and \( \beta_{SB} = -0.4 \)
  \( \nu_\ell \) is distributed standard Normal
  Link utility variance assumed proportional to length with parameter \( \sigma = 0.8 \)

- Path utilities are link additive

- 382 observations are generated after 500 realizations of the link utilities
Preliminary Numerical Results

- Biased random walk using 30 draws with $a = 2$ and $b = 1$ (382 choice sets)
  Generated choice sets include at least 7, maximum 19 and on average 13.5 paths
## Preliminary Numerical Results

<table>
<thead>
<tr>
<th>Sampling correction</th>
<th>MNL without</th>
<th>MNL with</th>
<th>PSL without</th>
<th>PSL with</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_L (-0.6)$</td>
<td>-0.627</td>
<td>-0.978</td>
<td>-0.619</td>
<td>-0.969</td>
</tr>
<tr>
<td>Scaled estimate</td>
<td>-0.600</td>
<td>-0.600</td>
<td>-0.600</td>
<td>-0.600</td>
</tr>
<tr>
<td>Robust std.</td>
<td>0.0397</td>
<td>0.032</td>
<td>0.0407</td>
<td>0.0358</td>
</tr>
<tr>
<td>Robust t-test</td>
<td>-15.79</td>
<td>-30.57</td>
<td>-15.22</td>
<td>-27.04</td>
</tr>
<tr>
<td>$\hat{\beta}_{SB} (-0.4)$</td>
<td>-0.0822</td>
<td>-0.0801</td>
<td>-0.347</td>
<td>-0.461</td>
</tr>
<tr>
<td>Scaled estimate</td>
<td>-0.0787</td>
<td>-0.0491</td>
<td>-0.336</td>
<td>-0.285</td>
</tr>
<tr>
<td>Robust std.</td>
<td>0.052</td>
<td>0.0559</td>
<td>0.182</td>
<td>0.158</td>
</tr>
<tr>
<td>Robust t-test</td>
<td>-1.58</td>
<td>-1.43</td>
<td>-1.90</td>
<td>-2.92</td>
</tr>
<tr>
<td>$\hat{\beta}_{PS}$</td>
<td></td>
<td>1.17</td>
<td>1.74</td>
<td></td>
</tr>
<tr>
<td>Scaled estimate</td>
<td></td>
<td>1.13</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>Robust std.</td>
<td></td>
<td>0.788</td>
<td>0.705</td>
<td></td>
</tr>
<tr>
<td>Robust t-test</td>
<td></td>
<td>1.49</td>
<td>2.47</td>
<td></td>
</tr>
</tbody>
</table>
## Preliminary Numerical Results

<table>
<thead>
<tr>
<th>Sampling correction</th>
<th>MNL without</th>
<th>MNL with</th>
<th>PSL without</th>
<th>PSL with</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null log-likelihood</td>
<td>-988.63</td>
<td>-2769.959</td>
<td>-988.63</td>
<td>-2769.959</td>
</tr>
<tr>
<td>Final log-likelihood</td>
<td>-676.111</td>
<td>-653.396</td>
<td>-674.481</td>
<td>-649.268</td>
</tr>
<tr>
<td>Adjusted $\rho^2$</td>
<td>0.314</td>
<td>0.337</td>
<td>0.315</td>
<td>0.340</td>
</tr>
</tbody>
</table>

BIOGEME has been used for all model estimations.
Conclusions and Future Work

- Stochastic path enumeration algorithms are viewed as an approach for importance sampling of alternatives
- We propose an algorithm that allows for computation of path selection probabilities and correction for sampling
- Ongoing research, further work will be dedicated, for example, to empirical results on real data and correction in prediction