Travel demand models: dealing with the curse of dimensionality

Michel Bierlaire and Anne-Valérie Preto

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

May 7, 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

$$p^T q \leq I, \; q \geq 0.$$

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

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

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

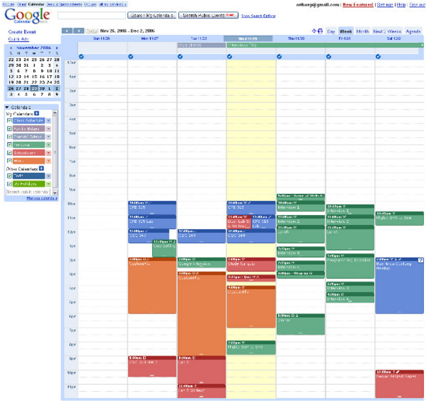
Combinatorial choices

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

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

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

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

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

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

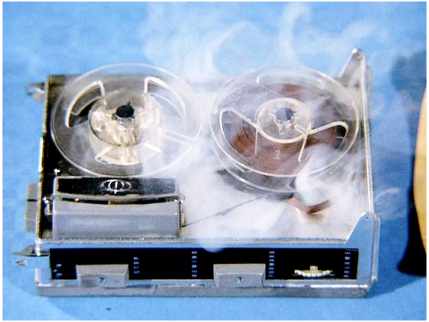
$$P_n(\phi|x_n; \theta) = \int_{\xi} P_n(\phi|x_n, \xi_n; \theta) d\xi.$$

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

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

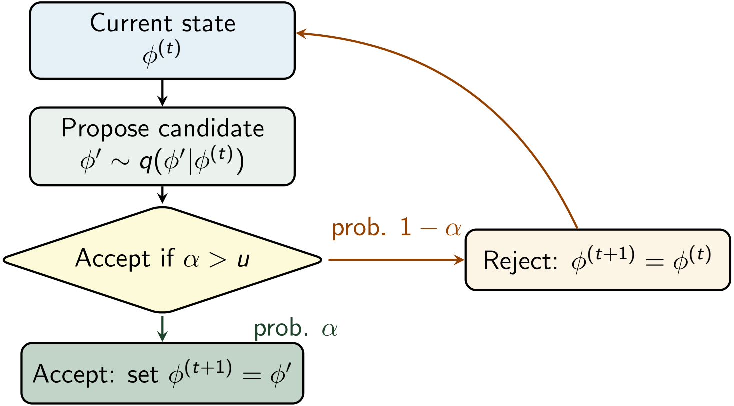
$$\alpha = \min\left(1, \frac{\pi(\phi')\, q(\phi^{(t)}|\phi')}{\pi(\phi^{(t)})\, q(\phi'|\phi^{(t)})}\right) = \min\left(1, \frac{e^{U_{\phi',n}}\, q(\phi^{(t)}|\phi')}{e^{U_{\phi^{(t)},n}}\, q(\phi'|\phi^{(t)})}\right).$$

- **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.

Sampling of alternatives

[McFadden, 1978]

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

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

- For logit

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

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

Proposed sampling protocol

Sample only one more alternative

- i_n is the chosen alternative for observation n.
- We need to sample a "competitor":

$$j_n \sim \Pr(j \mid \mathcal{C}_{-i}, i; \phi).$$

  where ϕ is a vector of values for the parameters.
- Define

$$\mathcal{D}_n = \{i_n, j_n\}.$$

- We now know how to do that without enumerating the choice set: Metropolis-Hastings.

Bayesian inference

We want to draw from

$$\Pr(\theta, \mathcal{D}_n | i_n; \phi) = \Pr(\theta | \mathcal{D}_n, i_n; \phi) \Pr(\mathcal{D}_n | i_n; \phi),$$

where

$$\Pr(\theta | \mathcal{D}_n, i_n; \phi) \propto \Pr(i_n | \mathcal{D}_n; \theta, \phi) \Pr(\theta).$$

Likelihood simplifies to

$$\Pr(i_n | \mathcal{D}_n; \theta, \phi) \approx \frac{\exp(V_{j_n}(\phi)) \exp(V_{i_n}(\theta))}{\exp(V_{j_n}(\phi)) \exp(V_{i_n}(\theta)) + \exp(V_{i_n}(\phi)) \exp(V_{j_n}(\theta))},$$

when $|\mathcal{C}|$ is large (see Appendix).

Adaptive sampling

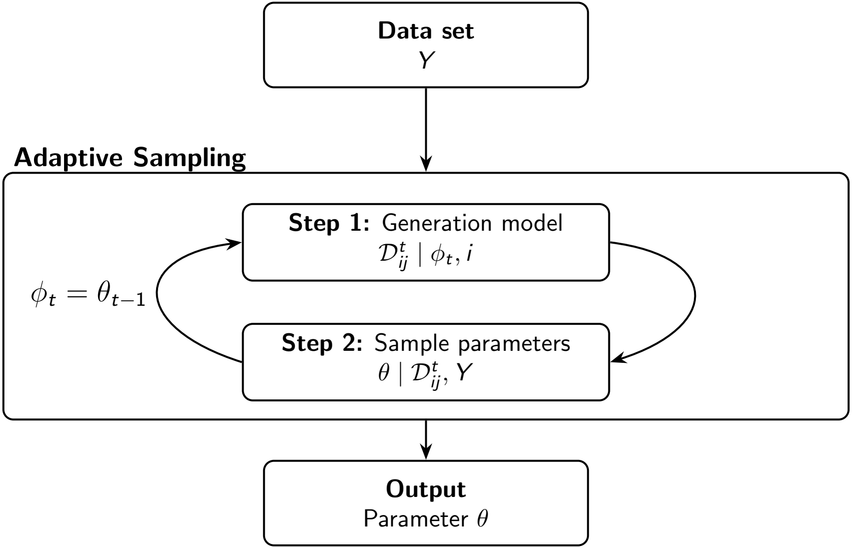
Issues with this methodology

- How do we know ϕ?
- How can two alternatives be representative of the full choice set?

Solution

- Integrate sampling into the simulation.
- We improve ϕ as we learn the parameters.
- The quality of the competing alternatives improves.

Choice of ϕ: Adaptive Sampling Procedure



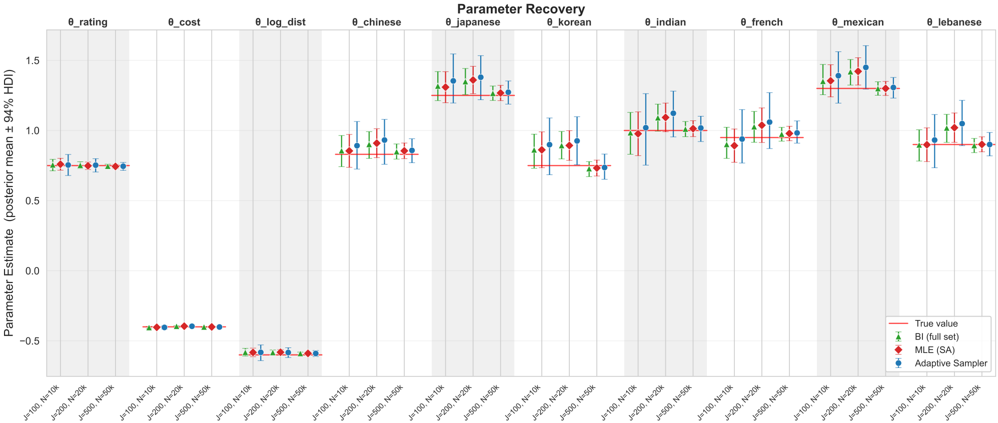
Case study

Restaurant choice model

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

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

Results comparison



The Adaptive Sampler 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.

Comparison of Estimation Methods

Configuration (J, N)	Wall time (mm:ss)			RAM (GB)		
	BI	MLE-SA	AS	BI	AS	MLE-SA
100, 10k	10:49	2:35	3:38	-	0.42	2.60
200, 20k	38:35	6:30	6:19	-	0.67	9.39
500, 50k	3:17:40	48:04	14:50	-	2.43	54.54

Adaptive Sampling (AS) maintains low memory footprint across all configurations. MLE with 40% sampling (MLE-SA) shows significant memory growth with problem size.
The Bayesian Inference with the full choice set is not shown here due to its prohibitive computational cost (more than 1h for the smallest configuration, and more than 3h for the largest one, with peak memory usage around 250 GB).

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.
- Results are particularly encouraging.

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.

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.

- Bierlaire, M. and Krueger, R. (2024).
  Sampling and discrete choice.
  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.

Bibliography II

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.

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

Bibliography III

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.

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.

Bibliography IV

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.

Detailed derivation

Let $\mathcal{C}_n$ be the choice set.

Logit model

For any $i \in \mathcal{C}_n$,

$$\Pr(i|\mathcal{C}_n, \theta) = \frac{\exp(V_{in}(\theta))}{\sum_{\ell \in \mathcal{C}_n} \exp(V_{\ell n}(\theta))} = \frac{\exp(V_{in}(\theta))}{Z(\mathcal{C}_n, \theta)}.$$

Competitor

Given the observed choice $i_n \in \mathcal{C}_n$, generate a competitor $j_n \in \mathcal{C}_{-i_n}$ with

$$\Pr(j_n|\mathcal{C}_n, i_n; \phi) = \frac{\exp(V_{j_n n}(\phi))}{\sum_{\ell \in \mathcal{C}_{-i_n}} \exp(V_{\ell n}(\phi))} = \frac{\exp(V_{j_n n}(\phi))}{Z(\mathcal{C}_{-i_n}, \phi)},$$

where $\mathcal{C}_{-i_n} = \mathcal{C}_n \setminus \{i_n\}$.

Likelihood simplification

Auxiliary binary set

Define

$$\mathcal{D}_n = \{i_n, j_n\},$$

where

$$\mathcal{D}_n \sim \Pr(\mathcal{D}_n | i_n; \phi), \qquad \Pr(\mathcal{D}_n | i_n; \phi) = \Pr(j_n | \mathcal{C}_n, i_n; \phi) = \frac{\exp(V_{j_n n}(\phi))}{Z(\mathcal{C}_{-i_n}, \phi)}.$$

Posterior update for individual n

We consider the joint distribution

$$\Pr(\theta, \mathcal{D}_n | i_n; \phi) = \Pr(\theta | \mathcal{D}_n, i_n; \phi) \Pr(\mathcal{D}_n | i_n; \phi).$$

Bayes rule

$$\Pr(\theta | \mathcal{D}_n, i_n; \phi) \propto \Pr(i_n | \mathcal{D}_n, \theta; \phi) \Pr(\theta).$$

Bayes rule again

$$\begin{aligned}\Pr(i_n | \mathcal{D}_n, \theta; \phi) &= \frac{\Pr(\mathcal{D}_n | i_n, \theta; \phi) \Pr(i_n | \theta)}{\Pr(\mathcal{D}_n | i_n, \theta; \phi) \Pr(i_n | \theta) + \Pr(\mathcal{D}_n | j_n, \theta; \phi) \Pr(j_n | \theta)} \\ &= \frac{\Pr(\mathcal{D}_n | i_n; \phi) \Pr(i_n | \theta)}{\Pr(\mathcal{D}_n | i_n; \phi) \Pr(i_n | \theta) + \Pr(\mathcal{D}_n | j_n; \phi) \Pr(j_n | \theta)},\end{aligned}$$

since $\mathcal{D}_n$ does not depend on θ.

Logit simplification

$$\Pr(i_n|\theta) = \frac{\exp(V_{i_n n}(\theta))}{Z(\mathcal{C}_n, \theta)}, \; \Pr(j_n|\theta) = \frac{\exp(V_{j_n n}(\theta))}{Z(\mathcal{C}_n, \theta)}.$$

$$\begin{aligned}\Pr(i_n|\mathcal{D}_n, \theta; \phi) &= \frac{Z(\mathcal{C}_n, \theta) \Pr(\mathcal{D}_n|i_n; \phi) \exp(V_{i_n n}(\theta))}{Z(\mathcal{C}_n, \theta)[\Pr(\mathcal{D}_n|i_n; \phi) \exp(V_{i_n n}(\theta)) + \Pr(\mathcal{D}_n|j_n; \phi) \exp(V_{j_n n}(\theta))]} \\ &= \frac{\Pr(\mathcal{D}_n|i_n; \phi) \exp(V_{i_n n}(\theta))}{\Pr(\mathcal{D}_n|i_n; \phi) \exp(V_{i_n n}(\theta)) + \Pr(\mathcal{D}_n|j_n; \phi) \exp(V_{j_n n}(\theta))}.\end{aligned}$$

Approximation

$$\Pr(i_n|\mathcal{D}_n,\theta;\phi) = \frac{\Pr(\mathcal{D}_n|i_n;\phi)\ \exp(V_{i_n n}(\theta))}{\Pr(\mathcal{D}_n|i_n;\phi)\ \exp(V_{i_n n}(\theta)) + \Pr(\mathcal{D}_n|j_n;\phi)\ \exp(V_{j_n n}(\theta))},$$

with

$$\Pr(\mathcal{D}_n|i_n;\phi) = \frac{\exp(V_{j_n n}(\phi))}{Z(\mathcal{C}_{-i_n},\phi)},\ \Pr(\mathcal{D}_n|j_n;\phi) = \frac{\exp(V_{i_n n}(\phi))}{Z(\mathcal{C}_{-j_n},\phi)}.$$

If $|\mathcal{C}_n|$ is large,

$$Z(\mathcal{C}_{-i_n},\phi) \approx Z(\mathcal{C}_{-j_n},\phi).$$

and therefore

$$\Pr(i_n|\mathcal{D}_n,\theta;\phi) \approx \frac{\exp(V_{j_n n}(\phi))\ \exp(V_{i_n n}(\theta))}{\exp(V_{j_n n}(\phi))\ \exp(V_{i_n n}(\theta)) + \exp(V_{i_n n}(\phi))\ \exp(V_{j_n n}(\theta))}.$$