Random Sampling of Alternatives in a Route Choice Context

Emma Frejinger

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

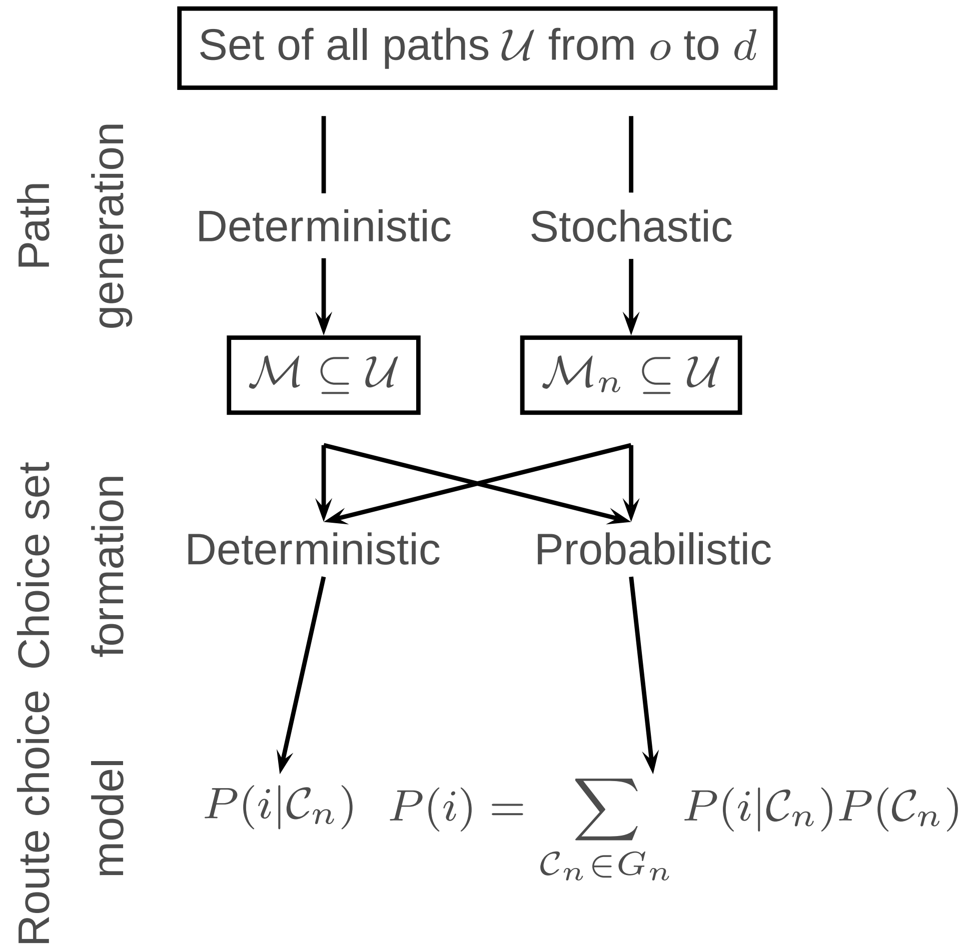
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

- Introduction to choice set generation
- Sampling of alternatives
- Stochastic path generation
- Derivation of sampling correction
- Numerical results
- Conclusions and future work

Introduction



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):

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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

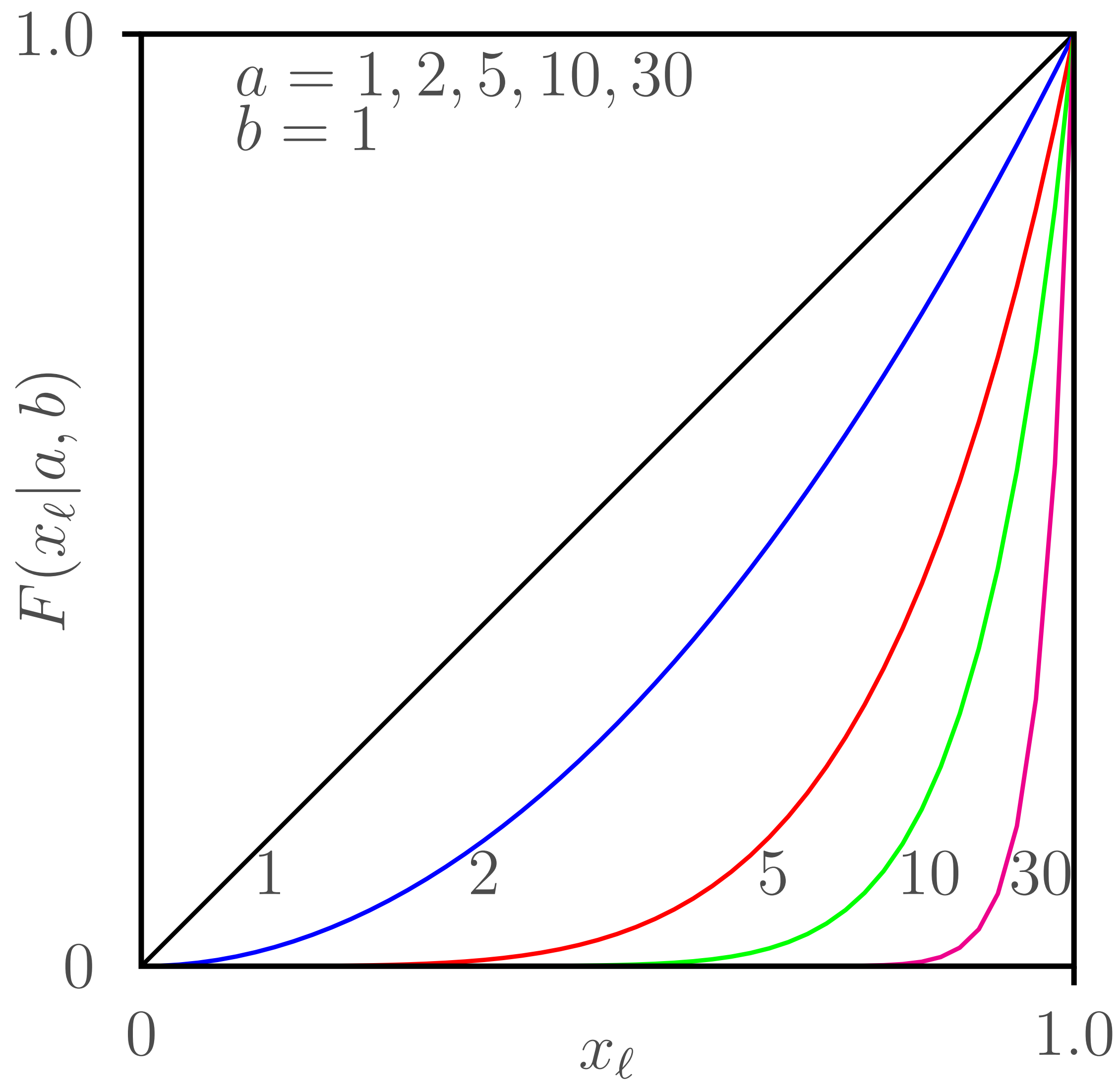
- Flexible approach that can be combined with various algorithms, here a biased random walk approach
- The probability of a link ℓ with source node v and sink node w is modeled in a stochastic way based on its distance to the shortest path
- Kumaraswamy distribution, cumulative distribution function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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Stochastic Path Enumeration



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

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

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

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

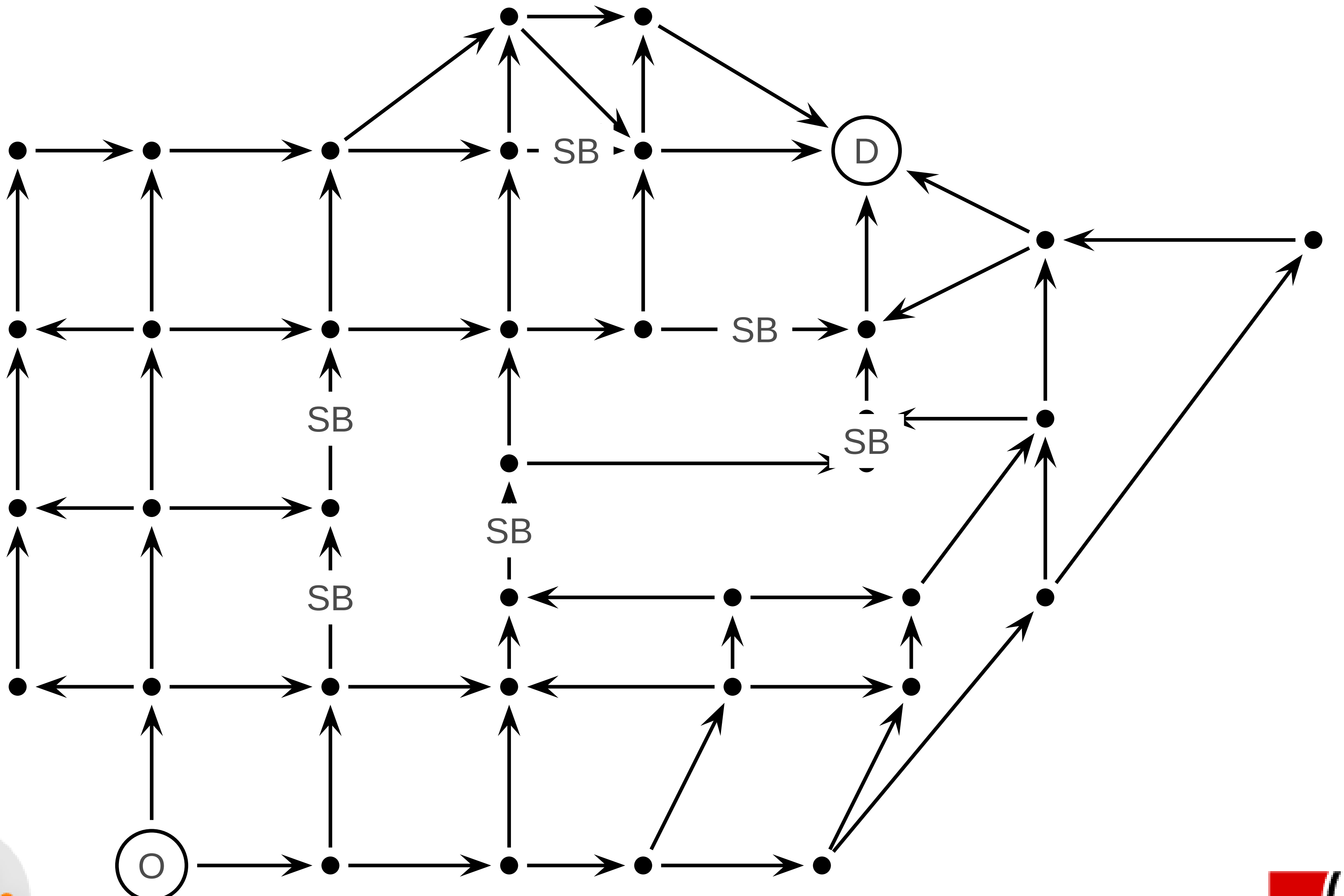
TRANSP-OR

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

  $U_j = \beta_{\text{PS}} \ln \text{PS}_j^{\mathcal{U}} + \beta_{\text{L}} \text{Length}_j + \beta_{SB} \text{SpeedBumps}_j + \varepsilon_j$

  $\beta_{\text{PS}} = 1$, $\beta_{\text{L}} = -0.3$, $\beta_{\text{SB}} = -0.1$

  ε_j distributed Extreme Value with scale 1 and location 0

  $\text{PS}_j^{\mathcal{U}} = \sum_{\ell \in \Gamma_j} \frac{L_\ell}{L_j} \frac{1}{\sum_{p \in \mathcal{U}} \delta_{\ell p}}$

- 3000 observations

Numerical Results

- Four model specifications

		Sampling Correction	
		Without	With
Path Size	$\mathcal{C}$	$M_{PS(\mathcal{C})}^{\text{NoCorr}}$	$M_{PS(\mathcal{C})}^{\text{Corr}}$
	$\mathcal{U}$	$M_{PS(\mathcal{U})}^{\text{NoCorr}}$	$M_{PS(\mathcal{U})}^{\text{Corr}}$

$$\text{PS}_i^{\mathcal{U}} = \sum_{\ell \in \Gamma_i} \frac{L_\ell}{L_i} \frac{1}{\sum_{j \in \mathcal{U}} \delta_{\ell j}}$$

$$\text{PS}_{in}^{\mathcal{C}} = \sum_{\ell \in \Gamma_i} \frac{L_\ell}{L_i} \frac{1}{\sum_{j \in \mathcal{C}_n} \delta_{\ell j}}$$

Numerical Results

- Model $M_{PS(\mathcal{C})}^{\text{NoCorr}}$:

$$V_{in} = \mu\left(\beta_{\text{PS}} \ln \text{PS}_{in}^{\mathcal{C}} - 0.3\text{Length}_i + \beta_{SB}\text{SpeedBumps}_i\right)$$

- Model $M_{PS(\mathcal{C})}^{\text{Corr}}$:

$$V_{in} = \mu\left(\beta_{\text{PS}} \ln \text{PS}_{in}^{\mathcal{C}} - 0.3\text{Length}_i + \beta_{SB}\text{SpeedBumps}_i + \ln(\frac{k_i}{q(i)})\right)$$

- Model $M_{PS(\mathcal{U})}^{\text{NoCorr}}$:

$$V_{in} = \mu\left(\beta_{\text{PS}} \ln \text{PS}_{in}^{\mathcal{U}} - 0.3\text{Length}_i + \beta_{SB}\text{SpeedBumps}_i\right)$$

- Model $M_{PS(\mathcal{U})}^{\text{Corr}}$:

$$V_{in} = \mu\left(\beta_{\text{PS}} \ln \text{PS}_{in}^{\mathcal{U}} - 0.3\text{Length}_i + \beta_{SB}\text{SpeedBumps}_i + \ln(\frac{k_i}{q(i)})\right)$$

Numerical Results

	True PSL	$M^{\text{NoCorr}}_{PS(\mathcal{C})}$ PSL	$M^{\text{Corr}}_{PS(\mathcal{C})}$ PSL	$M^{\text{NoCorr}}_{PS(\mathcal{U})}$ PSL	$M^{\text{Corr}}_{PS(\mathcal{U})}$ PSL
$\widehat{\beta}_{\text{L}}$ fixed	**-0.3**	**-0.3**	**-0.3**	**-0.3**	**-0.3**
$\widehat{\mu}$	**1**	**0.182**	**0.724**	**0.141**	**0.994**
Standard error		0.0277	0.0226	0.0263	0.0286
t-test w.r.t. 1		-29.54	-12.21	-32.64	-0.2
$\widehat{\beta}_{\text{PS}}$	**1**	**1.94**	**0.411**	**-1.02**	**1.04**
Standard error		0.428	0.104	0.383	0.0474
t-test w.r.t. 1		2.20	-5.66	-5.27	0.84
$\widehat{\beta}_{\text{SB}}$	**-0.1**	**-1.91**	**-0.226**	**-2.82**	**-0.0867**
Standard error		0.25	0.0355	0.428	0.0238
t-test w.r.t. -0.1		-7.24	-3.55	-6.36	0.56

Numerical Results

	True PSL	$M^{\text{NoCorr}}_{PS(\mathcal{C})}$ PSL	$M^{\text{Corr}}_{PS(\mathcal{C})}$ PSL	$M^{\text{NoCorr}}_{PS(\mathcal{U})}$ PSL	$M^{\text{Corr}}_{PS(\mathcal{U})}$ PSL
Final Log-likelihood		-6660.45	-6082.53	-6666.82	-5933.98
Adj. Rho-square		0.018	0.103	0.017	0.125

Null Log-likelihood: -6784.96, 3000 observations

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

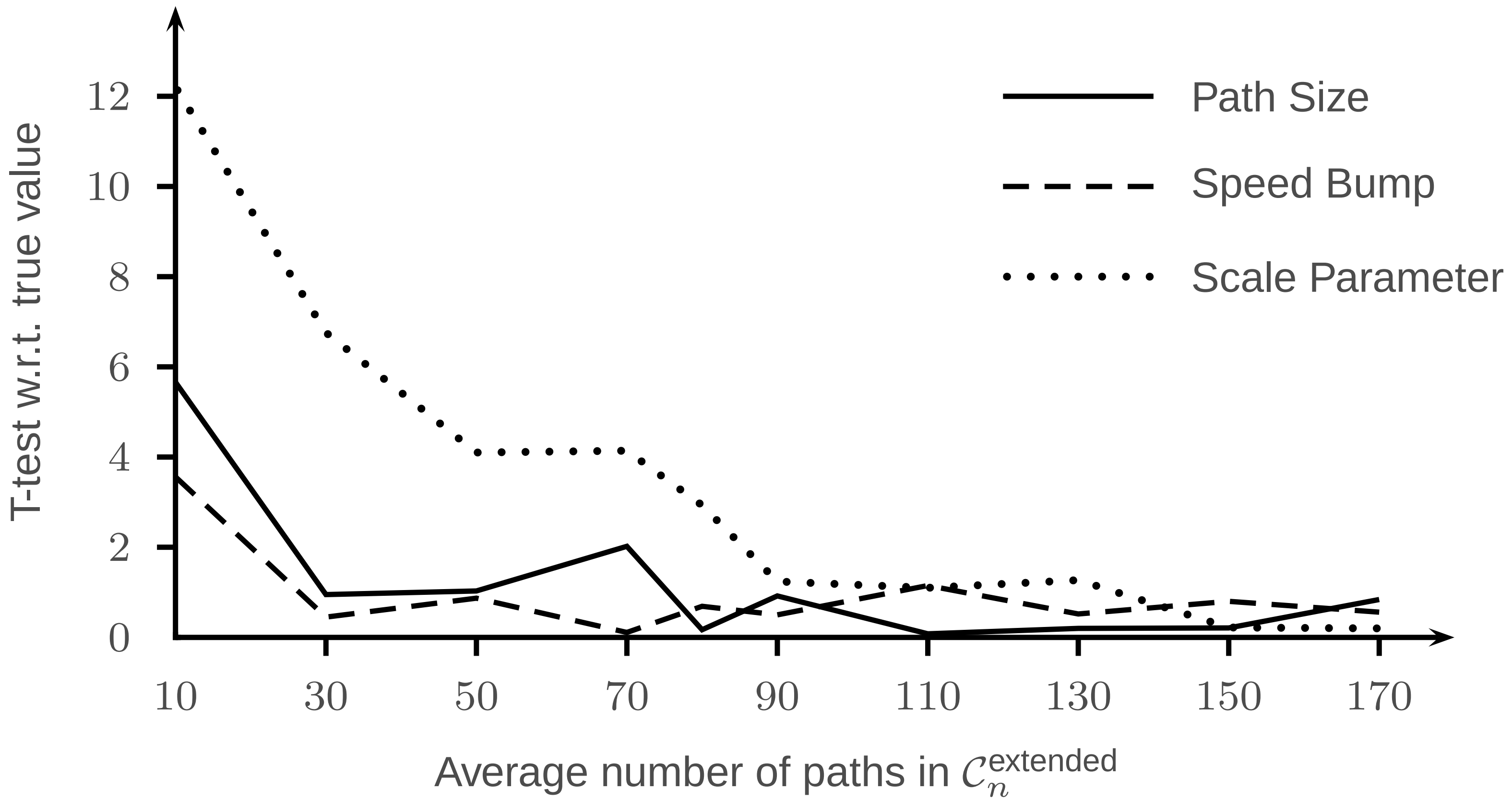
Average size of sampled choice sets: 9.66

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

Extended Path Size

- Compute Path Size attribute based on an *extended choice set* $\mathcal{C}_n^{\text{extended}}$
- Simple random draws from $\mathcal{U} \backslash \mathcal{C}_n$ so that $|\mathcal{C}_n| \leq |\mathcal{C}_n^{\text{extended}}| \leq |\mathcal{U}|$

Extended Path Size

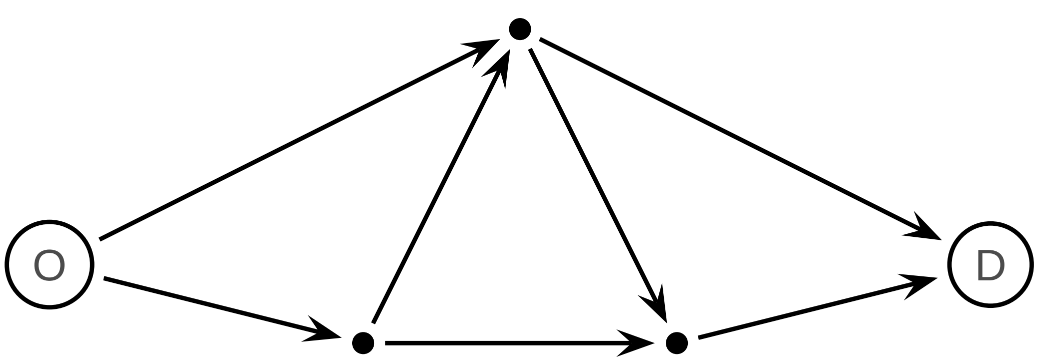


Extended Path Size

- Heuristic for finding an extended choice set $\mathcal{C}_n^{\text{extended}}$ (all paths in $\mathcal{C}_n$ are included)
- "Recursive gateway approach"
  - For each link in the network we generate a path
  - We count the number of times each link is used

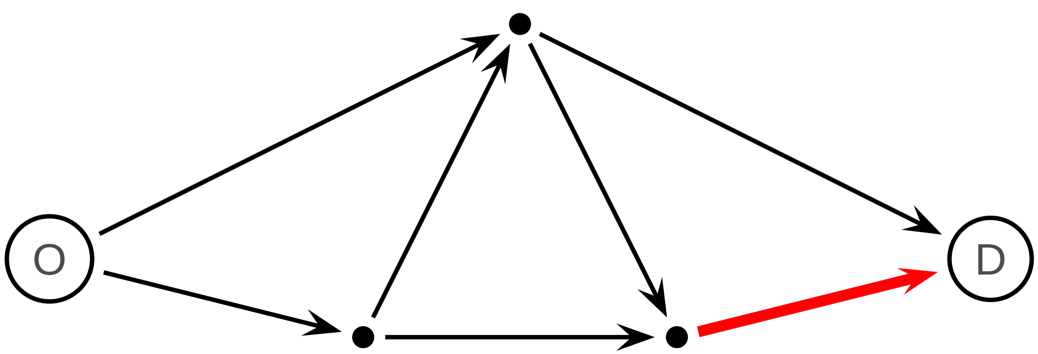
Extended Path Size

- Heuristic for finding an extended choice set $\mathcal{C}_n^{\text{extended}}$ (all paths in $\mathcal{C}_n$ are included)
- "Recursive gateway approach"
  - For each link in the network we generate a path
  - We count the number of times each link is used



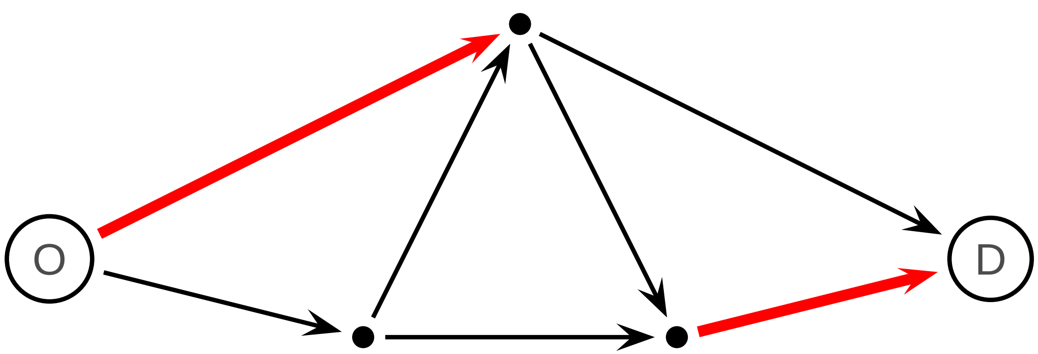
Extended Path Size

- Heuristic for finding an extended choice set $\mathcal{C}_n^{\text{extended}}$ (all paths in $\mathcal{C}_n$ are included)
- "Recursive gateway approach"
  - For each link in the network we generate a path
  - We count the number of times each link is used



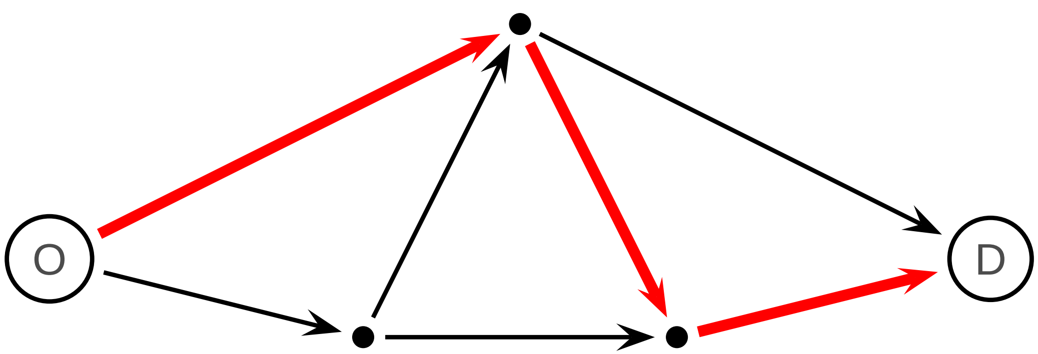
Extended Path Size

- Heuristic for finding an extended choice set $\mathcal{C}_n^{\text{extended}}$ (all paths in $\mathcal{C}_n$ are included)
- "Recursive gateway approach"
  - For each link in the network we generate a path
  - We count the number of times each link is used



Extended Path Size

- Heuristic for finding an extended choice set $\mathcal{C}_n^{\text{extended}}$ (all paths in $\mathcal{C}_n$ are included)
- "Recursive gateway approach"
  - For each link in the network we generate a path
  - We count the number of times each link is used



Extended Path Size

	True PSL	PS(C^{extended}) PSL	PS(C) PSL
$\widehat{\beta}_{\text{L}}$ fixed	**-0.3**	**-0.3**	**-0.3**
$\widehat{\mu}$	**1**	**0.885**	**0.724**
Standard error		0.0259	0.0266
t-test w.r.t. 1		-4.43	-12.21
$\widehat{\beta}_{\text{PS}}$	**1**	**1.52**	**0.411**
Standard error		0.102	0.104
t-test w.r.t. 1		5.10	-5.66
$\widehat{\beta}_{\text{SB}}$	**-0.1**	**-0.131**	**-0.266**
Standard error		0.0281	0.0355
t-test w.r.t. -0.1		-1.10	-3.55
Adj. Rho-Squared		0.114	0.103
Final Log-likelihood		-6006.96	-6082.53

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
  Heuristic for computing an approximation is proposed
- Numerical results are very promising