IIDE
Institut interdisciplinaire du Développement de l'entreprise

Stochastic adaptive resampling for the estimation of discrete choice models

23rd Swiss Transport Research Conference
10–12 May 2023

Nicola Ortelli[1,2], Matthieu de Lapparent[1], Michel Bierlaire[2]

[1] IIDE, HEIG-VD
[2] TRANSP-OR, EPFL

Flashback

Faster estimation of discrete choice models via dataset reduction

22nd Swiss Transport Research Conference
18–20 May 2022

Nicola Ortelli[1,2], Matthieu de Lapparent[1], Michel Bierlaire[2]

[1] IIDE, HEIG-VD
[2] TRANSP-OR, EPFL

Flashback

Discrete choice models (DCMs)

What are DCMs?

- Suppose N observations, each containing:
  - a vector of explanatory variables x_n;
  - the observed choice i_n.
- A DCM calculates the choice probabilities as a function of x_n and θ:

$$P\left(i \,|\, x_n; \theta\right),$$

- where θ is a vector of model parameters.

Flashback

Estimating DCMs

Maximum likelihood estimation (MLE)

- Find θ so as to maximize the joint probability of the observed choices:

$$\max_{\theta} \mathcal{L}(\theta) = \max_{\theta} \sum_{n=1}^{N} \log P\left(i_n \,|\, x_n; \theta\right).$$

  - Solved using iterative methods — Newton, BFGS, etc.
  - Each iteration is $\mathcal{O}(N)$.
- MLE is burdensome for large datasets!

Flashback

Intuition

Factoring out redundancy

- If the data contains groups of **identical observations**:

$$\mathcal{L}(\theta) = \sum_{u=1}^{U} N_u \cdot \log P\left(i_u \mid x_u; \theta\right),$$

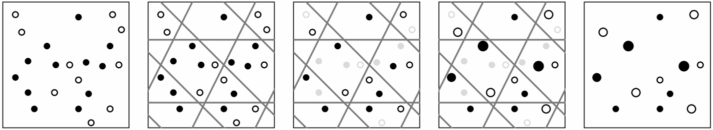
  - U unique observations.
  - Each appears N_u times in the original data.
- Can we extend this "factorization trick" to **nearly identical observations**?

Flashback

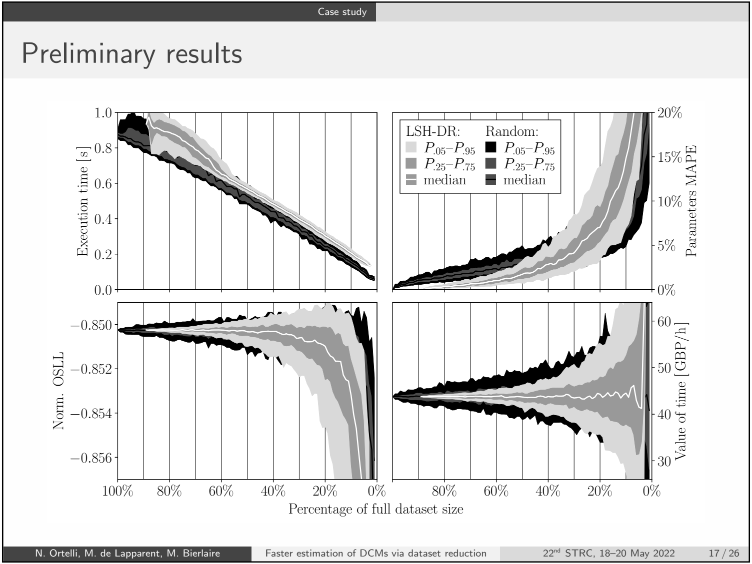
LSH-based dataset reduction (LSH-DR)

Procedure

- Use locality-sensitive hashing (LSH) to cluster observations.
- Sample one observation per cluster.
- Weight based on cluster sizes.



Flashback


Case study
Preliminary results
1.0
0.8
0.6
0.4
0.2
0.0
Execution time [s]
LSH-DR:
Random:
$P_{.05}$–$P_{.95}$
$P_{.25}$–$P_{.75}$
median
20%
15%
10%
5%
0%
Parameters MAPE
−0.850
−0.852
−0.854
−0.856
Norm. OSLL
60
50
40
30
Value of time [GBP/h]
100%
80%
60%
40%
20%
0%
Percentage of full dataset size
N. Ortelli, M. de Lapparent, M. Bierlaire
Faster estimation of DCMs via dataset reduction
22nd STRC, 18–20 May 2022
17 / 26

Resampling estimation of DCMs [Ortelli *et al.*, 2023]

TRANSP-OR EPFL

Resampling estimation of discrete choice models

Nicola Ortelli *† Matthieu de Lapparent * Michel Bierlaire †

March 30, 2023

Report TRANSP-OR 230330
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

*School of Management and Engineering Vaud, HES-SO // University of Applied Sciences and Arts Western Switzerland, {nicola.ortelli,matthieu.delapparent}@heig-vd.ch
†Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne (EPFL), {nicola.ortelli,michel.bierlaire}@epfl.ch

1

Spin-off

LSH-DR

- Substantial time savings when rough estimates are sufficient.
- What can we do when the full-dataset estimates are needed?

Spin-off

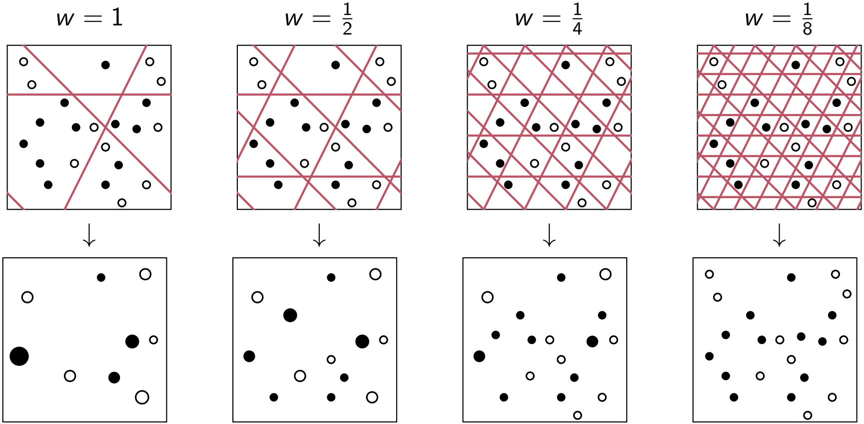
LSH-DR

- Substantial time savings when rough estimates are sufficient.
- What can we do when the full-dataset estimates are needed?

Stochastic adaptive resampling (STAR)

- Embed LSH-DR within the model estimation process.
- Generate batches for stochastic optimization. [Lederrey *et al.*, 2021]
- Start small and increase batch size dynamically.

Illustrative example



Stochastic adaptive resampling (STAR)

Generic algorithm

- Input:
  - $\mathcal{N}$: full dataset;
  - θ_0: initial solution;
- Initialization:
  - $k \leftarrow 0$;
- Repeat:
  1. $\theta_{k+1} \leftarrow \texttt{newCandidate}(\theta_k, \mathcal{N})$;
  2. $k \leftarrow k + 1$;
- Until $||\nabla_{\text{rel}} \mathcal{L}(\theta_k)|| < \varepsilon$.

Stochastic adaptive resampling (STAR)

Generic algorithm

- Input:
  - $\mathcal{N}$: full dataset;
  - θ_0: initial solution;
- Initialization:
  - $k \leftarrow 0$;
- Repeat:
  1. $\theta_{k+1} \leftarrow \texttt{newCandidate}(\theta_k, \mathcal{N})$;
  2. $k \leftarrow k+1$;
- Until $||\nabla_{\text{rel}}\mathcal{L}(\theta_k)|| < \varepsilon$.

Relative gradient

$$[\nabla_{\text{rel}}\mathcal{L}(\theta)]_j = \frac{[\nabla\mathcal{L}(\theta)]_j \cdot \theta_j}{\mathcal{L}(\theta)}$$

Stochastic adaptive resampling (STAR)

Generic algorithm + **STAR**

- Input:
  - $\mathcal{N}$: full dataset;
  - θ_0: initial solution;
  - w_0: initial bucket width.
- Initialization:
  - $k \leftarrow 0$;
- Repeat:
  1. $\theta_{k+1} \leftarrow \texttt{newCandidate}(\theta_k, \mathcal{N})$;
  2. $k \leftarrow k + 1$;
- Until $||\nabla_{\text{rel}}\mathcal{L}(\theta_k)|| < \varepsilon$.

Relative gradient

$$[\nabla_{\text{rel}}\mathcal{L}(\theta)]_j = \frac{[\nabla\mathcal{L}(\theta)]_j \cdot \theta_j}{\mathcal{L}(\theta)}$$

Stochastic adaptive resampling (STAR)

Generic algorithm + **STAR**

- Input:
  - $\mathcal{N}$: full dataset;
  - θ_0: initial solution;
  - w_0: initial bucket width.
- Initialization:
  - $k \leftarrow 0$;
- Repeat:
  1. $\mathcal{N}_k^* \leftarrow \texttt{LSH-DR}\,(w_k, \mathcal{N})$;
  2. $\theta_{k+1} \leftarrow \texttt{newCandidate}\,(\theta_k, \mathcal{N}_k^*)$;

  4. $k \leftarrow k + 1$;
- Until $||\nabla_{\text{rel}}\mathcal{L}(\theta_k)|| < \varepsilon$.

Relative gradient

$$[\nabla_{\text{rel}}\mathcal{L}(\theta)]_j = \frac{[\nabla \mathcal{L}(\theta)]_j \cdot \theta_j}{\mathcal{L}(\theta)}$$

Stochastic adaptive resampling (STAR)

Generic algorithm + **STAR**

- Input:
  - $\mathcal{N}$: full dataset;
  - θ_0: initial solution;
  - w_0: initial bucket width.
- Initialization:
  - $k \leftarrow 0$;
- Repeat:
  1. $\mathcal{N}_k^* \leftarrow \texttt{LSH-DR}(w_k, \mathcal{N})$;
  2. $\theta_{k+1} \leftarrow \texttt{newCandidate}(\theta_k, \mathcal{N}_k^*)$;
  3. $w_{k+1} \leftarrow \texttt{updateW}(w_k, \theta_k, \theta_{k+1})$;
  4. $k \leftarrow k+1$;
- Until $||\nabla_{\text{rel}} \mathcal{L}(\theta_k)|| < \varepsilon$.

Relative gradient

$$[\nabla_{\text{rel}} \mathcal{L}(\theta)]_j = \frac{[\nabla \mathcal{L}(\theta)]_j \cdot \theta_j}{\mathcal{L}(\theta)}$$

Stochastic adaptive resampling (STAR)

Generic algorithm + **STAR**

- Input:
  - $\mathcal{N}$: full dataset;
  - θ_0: initial solution;
  - w_0: initial bucket width.
- Initialization:
  - $k \leftarrow 0$;
- Repeat:
  1. $\mathcal{N}_k^* \leftarrow \texttt{LSH-DR}\,(w_k, \mathcal{N})$;
  2. $\theta_{k+1} \leftarrow \texttt{newCandidate}\,(\theta_k, \mathcal{N}_k^*)$;
  3. $w_{k+1} \leftarrow \texttt{updateW}\,(w_k, \theta_k, \theta_{k+1})$;
  4. $k \leftarrow k + 1$;
- Until $||\nabla_{\text{rel}}\mathcal{L}(\theta_k)|| < \varepsilon$.

Relative gradient

$$[\nabla_{\text{rel}}\mathcal{L}(\theta)]_j = \frac{[\nabla\mathcal{L}(\theta)]_j \cdot \theta_j}{\mathcal{L}(\theta)}$$

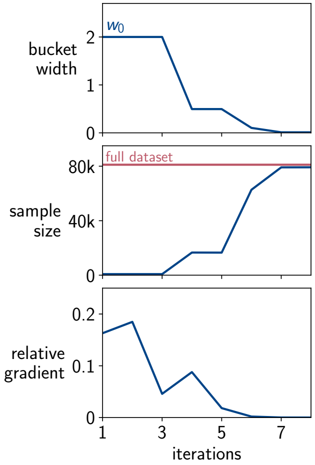
Bucket width update

$$w_{k+1} = w_k \cdot \min\left(1, \frac{||\nabla_{\text{rel}}\mathcal{L}(\theta_{k+1})||}{||\nabla_{\text{rel}}\mathcal{L}(\theta_k)||}\right)$$

Stochastic adaptive resampling (STAR)

Generic algorithm + **STAR**

- Input:
  - $\mathcal{N}$: full dataset;
  - θ_0: initial solution;
  - w_0: initial bucket width.
- Initialization:
  - $k \leftarrow 0$;
- Repeat:
  1. $\mathcal{N}_k^* \leftarrow \texttt{LSH-DR}\,(w, \mathcal{N})$;
  2. $\theta_{k+1} \leftarrow \texttt{newCandidate}\,(\theta_k, \mathcal{N}_k^*)$;
  3. $w_{k+1} \leftarrow \texttt{updateW}\,(w_k, \theta_k, \theta_{k+1})$;
  4. $k \leftarrow k + 1$;
- Until $||\nabla_{\text{rel}}\mathcal{L}(\theta_k)|| < \varepsilon$.



Dataset & models

LPMC data [Hillel *et al.*, 2018]

- Mode choice, 4 alternatives: walk, cycle, drive, public transport.
- **81’086 observations**.

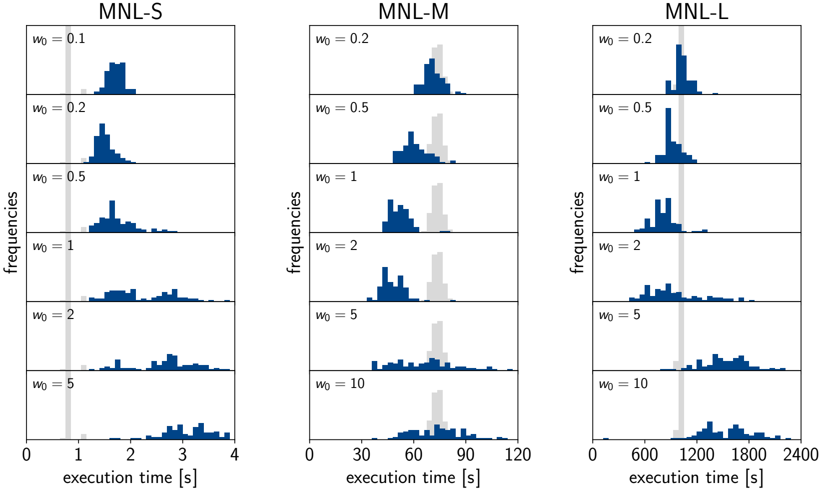
Models [Hillel, 2019]

MNL-S
- 10 cont. variables.
- 0 dummies.
- **13 parameters**.

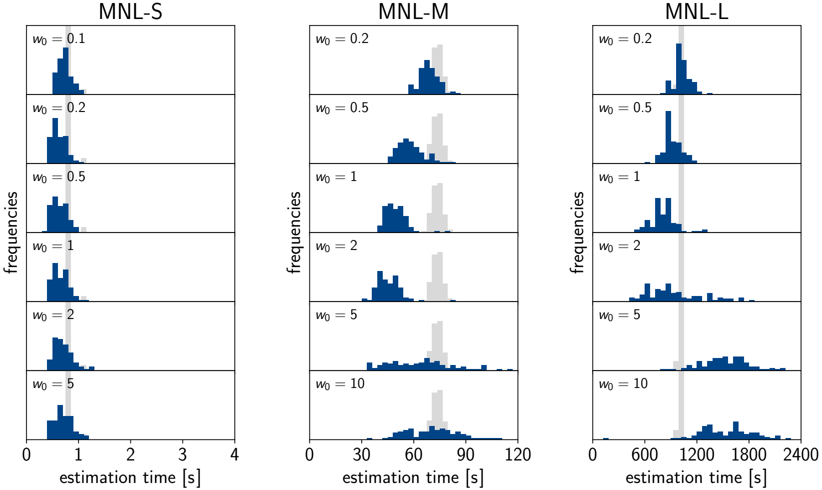
MNL-M
- 11 cont. variables.
- 15 dummies.
- **53 parameters**.

MNL-L
- 13 cont. variables.
- 18 dummies.
- **100 parameters**.

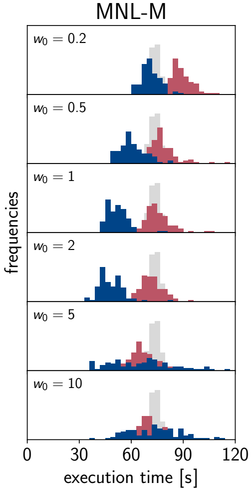
Execution time



Estimation time



Execution time — STAR vs. stochastic



Conclusion

Summary

- Embed LSH-DR within the model estimation process.
- Significant time savings without compromising the quality of results.

Next steps

- Bucket width update.
- Advanced DCMs.

References

LSH-DR

- Ortelli, N., de Lapparent, M. and Bierlaire, M. (2023). Resampling estimation of discrete choice models, Technical Report, TRANSP-OR 230330. Transport and Mobility Laboratory, ENAC, EPFL.
- Ortelli, N., de Lapparent, M. and Bierlaire, M. (2022). Faster estimation of discrete choice models via dataset reduction, Proceedings of the 23rd Swiss Transportation Research Conference.

Direct precedent

- Lederrey, G., Lurkin, V., Hillel, T. and Bierlaire, M. (2021). Estimation of discrete choice models with hybrid stochastic adaptive batch size algorithms, Journal of choice modelling 38.

Dataset & models

- Hillel, T., Elshafie, M. Z. and Jin, Y. (2018). Recreating passenger mode choice-sets for transport simulation: A case study of London, UK, Proceedings of the Institution of Civil Engineers-Smart Infrastructure and Construction 171(1).
- Hillel, T. (2019). Understanding travel mode choice: A new approach for city scale simulation, PhD thesis, University of Cambridge.

HEIG-VD IIDE Institut interdisciplinaire du Développement de l'entreprise

EPFL

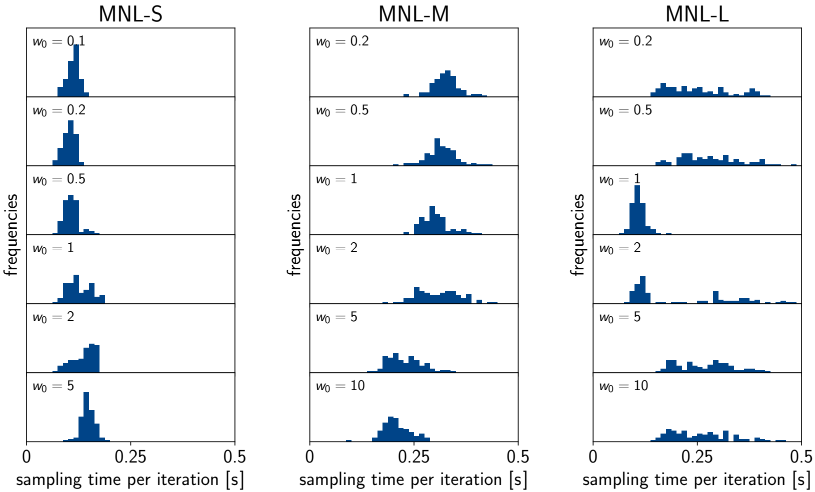
Stochastic adaptive resampling for the estimation of discrete choice models

23rd Swiss Transport Research Conference
10–12 May 2023

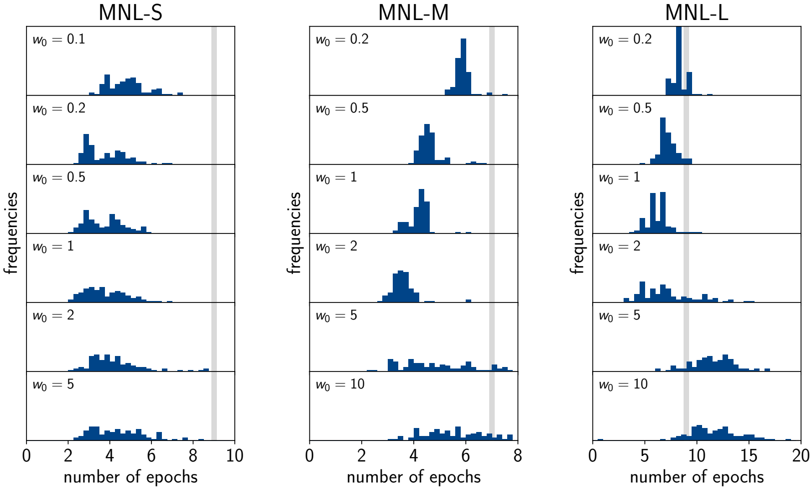
Nicola Ortelli[1,2], Matthieu de Lapparent[1], Michel Bierlaire[2]

[1] IIDE, HEIG-VD
[2] TRANSP-OR, EPFL

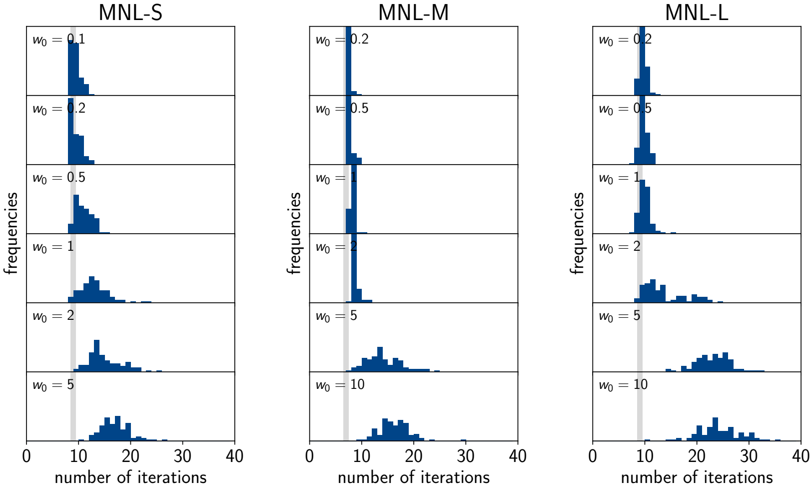
Sampling time per iteration



Number of epochs



Number of iterations



Model	Execution time [s]		Ratio
	Newton-TR*	Newton-TR	
MNL-S	1.5 ± 0.2	0.8 ± 0.1	**1.89**
MNL-M	48.2 ± 6.8	73.6 ± 2.8	**0.65**
MNL-L	811.9 ± 141.6	1'003.8 ± 10.9	**0.81**