

Bayesian Estimation of Large Choice Models with Adaptive Sampling and Likelihood-Free Inference

Anne-Valérie Preto, Xinwei Li, Prateek Bansal, Michel Bierlaire

Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne (EPFL)

Behavioural Computational Science (BeCoS) Lab
National University of Singapore (NUS)

September 30, 2026



Outline

- 1 Motivation
- 2 Bayesian Inference: Adaptive Sampling
- 3 TuRBOLFI
- 4 Case study



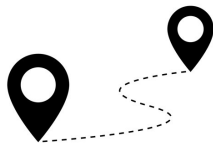
When enumeration becomes impossible

Consumer choice



A famous coffee chain has 383 billion unique latte combinations
[Merritt, 2023]

Route choice



Number of paths explodes with network size
Loops make enumeration intractable

Activity-based models

- Activity type
- Location
- Timing
- Duration
- Mode
- Route
- Joint decisions
- Rescheduling
- ...

Current approaches: The disconnection problem

Sampling of alternatives

Generate a fixed subset $\mathcal{D} \subset \mathcal{C}$ *before* estimation, then apply McFadden's correction factor [McFadden, 1978]:

$$\pi_n(i \mid \mathcal{D}) = \frac{\exp(\mu V_{in} + \ln(\pi_n(\mathcal{D} \mid i)))}{\sum_{j \in \mathcal{D}} \exp(\mu V_{jn} + \ln(\pi_n(\mathcal{D} \mid j)))}$$

- Mostly developed for MLE framework ([Guevara and Ben-Akiva, 2013a],[Guevara and Ben-Akiva, 2013b]); recently validated in Bayesian Inference [Dekker et al., 2025]
- In order to have a reliable, efficient estimation, we still need a large subset
- Sampling distribution is fixed independently of the model being estimated
- **Generation and estimation remain disconnected**

Outline

- 1 Motivation
- 2 Bayesian Inference: Adaptive Sampling
- 3 TuRBOLFI
- 4 Case study



The intuition behind adaptive sampling

1. Logit model with sampling of alternatives

Simple behavioral model with McFadden's correction factor.

$$P(i \mid \mathcal{D}, \theta)$$



2. Exchange idea

Find a generator such that the intractable normalizing constant cancels.

$$j^t \sim q(j \mid \phi_t)$$

$$q(j \mid \phi_t) = \Pr(j^t \mid \mathcal{C}, i; \phi_t)$$

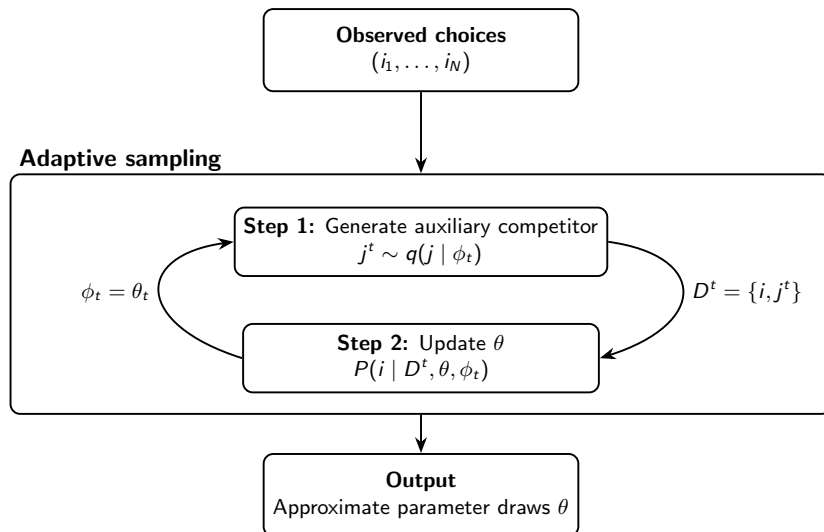


3. Binary comparison

Compare the chosen alternative with one relevant competitor.

$$\mathcal{D}^t = \{i, j^t\}$$

Proposed approach: Adaptive Sampling Procedure



Adaptive Sampling

Only two alternatives are involved

$$P_n(i_n \mid \{i_n, j_n^t\}, \theta, \phi_t) \approx \frac{\exp\{V_{i_n}(\theta) + V_{j_n^t}(\phi_t)\}}{\exp\{V_{i_n}(\theta) + V_{j_n^t}(\phi_t)\} + \exp\{V_{j_n^t}(\theta) + V_{i_n}(\phi_t)\}}.$$

Warning

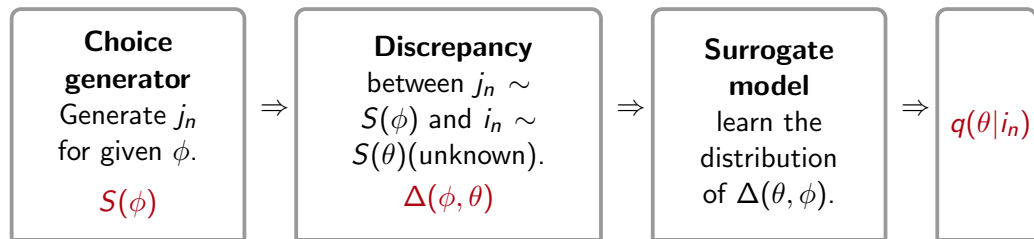
This is an adaptive approximate Bayesian update based on binary comparisons. We change the generator as θ_t changes.

Outline

- 1 Motivation
- 2 Bayesian Inference: Adaptive Sampling
- 3 TuRBOLFI**
- 4 Case study



What does likelihood-free inference do?

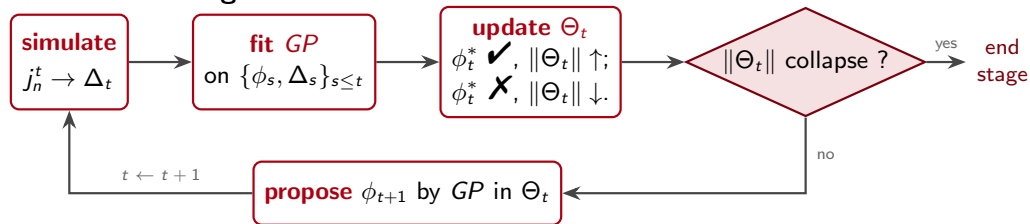


TuRBOLFI—**T**rusted-region **B**ayesian **o**ptimization **l**ikelihood-free **i**nference

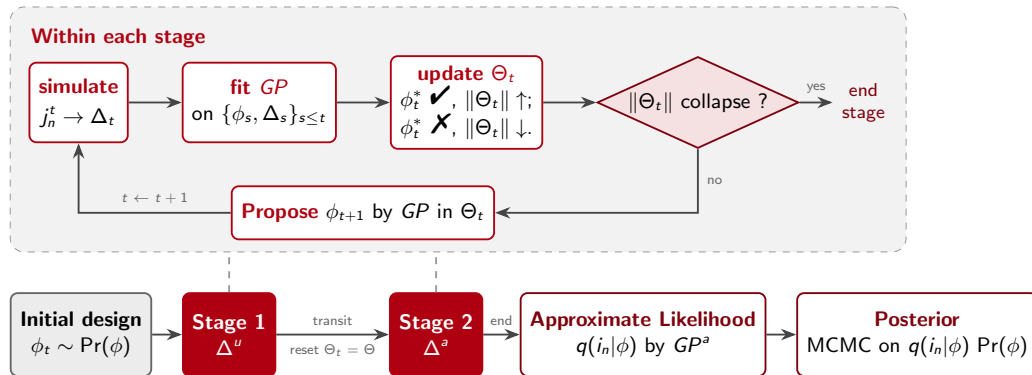
- Propose ϕ by Bayesian optimization (BO) under Gaussian Process surrogate $GP(\phi)$.
- two-stage** discrepancy: Δ^u and Δ^a .
- trust-region searching and stop rule.

TuRBOLFI: flow chart I

Within each stage



TuRBOLFI: flow chart II



TuRBOLFI: coarse-to-fine discrepancies

Stage 1: utility discrepancy Δ^u

to locate the promising Θ_t

$$\Delta_n^u(\theta, \phi) = \mu V_{j_n}(\phi) - \mu V_{i_n}(\phi)$$

$$\Delta^u(\theta, \phi) = \sum_{n=1}^N \Delta_n^u(\theta, \phi)$$

- Only includes **pairwise** utility difference;
- $\Delta^u(\theta, \phi)$ is **convex** over ϕ .

Stage 2: attribute discrepancy Δ^a

to refine the estimate precision

$$\Delta^a(\theta, \phi) = (T_\theta - T_\phi)^T W (T_\theta - T_\phi),$$

$$T_\phi = \sum_{n=1}^N x_{j_n}, \quad W = \text{Var}^{-1}(T_\phi^*).$$

- Only includes **pairwise** attribute vector;
- T is a **sufficient summary statistic** for MNL.

Outline

- 1 Motivation
- 2 Bayesian Inference: Adaptive Sampling
- 3 TuRBOLFI
- 4 Case study



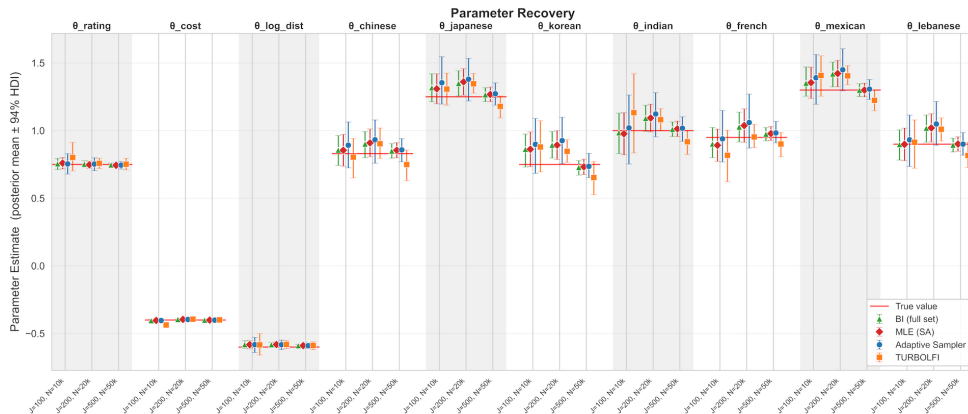
Synthetic restaurant choice model

Dataset

- Synthetic data similar to [Bierlaire and Paschalidis, 2023] and [Bierlaire and Krueger, 2024]
- Full choice set $\mathcal{C}$: $J \in \{100, 200, 500\}$ restaurants
- $N \in \{10k, 20k, 50k\}$ individuals
- Observed choices generated synthetically with a logit model

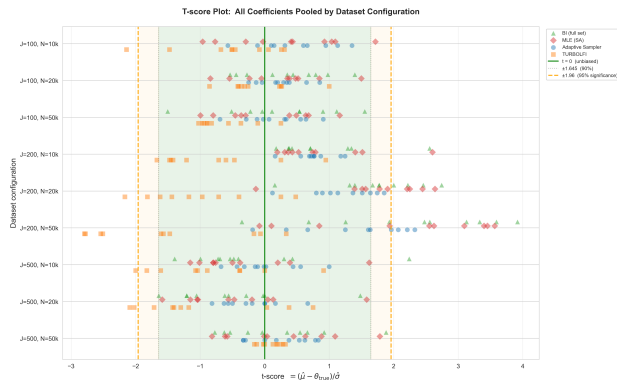
$$V_{jn} = \theta_{\text{cost}} \text{price}_j + \theta_{\text{rate}} \text{rating}_j + \theta_{\log d} \log(\text{dist}_{jn}) + \sum_{c \in \mathcal{C}_{\text{cuisine}}} \theta_c \mathbb{1}\{j \text{ is cuisine } c\}$$

Comparison of Estimation Methods



Both AS and TuRBOLFI recover the parameters accurately, even though they never enumerate the choice set.

Comparison of Estimation Methods



- Comparison of significance of parameter estimates against the true value

$$t_{score} = \frac{\hat{\theta} - \theta_{true}}{\hat{\sigma}}$$

- Under misspecification the benchmark methods give too narrow confidence interval

Comparison of Estimation Methods

Config. (J, N)	Wall time (mm:ss)				RAM (GB)			
	BI	MLE-SA	AS	TuRBOLFI	BI	MLE-SA	AS	TuRBOLFI
100, 10k	10:49	2:35	3:38	42:28	121.6	2.60	0.42	11.22
200, 20k	38:35	6:30	6:19	1:02:97	349.3	9.39	0.67	15.23
500, 50k	3:17:40	48:04	14:50	1:53:14	~800	54.54	2.43	29.11

Adaptive Sampling (AS) maintains low memory footprint across all configurations. MLE with 40% sampling (MLE-SA) shows significant memory growth with problem size. BI (Bayesian Inference with full choice set) has prohibitive RAM requirements.

Conclusion: inference without enumeration

Adaptive Sampling

- Approximate, exchange-inspired inference
- Each observed choice is compared with a simulated competing alternative
- Low computational cost

Limitation: Performance can deteriorate with dominating alternative.

TuRBOLFI

- Likelihood-free inference through a surrogate model
- Promising robustness under model misspecification
- Naturally extendable to mixed logit models

Limitation: Higher simulation cost and longer wall-clock time.

We assume we can always simulate an alternative

Next steps

Empirical validation

- Apply the methods to real choice data with large or implicit choice sets.
- Characterize how utility concentration affect adaptive sampling.

Methodological extensions

- Connect the adaptive binary-comparison framework with likelihood-based estimation.
- Extend TuRBOLFI to mixed logit models with random preference heterogeneity.

Thank you!

Code and reproducibility



[https://github.com/transp-or/
TooManyAlternativesToHandle](https://github.com/transp-or/TooManyAlternativesToHandle)



Questions?

`anne-valerie.preto@epfl.ch`

`xinwei.li@u.nus.edu`

Outline

5 Appendix

6 Bibliography

Step 1: Generation model

Sample competitor j^t

- Draw from generation model with parameter ϕ_t :

$$\begin{aligned} \Pr(j^t \mid \mathcal{C}; \phi_t) \\ = \frac{\exp(\mu V_j(\phi_t))}{\sum_{k \in \mathcal{C}} \exp(\mu V_k(\phi_t))} \end{aligned}$$

- Use accept-reject methods, no computation of the denominator

Multiple-Try Metropolis (MTM)

Procedure for each individual

- Draw m candidates $\{j_1, \dots, j_m\}$
- Assign weights $w(j) \propto \exp(V_{jn})$
- Select j^* with probability $\propto w(j^*)$
- Accept using MTM acc. ratio
- Higher acceptance than single-try MH

Warning

We're in an approximate setting, since we never wait for convergence in this step.

Step 2: Why the binary likelihood simplifies

We want to draw from

$$\Pr(\theta \mid \mathcal{D}_{ij}^t, i; \phi_t) \propto \Pr(i \mid \mathcal{D}_{ij}^t, \theta, \phi_t) \cdot \Pr(\theta)$$

- General Bayesian form $\rightarrow$ pairwise logit.
- Both i and j^t drawn from same logit denominator over $\mathcal{C}$: cancels in ratio.

$$\begin{aligned} P(i \mid \mathcal{D}_{ij}^t, \theta, \phi_t) &= \frac{\Pr(\mathcal{D}_{ij}^t \mid i; \phi_t) \Pr(i \mid \theta)}{\Pr(\mathcal{D}_{ij}^t \mid i; \phi_t) \Pr(i \mid \theta) + \Pr(\mathcal{D}_{ij}^t \mid j; \phi_t) \Pr(j \mid \theta)} \\ &\approx \frac{\exp(V_j(\phi_t)) \exp(V_i(\theta))}{\exp(V_j(\phi_t)) \exp(V_i(\theta)) + \exp(V_i(\phi_t)) \exp(V_j(\theta))} \end{aligned}$$

Comparison of Estimation Methods

Configuration	Estimator	Cov. (%)	RMSE	MAE	Rel. Width
$J = 100, N = 10k$	BI (Full)	100.0	0.0483	0.0352	0.1783
	MLE-SA	100.0	0.0489	0.0363	0.1893
	Adaptive Sampling	100.0	0.0686	0.0495	0.3021
	TuRBOLFI	90.0	0.0641	0.0467	0.3572
$J = 200, N = 20k$	BI (Full)	60.0	0.0883	0.0741	0.1461
	MLE-SA	60.0	0.0920	0.0779	0.1571
	Adaptive Sampling	90.0	0.1142	0.0964	0.2519
	TuRBOLFI	90.0	0.1065	0.0896	0.2961
$J = 500, N = 50k$	BI (Full)	100.0	0.0138	0.0113	0.0775
	MLE-SA	100.0	0.0154	0.0121	0.0815
	Adaptive Sampling	100.0	0.0176	0.0139	0.1279
	TuRBOLFI	100.0	0.0194	0.0159	0.3793

Outline

5 Appendix

6 Bibliography

Bibliography I



Bierlaire, M. and Krueger, R. (2024).

Sampling and discrete choice.

In Hess, S. and Daly, A., editors, *Handbook of Choice Modelling*, pages 693–719, Cheltenham, UK. Edward Elgar Publishing.



Bierlaire, M. and Paschalidis, E. (2023).

Estimating MEV models with samples of alternatives.

Technical Report TRANSP-OR 231225, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.



Dekker, T., Bansal, P., and Huo, J. (2025).

Revisiting McFadden's correction factor for sampling of alternatives in multinomial logit and mixed multinomial logit models.

Transportation Research Part B: Methodological, 192:103129.

Bibliography II

-  Guevara, C. A. and Ben-Akiva, M. (2013a).
Sampling of alternatives in multivariate extreme value (MEV) models.
Transportation Research Part B, 48:31– 52.
-  Guevara, C. A. and Ben-Akiva, M. E. (2013b).
Sampling of alternatives in logit mixture models.
Transportation Research Part B: Methodological, 58:185–198.
-  McFadden, D. (1978).
Modelling the choice of residential location.
In A. Karlquist *et al.*, editor, *Spatial interaction theory and residential location*,
pages 75–96, Amsterdam. North-Holland.

Bibliography III



Merritt, M. (2023).

Starbucks has 383 billion drink combinations and wants to make them faster.

Accessed: 2025-06-09.