Choice with many alternatives

Michel Bierlaire and Anne-Valérie Preto

Transport and Mobility Laboratory, EPFL
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

September 23, 2026

Outline

Choice model as an optimization problem

Travel demand: activity based models

Combinatorial choices

Prediction

Estimation

Predicting choice behavior

Decision rule

Homo economicus

Rational and narrowly self-interested economic actor who is optimizing her outcome

Behavioral assumptions

- The decision maker solves an optimization problem.
- The analyst needs to define
  - the decision variables,
  - the objective function,
  - the constraints.

Continuous case: classical microeconomics

Optimization problem

$$\max_q \widetilde{U}(q;\theta)$$

subject to

$$p^T q \leqslant I, \; q \geqslant 0.$$

Demand function

- Solution of the optimization problem.
- KKT optimality conditions:

$$q^* = f(I, p; \theta)$$

Discrete choices

How does it work for discrete choices?

Utility maximization

Optimization problem

$$\max_{q,w} \widetilde{U}(q, w; \theta)$$

subject to

$$\begin{aligned} & p^T q + c^T w \leqslant I \\ & \textstyle\sum_j w_j = 1 \\ & w_j \in \{0, 1\}, \forall j. \end{aligned}$$

where $c^T = (c_1, \ldots, c_j, \ldots, c_J)$ contains the cost of each alternative.

Derivation of demand functions

- Mixed integer optimization problem
- No optimality condition
- Impossible to derive demand functions directly

Derivation of the demand functions

Step 1: condition on the choice of the discrete good

- Fix the discrete good(s), that is select a feasible w.
- Derive the conditional demand functions from KKT.

Step 2: enumerate all alternatives

- Enumerate all alternatives.
- Compute the conditional indirect utility function U_i.
- Select the alternative with the highest U_i.

Enumerate all alternatives ??????

Starbucks has 383 billion unique latte combinations, [Merritt, 2023]

Activity-based models

- Activity participation
- Activity type
- Activity location
- Activity timing
- Activity duration
- Activity scheduling
- Activity frequency
- Travel mode choice
- Route choice
- Departure time choice
- Trip chaining / Tour formation
- Vehicle usage
- Parking choice
- Joint activity participation
- Ride-sharing / Carpooling decision
- Household resource allocation
- Teleworking decision
- Trip cancellation or rescheduling
- Use of on-demand mobility services
- ... and many more

Outline

Choice model as an optimization problem

Travel demand: activity based models

Combinatorial choices

Prediction

Estimation

Activities

Why do people travel?

- Most of the time, not for the sake of it.
- Activities.
- Spread in space and time.

Activity-based models: literature

Econometric models

- Discrete choice models.
- Curse of dimensionality.
- Decomposition: sequence of choices
  - Activity pattern
  - Primary tour: time of day
  - Primary tour: destination and mode
  - Secondary tour: time of day
  - Secondary tour: destination and mode
  - e.g. [Bowman and Ben-Akiva, 2001]

Rule-based models

- If the selected activity is at location L,
- and the travel time from current location C to L exceeds T_{max},
- then reject the activity–location combination,
- unless it is a high-utility or infrequent activity (e.g., doctor appointment).
- e.g. [Arentze et al., 2000]

Research question: can we combine the two?

	Econometric	Rule-based
Micro-economic theory	X	—
Parameters inference	X	—
Testing/validation	X	—
Joint decisions	—	X
Complex rules	—	X
Complex constraints	—	X

Outline

Choice model as an optimization problem

Travel demand: activity based models

Combinatorial choices

Prediction

Estimation

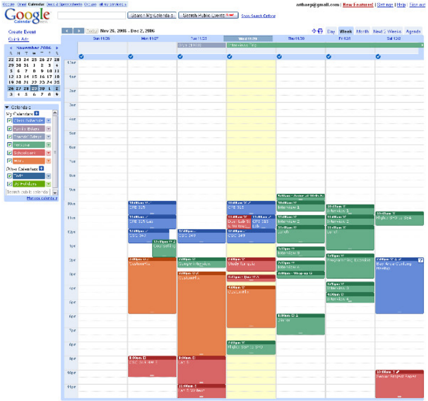
Combinatorial choices

Mathematical optimization

- Each individual is solving a combinatorial optimization problem.
- Decisions: see the long list before...
- Objective function: utility (to be maximized).
- Constraints: complex rules.

[Pougala et al., 2022]

Example: activity schedule



The context

- Given a list of potential activities,
- with preferred starting time and duration,
- identify a feasible schedule,
- that maximizes utility.

Example: activity schedule

Decision variables

- Activity participation: $\phi_a \in \{0, 1\}$,
- Activity scheduling: $\phi_{ab} \in \{0, 1\}$,
- Activity start time: $s_a \in \mathbb{R}$,
- Activity duration: $\tau_a \in \mathbb{R}$,

Schedule utility

$$U_S = \sum_{a=0}^{A} \phi_a \left(U_a^{\text{participation}} + U_a^{\text{start time}} + U_a^{\text{duration}} + \sum_{b=0}^{A} \phi_{ab} U_{a,b}^{\text{travel}} \right)$$

[Pougala et al., 2022]

Example: utility schedule

Utility components

$$U_a^{\text{duration}} = \theta_a^{\text{short}} \max(0, \tau_a^* - \tau_a) + \theta_a^{\text{long}} \max(0, \tau_a - \tau_a^*) + \xi_{\text{duration}}$$

$$U_a^{\text{starting}} = \theta_a^{\text{early}} \max(0, s_a^* - s_a) + \theta_a^{\text{late}} \max(0, s_a - s_a^*) + \xi_{\text{starting}}$$

[Pougala et al., 2022]

Example: utility schedule

Constraints

$$
\begin{aligned}
T &= 24\text{h}, \\
\sum_a \phi_a \tau_a &= T, \\
x_b &\geqslant x_a + \tau_a + d_{ab} - T(1 - \phi_{ab}) && \forall a, b, \\
x_b &\leqslant x_a + \tau_a + d_{ab} + T(1 - \phi_{ab}) && \forall a, b,
\end{aligned}
$$

and many others...

[Pougala et al., 2022]

Combinatorial choice model

Decision variables

$$\phi \in \{0, 1\}^K$$

Total: 2^K combinations.

Choice set: defined by constraints

$$\mathcal{C}_n = \{\phi \mid g_n(\phi) \leqslant 0, h_n(\phi) = 0\}.$$

Combinatorial choice model

Utility

$$U_{\phi,n} = V_n(\phi, x_n, \xi_n; \theta) + \nu_{\phi,n},$$

where

- $\nu_{\phi,n}$ are i.i.d. extreme value,
- ξ_n is a random vector, capturing
  - correlation among alternatives,
  - taste heterogeneity,
  - etc.
- θ is a vector of unknown parameters, to be estimated from data.

Combinatorial choice model

Mixture of logit models

$$P_n(\phi | x_n, \xi_n; \theta) = \frac{e^{U_{\phi,n}}}{\sum_{\ell \in \mathcal{C}_n} e^{U_{\ell,n}}},$$

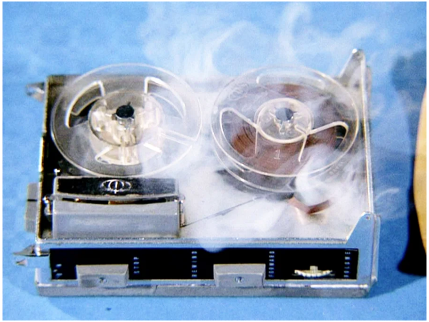
$$P_n(\phi | x_n; \theta) = \int_{\xi} P_n(\phi | x_n, \xi_n; \theta) d\xi.$$

Main challenge

- Enumeration of $\mathcal{C}_n$ is impossible.
- Calculating the probability is impossible.

MISSION: IMPOSSIBLE

Challenges



Your mission, should you choose to accept it

- Use the model for prediction.
- Estimate its parameters from data.

Outline

Choice model as an optimization problem

Travel demand: activity based models

Combinatorial choices

Prediction

Estimation

Use the model

Idea

- We cannot calculate the probability.
- But we do not really need it, do we?
- In order to identify the chosen alternative, we rely on simulation.
- Actually, simulation would be necessary anyway for the integral.

Methodology

- Draw independent realizations ξ_{nr} of ξ_n.
- Draw the chosen alternative from the logit model

$$\pi(\phi) = P_n(\phi | x_n, \xi_{nr}; \theta).$$

- Metropolis-Hastings uses only the numerator $e^{U_{\phi,n}}$.

Metropolis–Hastings: general idea

Goal

- We want to generate draws from a complicated target distribution $\pi(\cdot)$.
- Direct sampling is impossible.
- Metropolis–Hastings constructs a Markov chain whose stationary distribution is $\pi(\cdot)$.
- Markov chain Monte-Carlo method.

Algorithm ingredients

- Target distribution: $\pi(\phi) \propto e^{U_{\phi,n}}$.
- Proposal distribution: $q(\phi'|\phi)$, easy to sample from.

Metropolis–Hastings: Steps

Iteration $t \to t+1$

- Current state: $\phi^{(t)}$.
- **Step 1.** Draw a candidate ϕ' from $q(\cdot|\phi^{(t)})$.
- **Step 2.** Compute the acceptance probability

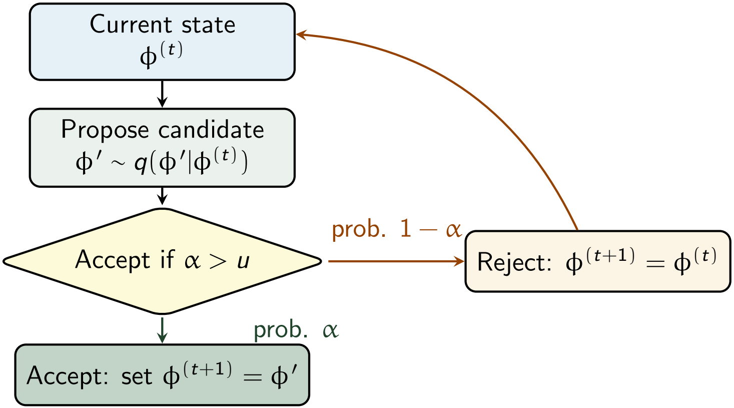
$$\alpha = \min\left(1, \frac{\pi(\phi')\, q(\phi^{(t)}|\phi')}{\pi(\phi^{(t)})\, q(\phi'|\phi^{(t)})}\right) = \min\left(1, \frac{e^{U_{\phi',n}}\, q(\phi^{(t)}|\phi')}{e^{U_{\phi^{(t)},n}}\, q(\phi'|\phi^{(t)})}\right).$$

- **Step 3.** With probability α, accept and set $\phi^{(t+1)} = \phi'$. Otherwise, reject and keep $\phi^{(t+1)} = \phi^{(t)}$.

Key property

- The chain has stationary distribution $\pi(\cdot)$.
- Here, $\pi(\phi) \propto e^{U_{\phi,n}}$, so only the numerator of the logit formula is needed.

Metropolis–Hastings: Illustration



Acceptance probability: $\alpha = \min\left(1, \dfrac{e^{U_{\phi',n}}\, q(\phi^{(t)}|\phi')}{e^{U_{\phi^{(t)},n}}\, q(\phi'|\phi^{(t)})}\right).$

Metropolis–Hastings

Is it magic?

- Key to success: well-designed proposal distribution $q(\cdot)$.
- Smart exploration of the choice set.
- Must exploit its structure to generate high probability alternatives.

Optimization-based choice problem

- By design, an optimization approach generates “good” alternatives.
- Main challenge: ergodicity.
- It means that the Markov chain must potentially cover the whole choice set.

[Flötteröd and Bierlaire, 2013], [Pougala et al., 2023]

Challenges

Your mission, should you choose to accept it

- ✓ Use the model for prediction.
- ❏ Estimate its parameters from data.

Outline

Choice model as an optimization problem

Travel demand: activity based models

Combinatorial choices

Prediction

Estimation

Parameters inference

Research question

- Given the values of the parameters, we can simulate the chosen alternative.
- What about the opposite?
- Given observed choices, we want to infer the value of the parameters θ.

Too many alternatives

Choice model 9:41

Do we really need every alternative to learn the trade-off parameters?

Not necessarily. Dominated alternatives do not tell much.
9:43

Then consider only plausible competitors.

But how do we identify those competitors?
9:45

One competing alternative

Choice model 9:46

By simulating from the choice model.

But we do not know the choice model yet.

9:48

Then use the best approximation of the model available at the current iteration.

9:49

Method 1: Bayesian Inference Adaptive Sampling

Sampling of alternatives: [McFadden, 1978]

- Assign to each observation n a subset D_n of alternatives.
- $\pi_n(D|i)$ the probability to generate subset D, when i is observed.
- Contribution to the likelihood of observation n:

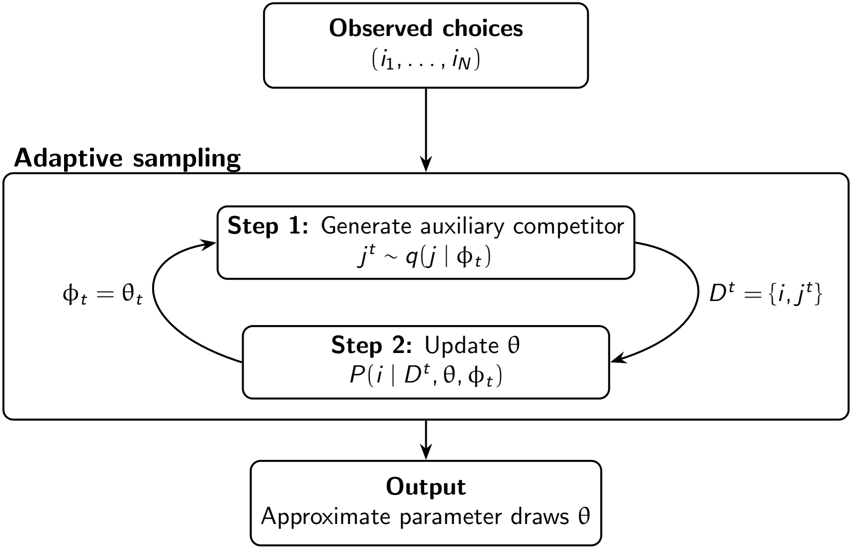
$$\pi_n(i|D) = \frac{\pi_n(D|i)P_n(i)}{\sum_{j\in D}\pi_n(D|j)P_n(j)} \quad \text{[Bayes]}$$

- For logit

$$\pi_n(i|D) = \frac{e^{V_{in}+\ln \pi_n(D|i)}}{\sum_{j\in D} e^{V_{jn}+\ln \pi_n(D|j)}}.$$

- No more sum over $\mathcal{C}_n$.
- It works also for Bayesian estimation [Dekker et al., 2025].

Adaptive approximate Bayesian estimation



Bayesian Inference Adaptive Sampling

Embedded in the Gibbs sampler

- Draw alternative j_n^t using the current parameter estimates ϕ_t.
- Include its generation probability in the parameter update (McFadden's correction).

Only two alternatives are involved

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

Method 2: Maximum likelihood with approximated gradient

Monte Carlo approximation of the full-choice-set score

- The likelihood gradient contains an expectation over the whole choice set.

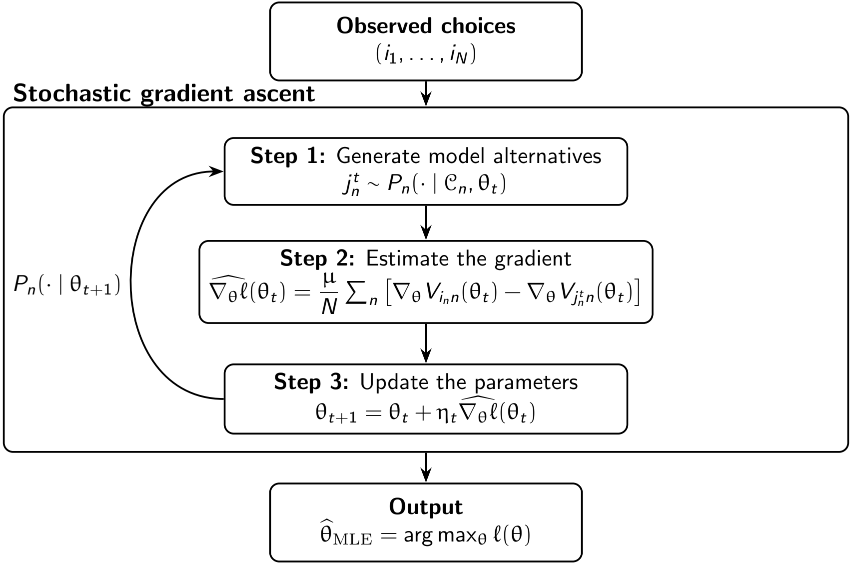
$$\nabla_\theta \bar{\ell}(\theta_t) = \frac{\mu}{N} \sum_{n=1}^{N} \left[\underbrace{\nabla_\theta V_{i_n n}(\theta_t)}_{\text{observed behaviour}} - \underbrace{\mathbb{E}_{j_n \sim P_n(\cdot|\mathcal{C}_n,\theta_t)} \left[\nabla_\theta V_{j_n n}(\theta_t)\right]}_{\text{model-predicted behaviour}} \right].$$

- We do not need to calculate this expectation exactly.

Use the competitor j_n^t

$$\nabla_\theta \bar{\ell}(\theta_t) \approx \frac{\mu}{N} \sum_{n=1}^{N} \left[\nabla_\theta V_{i_n n}(\theta_t) - \nabla_\theta V_{j_n^t n}(\theta_t)\right]$$

Adaptive stochastic-gradient estimation



Case study

Restaurant choice model

- Synthetic dataset similar to [Bierlaire and Paschalidis, 2023] and [Bierlaire and Krueger, 2024]
- Full choice set $\mathcal{C}$: $J = 100, 200$ or 500 restaurants
- $N = 10'000, 20'000$ or $50'000$ individuals
- Observed choices generated synthetically with a logit model

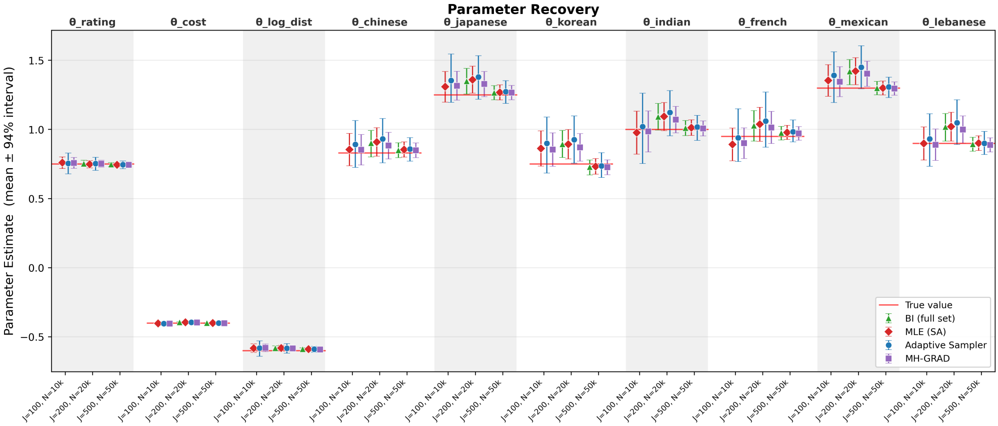
$$V_{jn} = \theta_{\text{cost}}\, \text{price}_j + \theta_{\text{rate}}\, \text{rating}_j + \theta_{\log d}\, \log(\text{dist}_{jn}) + \sum_{c \in \mathcal{C}_{\text{cuisine}}} \theta_c\, \mathbb{1}\{j \text{ is cuisine } c\}$$

Comparison of Estimation Methods

Configuration (J, N)	Wall time ([h:]mm:ss) BI	MLE-SA	BI-AS	MH-GRAD
100, 10k	10:49	2:35	3:38	0:01.80
200, 20k	38:35	6:30	6:19	0:03.86
500, 50k	3:17:40	48:04	14:50	0:10.69

Both adaptive sampling (BI-AS) and the stochastic gradient method (MH-GRAD) avoid enumerating alternatives and maintain a low memory footprint.
Maximum likelihood estimation with sampling of alternatives (MLE-SA) uses 40% of the choice set.
BI, Bayesian inference with the full choice set, has prohibitive RAM requirements (around 250 GB for the largest configuration).

Comparison of Estimation Methods



The stochastic gradient method (MH-GRAD) recovers parameters as accurately as Bayesian Inference with the full choice set, and as Maximum Likelihood Estimation with 40% uniform sampling of alternatives, but with much lower computational cost.

Case study: international migration

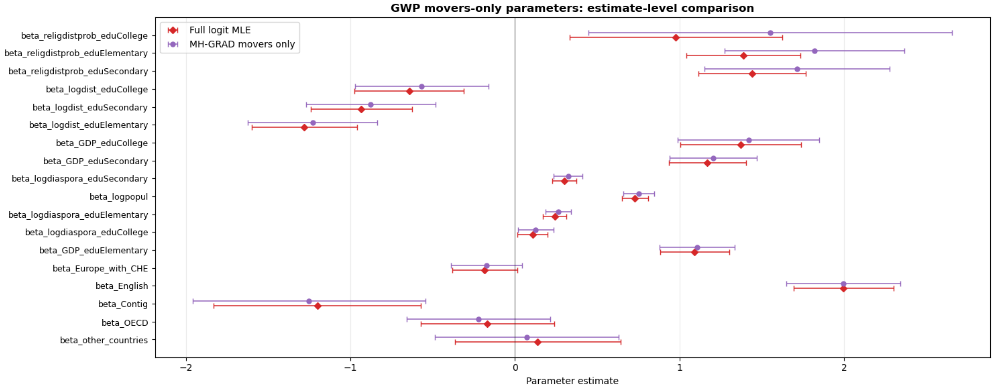
Gallup World Poll

- Real migration choices: stay in the origin country or migrate to a destination country.
- Model specification based on [Beine et al., 2021].
- We use the simpler multinomial logit version.

Objective

- Apply MH-GRAD to a real large-scale choice model.
- Compare its estimates with the full-choice-set logit MLE.

Comparison of Estimation Methods



MH-GRAD completed the optimization in 0.4 s (movers only), compared with 165.1 s for full logit MLE, respectively.

Summary

- Large scale choice models.
- Numerical solution of the utility maximization problem.
- Impossibility to enumerate alternatives in the choice set.
- Main idea: rely on Markov chain Monte-Carlo methods, both for prediction and estimation.
- First results are particularly encouraging.
- Real data: MH-GRAD might have some difficulties when one alternative (stay) dominates the dataset
- Other parameters: scale, nests, etc.

Challenges

Your mission, should you choose to accept it

- ✓ Use the model for prediction.
- ✓ Estimate its parameters from data.

This presentation will self-destruct in 5 seconds.

We are grateful to Prateek Bansal for helpful discussions that contributed to this work.

Bibliography I

- Arentze, T., Hofman, F., van Mourik, H., and Timmermans, H. (2000).
  ALBATROSS: Multiagent, rule-based model of activity pattern decisions.
  Transportation Research Record: Journal of the Transportation Research Board, 1706(1):136–144.

- Beine, M., Bierlaire, M., and Docquier, F. (2021).
  New york, abu dhabi, london or stay at home? using a cross-nested logit model to identify complex substitution patterns in migration.
  Technical Report TRANSP-OR 210130, Lausanne, Switzerland.

- 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.

Bibliography II

- 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.
- Bowman, J. L. and Ben-Akiva, M. E. (2001).
  Activity-based disaggregate travel demand model system with activity schedules.
  Transportation Research Part A: Policy and Practice, 35(1):1–28.
- 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 III

Flötteröd, G. and Bierlaire, M. (2013).
Metropolis-Hastings sampling of paths.
Transportation Research Part B: Methodological, 48:53–66.
Accepted on Oct 17, 2012.

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 IV

Merritt, M. (2023).
Starbucks has 383 billion drink combinations and wants to make them faster.
https://www.morningbrew.com/daily/stories/starbucks-has-383-billion-drink-combinations.
Accessed: 2025-06-09.

Pougala, J., Hillel, T., and Bierlaire, M. (2022).
Capturing trade-offs between daily scheduling choices.
Journal of Choice Modelling, 43(100354).
Accepted on Mar 19, 2022.

Pougala, J., Hillel, T., and Bierlaire, M. (2023).
OASIS: Optimisation-based activity scheduling with integrated simultaneous choice dimensions.
Transportation Research Part C: Emerging Technologies, 155.

Bibliography V