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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Sampling of Alternatives

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

$$P(i|\mathcal{C}_n) = \frac{q(\mathcal{C}_n|i)P(i)}{\sum_{j\in\mathcal{C}_n} q(\mathcal{C}_n|j)P(j)} = \frac{e^{V_{in}+\ln q(\mathcal{C}_n|i)}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}+\ln q(\mathcal{C}_n|j)}}$$

$\mathcal{C}_n$: set of sampled alternatives
$q(\mathcal{C}_n|j)$: probability of sampling $\mathcal{C}_n$ given that j is the chosen alternative

Importance Sampling of Alternatives

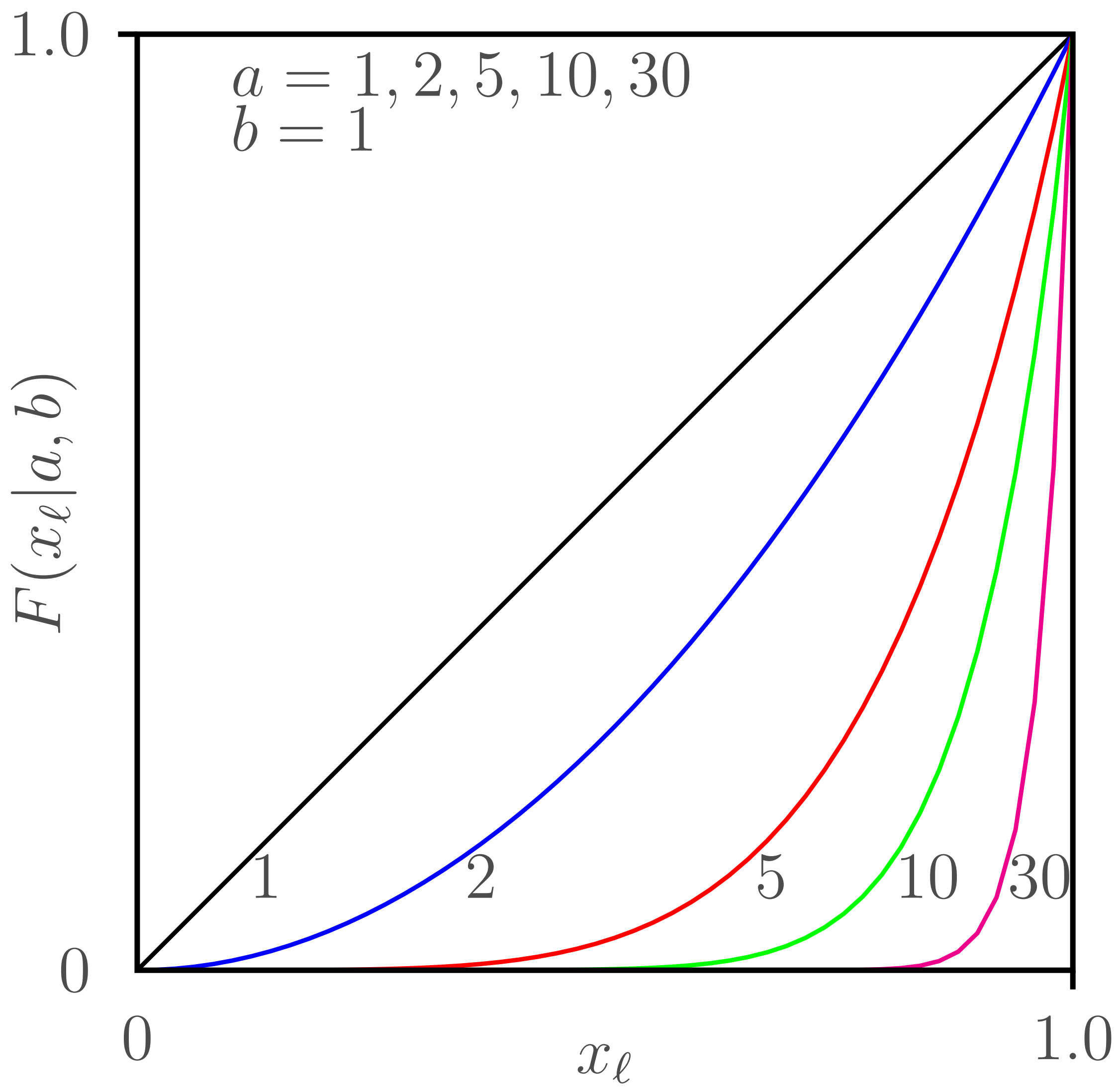
- Attractive paths have higher probability of being sampled than unattractive paths
- Path utilities must be corrected in order to obtain unbiased estimation results

Stochastic Path Enumeration

- Flexible approach that can be combined with various algorithms, here a biased random walk approach
- The probability of a link ℓ 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



Stochastic Path Enumeration

- Probability for path j to be sampled

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

- Γ_j: ordered set of all links in j
- v: source node of j
- $\mathcal{E}_v$: 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 $\widetilde{\mathcal{C}}_n$ is generated by drawing R paths with replacement from the universal set of paths $\mathcal{U}$
  2. Add chosen path to $\widetilde{\mathcal{C}}_n$
- Outcome of sampling: $(\widetilde{k}_1, \widetilde{k}_2, \ldots, \widetilde{k}_J)$ and $\sum_{j=1}^{J} \widetilde{k}_j = R$

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

- Alternative j appears $k_j = \widetilde{k}_j + \delta_{cj}$ in $\widetilde{\mathcal{C}}_n$

Sampling of Alternatives

- Let $\mathcal{C}_n = \{j \in \mathcal{U} \mid k_j > 0\}$

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

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

$$P(i|\mathcal{C}_n) = \frac{e^{V_{in}+\ln\left(\frac{k_i}{q(i)}\right)}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}+\ln\left(\frac{k_j}{q(j)}\right)}}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

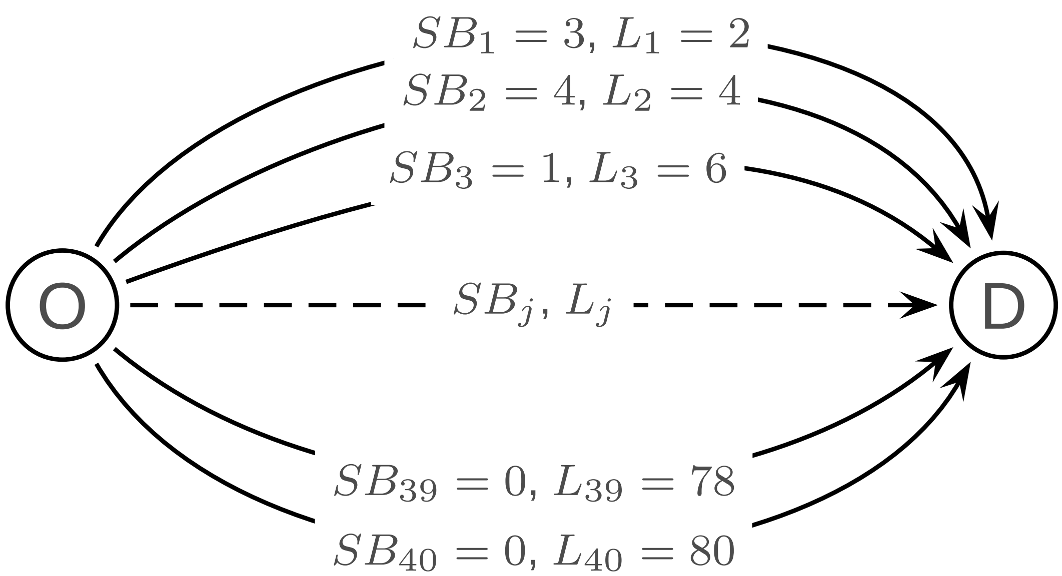
Preliminary Numerical Results

- 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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Preliminary Numerical Results



Preliminary Numerical Results

- True model: multinomial logit
  $U_j = \beta_{\mathsf{L}}$ length$_j + \beta_{\mathsf{SB}}$ nbspeedbumps$_j + \varepsilon_j$
  $\beta_{\mathsf{L}} = -0.6$ and $\beta_{\mathsf{SB}} = -0.3$
  ε_j 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

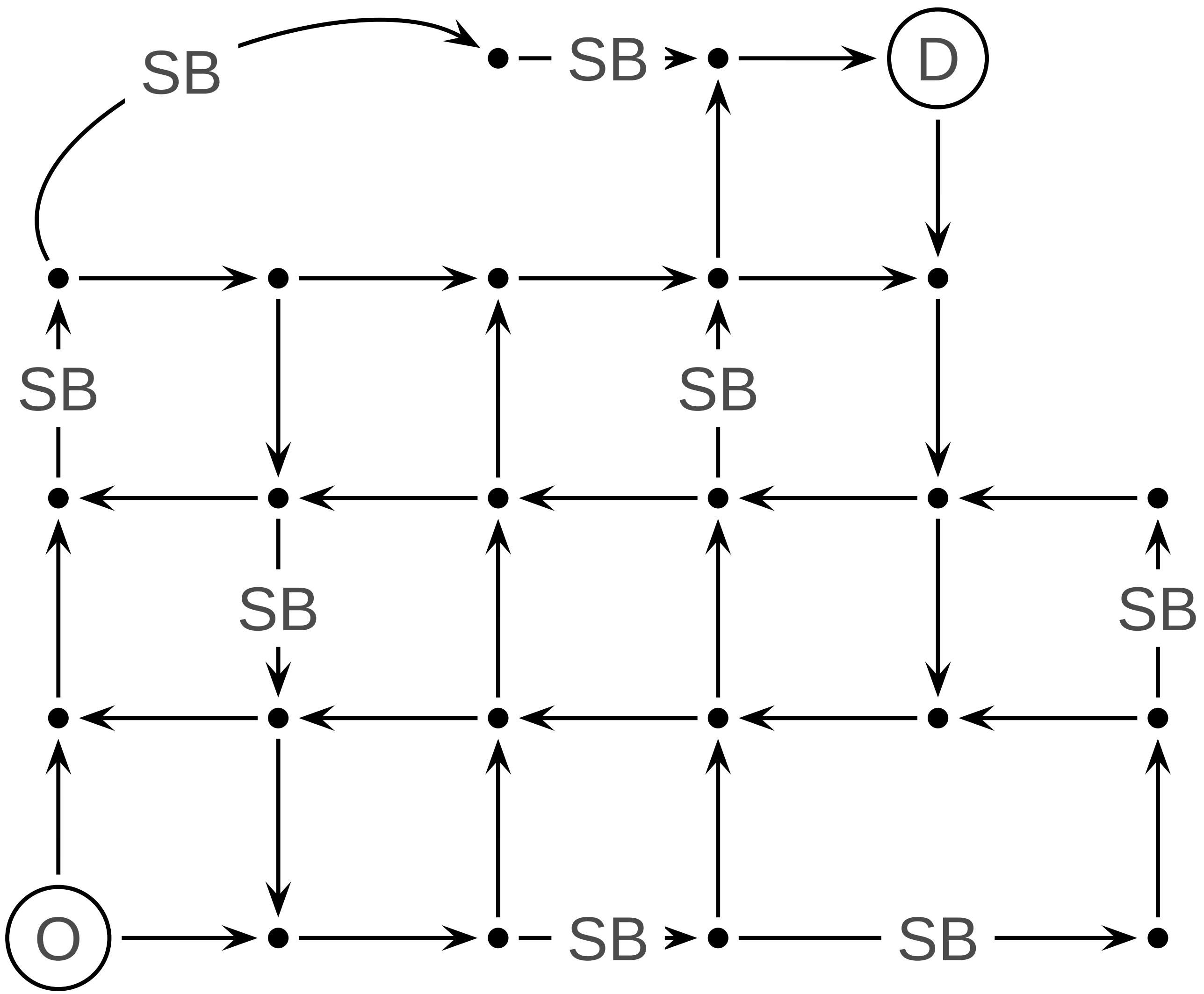
Preliminary Numerical Results

Sampling correction	MNL without	MNL with
$\widehat{\beta}_{\text{L}}$ (-0.6)	-0.203	-0.286
Scaled estimate	-0.600	-0.600
Robust std.	0.0193	0.019
Robust t-test	-10.53	-15.01
$\widehat{\beta}_{\text{SB}}$ (-0.3)	-0.0194	-0.143
Scaled estimate	-0.0573	-0.300
Robust std.	0.0662	0.0661
Robust t-test	-0.29	-2.17
Null log-likelihood	-1069.453	-1633.501
Final log-likelihood	-788.42	-759.848
Adjusted $\bar{\rho}^2$	0.261	0.288
BIOGEME has been used for all model estimations.		

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Preliminary Numerical Results



Preliminary Numerical Results

- True model: probit (Burrell, 1968)
  $U_\ell = \beta_{\mathsf{L}}\ \text{length}_\ell + \beta_{\mathsf{SB}}\ \text{nbspeedbumps}_\ell + \sigma\sqrt{L_\ell}\nu_\ell$
  $\beta_{\mathsf{L}} = -0.6$ and $\beta_{\mathsf{SB}} = -0.4$
  ν_ℓ 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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Preliminary Numerical Results

Sampling correction	MNL without	MNL with	PSL without	PSL with
$\widehat{\beta}_{\mathrm{L}}$ (-0.6)	-0.627	-0.978	-0.619	-0.969
Scaled estimate	-0.600	-0.600	-0.600	-0.600
Robust std.	0.0397	0.032	0.0407	0.0358
Robust t-test	-15.79	-30.57	-15.22	-27.04
$\widehat{\beta}_{\mathrm{SB}}$ (-0.4)	-0.0822	-0.0801	-0.347	-0.461
Scaled estimate	-0.0787	-0.0491	-0.336	-0.285
Robust std.	0.052	0.0559	0.182	0.158
Robust t-test	-1.58	-1.43	-1.90	-2.92
$\widehat{\beta}_{\mathrm{PS}}$			1.17	1.74
Scaled estimate			1.13	1.08
Robust std.			0.788	0.705
Robust t-test			1.49	2.47

Preliminary Numerical Results

Sampling correction	MNL without	MNL with	PSL without	PSL with
Null log-likelihood	-988.63	-2769.959	-988.63	-2769.959
Final log-likelihood	-676.111	-653.396	-674.481	-649.268
Adjusted $\bar{\rho}^2$	0.314	0.337	0.315	0.340
BIOGEME has been used for all model estimations.				

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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