Choice set generation for route choice models using a sampling approach

Emma Frejinger and Michel Bierlaire

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

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

Introduction

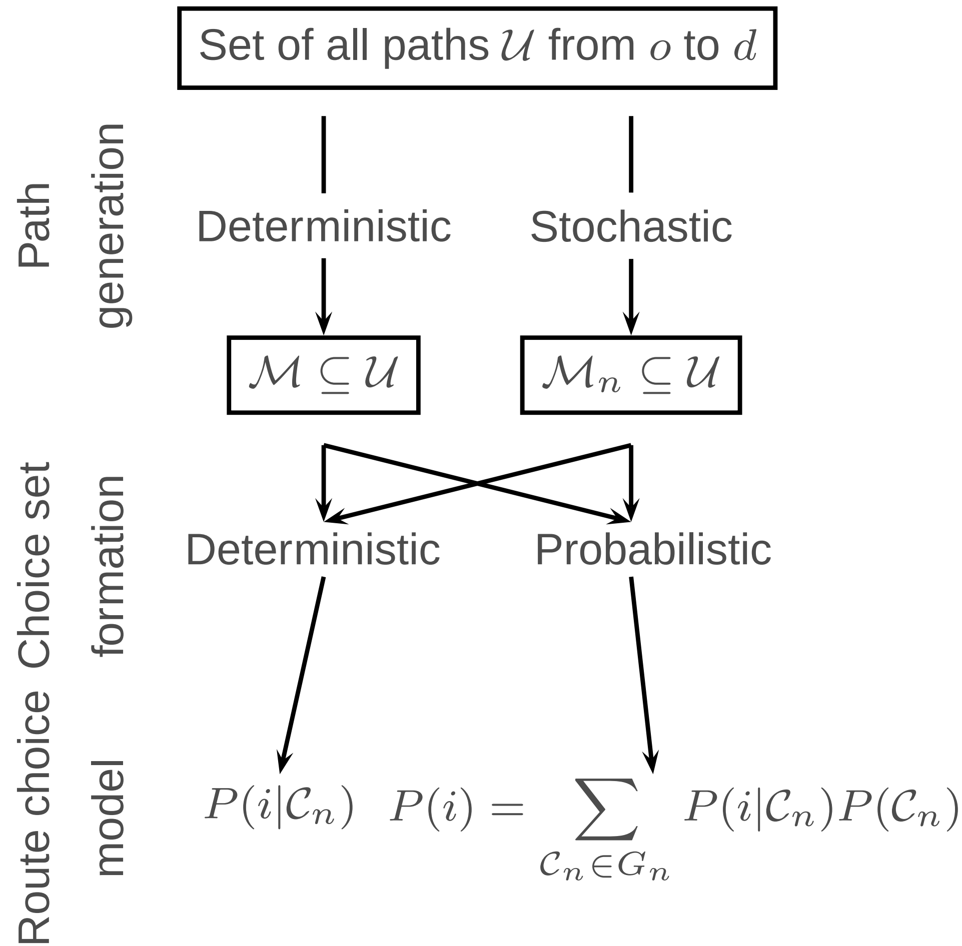
- Route choice problem

  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



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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

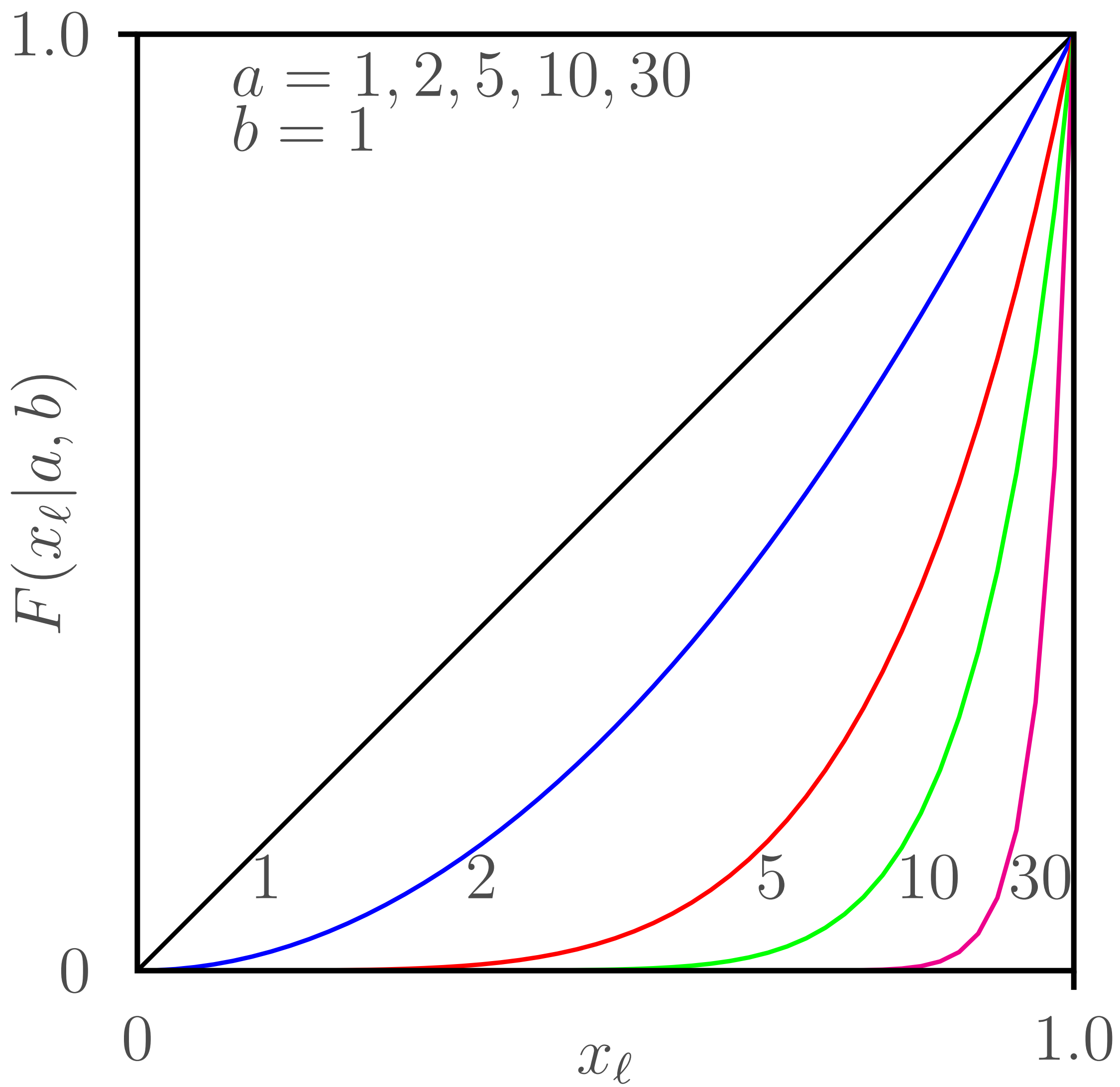
- Flexible approach that can be combined with various algorithms, here a biased random walk approach
- The probability of a link ℓ with source node v and sink node w is modeled in a stochastic way based on its distance to the shortest path
- Kumaraswamy distribution, cumulative distribution function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



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.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

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

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

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

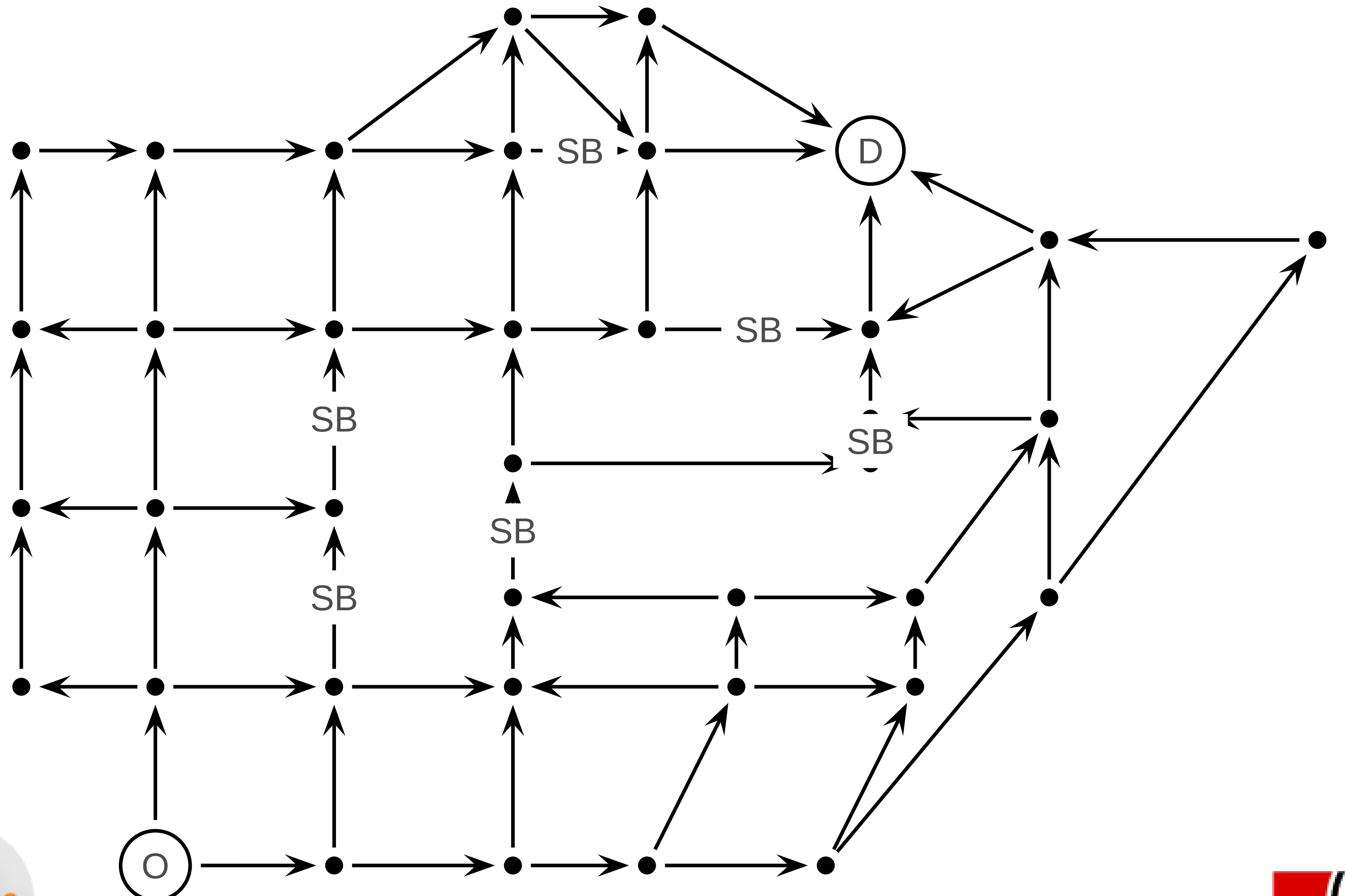
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

  $V_j = \beta_{\mathsf{PS}} \mathsf{PS}_j^{\mathcal{U}} + \beta_{\mathsf{L}} \mathsf{Length}_j + \beta_{SB} \mathsf{SpeedBumps}_j$

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

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

  $P(i|\mathcal{U}) = \frac{e^{V_i}}{\sum_{j \in \mathcal{U}} e^{V_j}}$

- 3000 observations

Numerical Results

- Four model specifications
  - Model $M^{\text{NoCorr}}_{PS(\mathcal{C})}$:
    $V_{in} = \beta_{\text{PS}}\text{PS}^{\mathcal{C}}_{in} + \beta_{\text{L}}\text{Length}_i + \beta_{SB}\text{SpeedBumps}_i$
  - Model $M^{\text{Corr}}_{PS(\mathcal{C})}$: $V_{in} =$
    $\beta_{\text{PS}}\text{PS}^{\mathcal{C}}_{in} + \beta_{\text{L}}\text{Length}_i + \beta_{SB}\text{SpeedBumps}_i + \ln(\frac{k_i}{q(i)})$
  - Model $M^{\text{NoCorr}}_{PS(\mathcal{U})}$:
    $V_i = \beta_{\text{PS}}\text{PS}^{\mathcal{U}}_{i} + \beta_{\text{L}}\text{Length}_i + \beta_{SB}\text{SpeedBumps}_i$
  - Model $M^{\text{Corr}}_{PS(\mathcal{U})}$: $V_j =$
    $\beta_{\text{PS}}\text{PS}^{\mathcal{U}}_{i} + \beta_{\text{L}}\text{Length}_i + \beta_{SB}\text{SpeedBumps}_i + \ln(\frac{k_i}{q(i)})$

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

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
$\widehat{\beta}_{\text{PS}}$	**1**	**0.363**	**0.443**	**-0.203**	**1.03**
Standard error		0.0729	0.086	0.0487	0.0465
t-test w.r.t. 1		-8.74	-6.48	-24.70	0.65
$\widehat{\beta}_{L}$	**-0.3**	**-0.0529**	**-0.326**	**-0.0453**	**-0.291**
Standard error		0.00864	0.0085	0.00828	0.00788
t-test w.r.t. -0.3		28.60	-3.06	30.76	1.14
$\widehat{\beta}_{\text{SB}}$	**-0.1**	**-0.345**	**-0.134**	**-0.404**	**-0.0773**
Standard error		0.0315	0.0259	0.0298	0.0258
t-test w.r.t. -0.1		-7.78	-1.31	-10.20	0.88

Numerical Results

	True PSL	$M_{PS(\mathcal{C})}^{\text{NoCorr}}$ PSL	$M_{PS(\mathcal{C})}^{\text{Corr}}$ PSL	$M_{PS(\mathcal{U})}^{\text{NoCorr}}$ PSL	$M_{PS(\mathcal{U})}^{\text{Corr}}$ PSL
Final Log-likelihood		-6596.22	-6047.14	-6598.46	-5840.80
Adj. rho square		0.02	0.10	0.02	0.13

Null Log-likelihood: -6719.733, 3000 observations

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

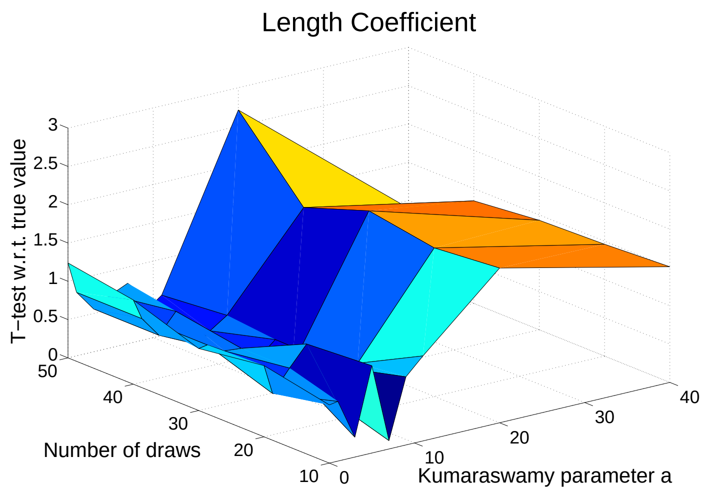
Average size of sampled choice sets: 9.43

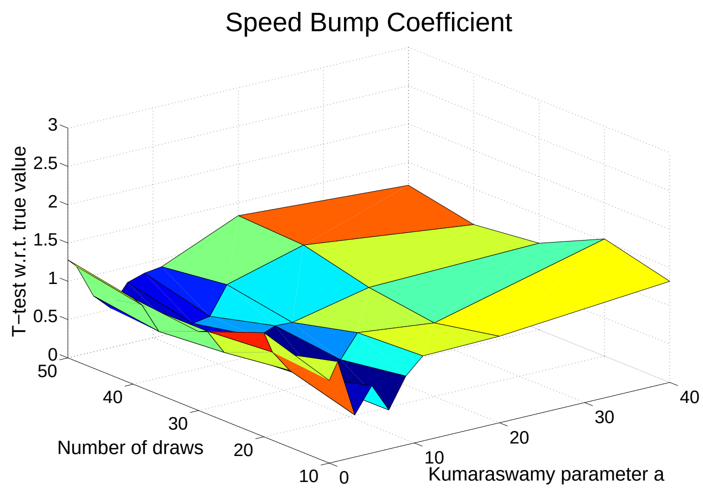
BIOGEME (`biogeme.epfl.ch`) has been used for all model estimations

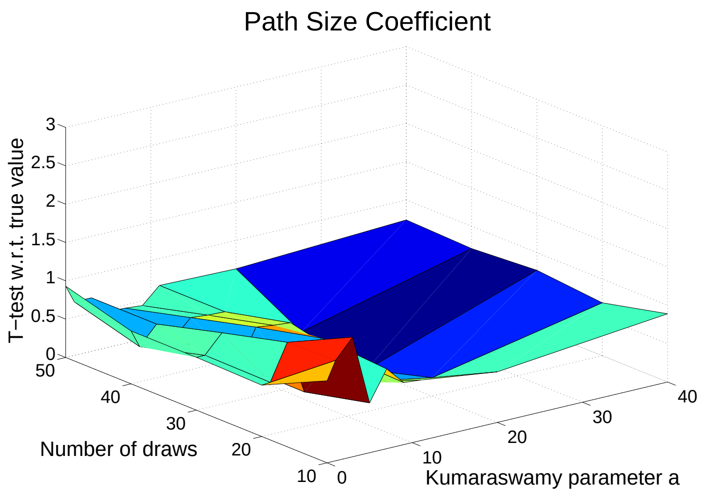
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Numerical Results



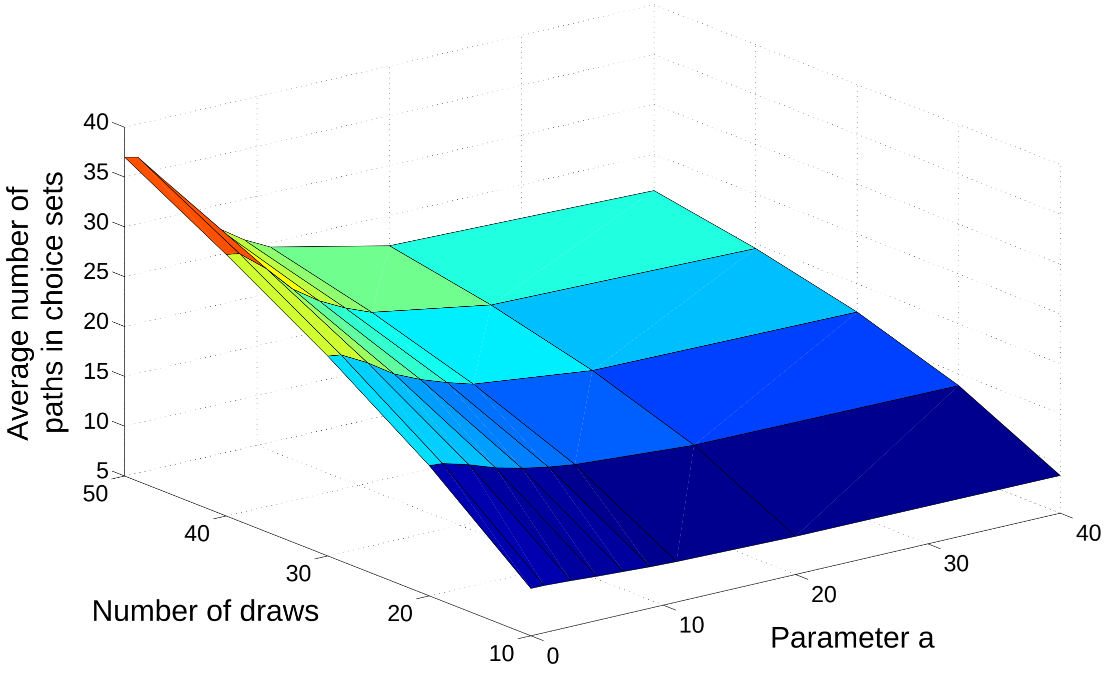




TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Numerical Results



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