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

EPFL

Stochastic adaptive resampling for the estimation of discrete choice models

15th Workshop on Discrete Choice Models
1–3 June 2023 | EPFL, Switzerland

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

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

DCMs in the era of big data

Ever-larger choice datasets

- Exponential growth of collected data:
  - "Wider" data — more variables;
  - "Taller" data — more observations.
- Two distinct problems: specification and estimation.

Model specification is iterative

- More candidates are tested before reaching an adequate model...
- ... and each of them takes longer to estimate.

Estimating DCMs

Maximum likelihood estimation (MLE)

- Suppose a dataset $\mathcal{N} = \{(\boldsymbol{x_n}, i_n) : n = 1, \ldots, N\}$.
- Maximize the joint probability of the observed choices:

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

  - Solved using iterative methods: BFGS, Newton, etc.
  - Each iteration is $\mathcal{O}(N)$.

Intuition

Factoring-out redundancy

- If the data contains identical observations:

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

  - U unique observations ($U \leq N$).
  - Each appears N_u times in the original data.
- Evaluating $\mathcal{L}(\boldsymbol{\theta})$ is faster by a ratio of approximately $\frac{U}{N}$.

$\Rightarrow$ **Extend this trick to "nearly identical" observations!**

Resampling estimation of DCMs

Procedure

- Group similar observations.
- Sample from each group.
- Associate weights based on group sizes.
- Use MLE on the weighted subsample.

Challenges

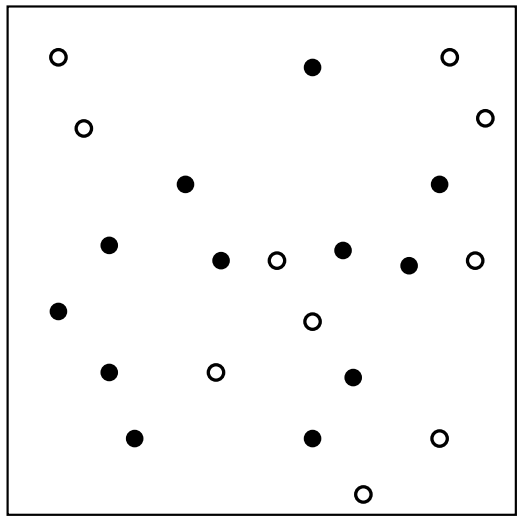
- Minimize information loss.
- Still, clustering must be fast!

$\Rightarrow$ **Use locality-sensitive hashing (LSH).**

LSH-based dataset reduction (LSH-DR)

Toy data

- 2 alternatives.
- 2 explanatory variables.

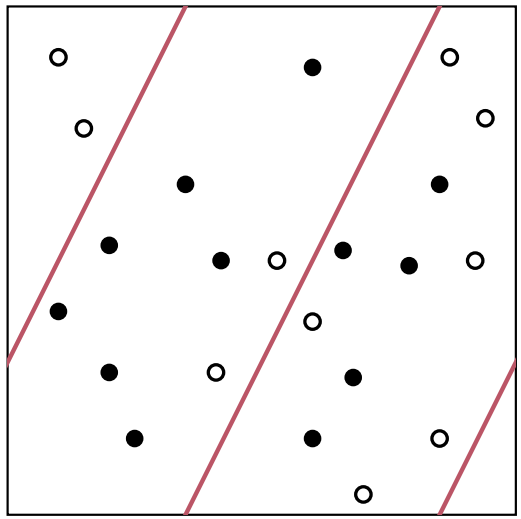


LSH-based dataset reduction (LSH-DR)

LSH function

$$h_{\boldsymbol{a},b}(\boldsymbol{x}_n) = \left\lfloor \frac{\boldsymbol{a} \cdot \boldsymbol{x}_n + b}{w} \right\rfloor$$

- $\boldsymbol{a} \overset{iid}{\sim} \mathcal{N}(0,1)$.
- $b \sim \mathcal{U}(0,w)$.
- w is the bucket width.

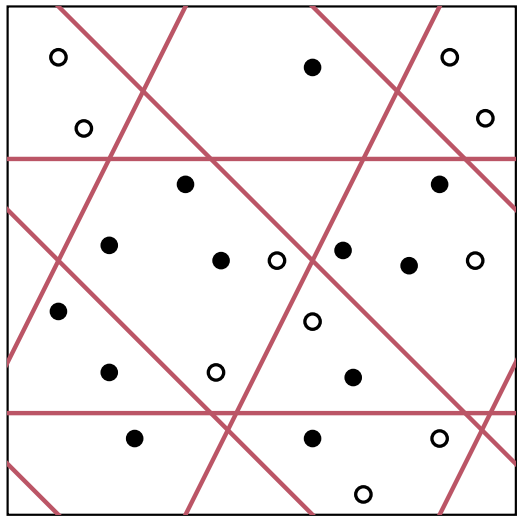


LSH-based dataset reduction (LSH-DR)

Combined LSH functions

$$H_{\boldsymbol{A},B}(\boldsymbol{x}_n) = H_{\boldsymbol{A},B}(\boldsymbol{x}_p) \Leftrightarrow$$
$$h_{\boldsymbol{a}_r,b_r}(\boldsymbol{x}_n) = h_{\boldsymbol{a}_r,b_r}(\boldsymbol{x}_p) \forall r = 1,\dots,R.$$

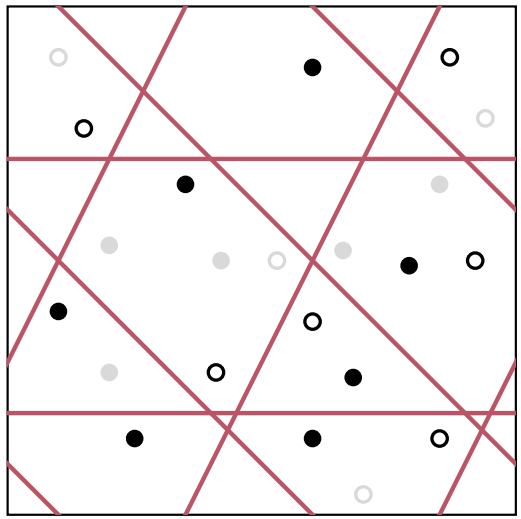
- $\boldsymbol{A} = (\boldsymbol{a}_1,\dots,\boldsymbol{a}_R)$.
- $B = (b_1,\dots,b_R)$.



LSH-based dataset reduction (LSH-DR)

Sampling

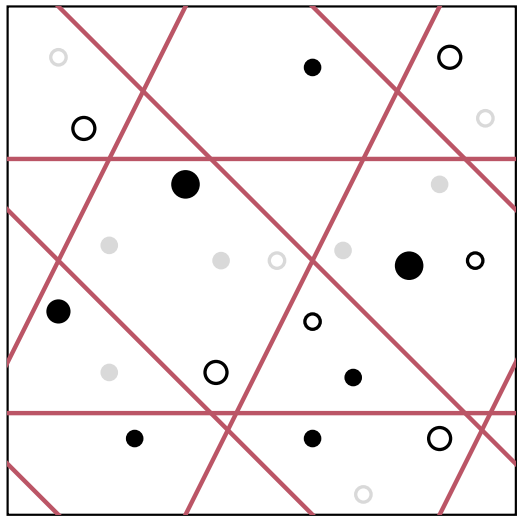
In each bucket, sample one observation $(\boldsymbol{x}_g, i_g)$ per alternative.



LSH-based dataset reduction (LSH-DR)

Associated weights N_g

$$N_g = |\{(\boldsymbol{x}_n, i_n) : H_{\boldsymbol{A},B}(\boldsymbol{x}_g) = H_{\boldsymbol{A},B}(\boldsymbol{x}_n),\ i_g = i_n\}|$$



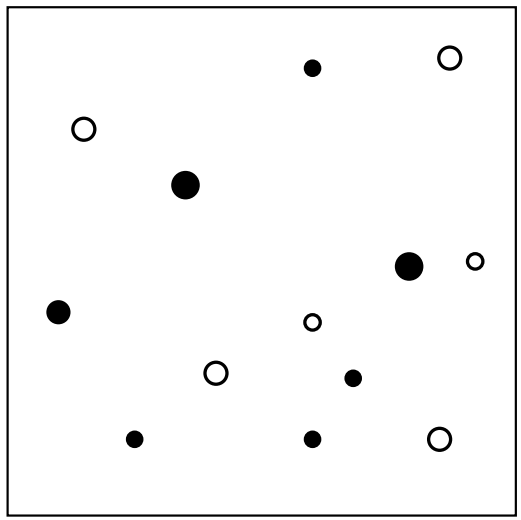
LSH-based dataset reduction (LSH-DR)

Subsample, weighted

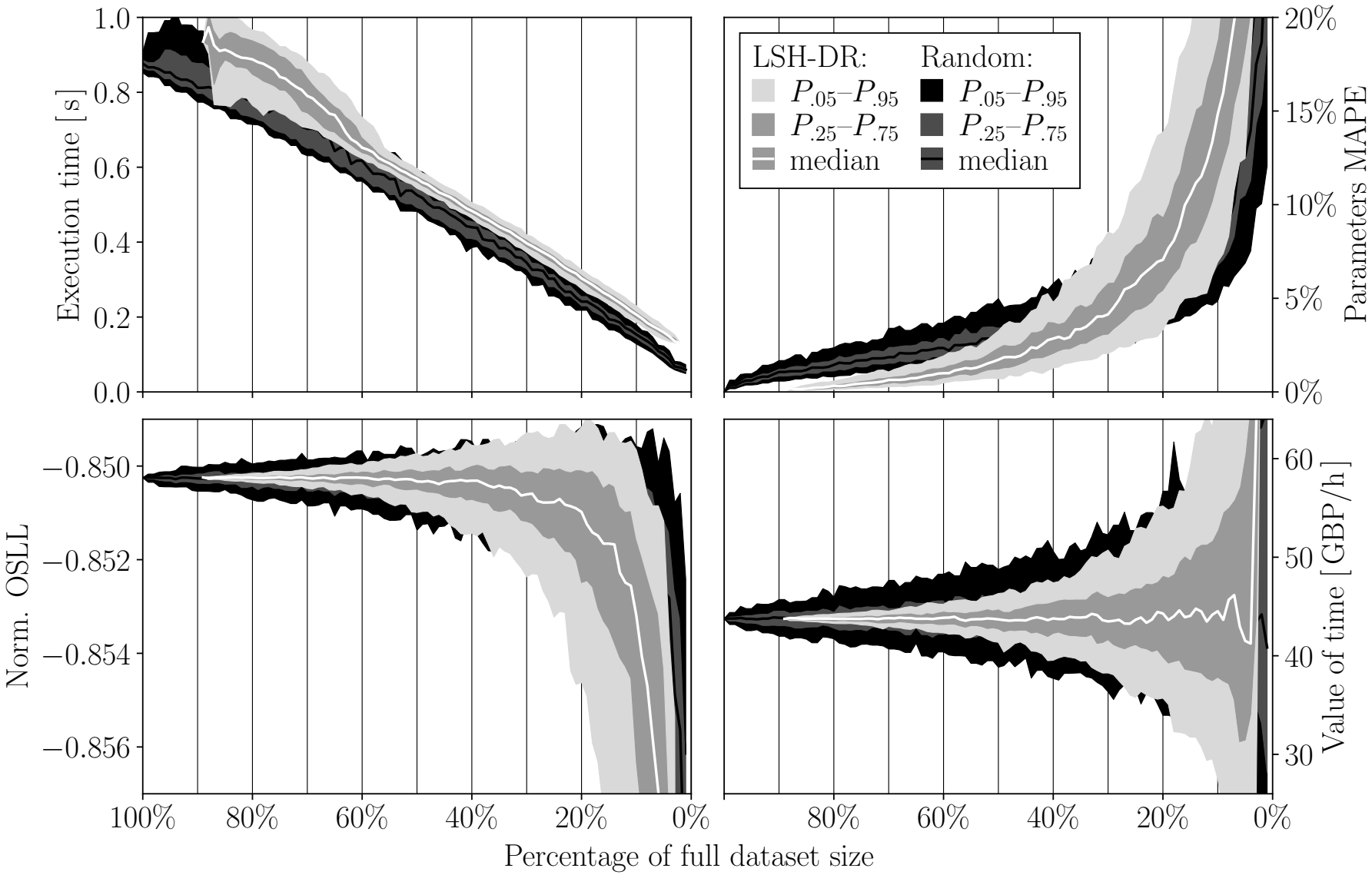
$$\mathcal{N}^* = \{(\boldsymbol{x_g}, i_g, N_g) : g = 1, \ldots, G\}$$

Weighted MLE

$$\mathcal{L}^*(\boldsymbol{\theta}) = \sum_{g=1}^{G} N_g \cdot \log P\left(i_g \mid \boldsymbol{x_g}; \boldsymbol{\theta}\right)$$



Results


Execution time [s]
1.0
0.8
0.6
0.4
0.2
0.0
LSH-DR:
$P_{.05}$–$P_{.95}$
$P_{.25}$–$P_{.75}$
median
Random:
$P_{.05}$–$P_{.95}$
$P_{.25}$–$P_{.75}$
median
Parameters MAPE
20%
15%
10%
5%
0%
Norm. OSLL
−0.850
−0.852
−0.854
−0.856
Value of time [GBP/h]
60
50
40
30
100%
80%
60%
40%
20%
0%
Percentage of full dataset size

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

Spin-off

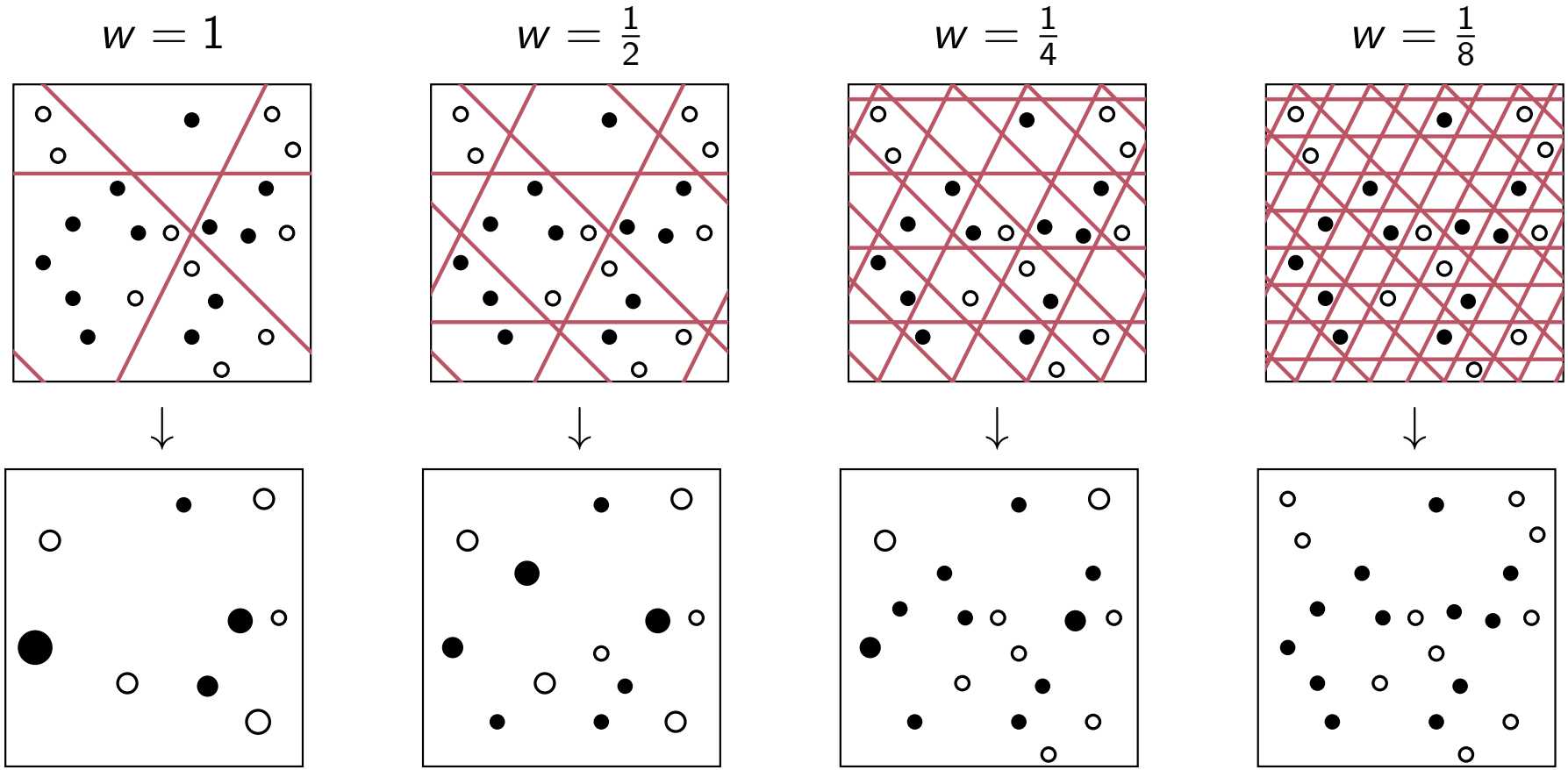
LSH-DR

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

Stochastic adaptive resampling (STAR)

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

Illustration



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

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

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

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

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

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

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

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

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

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

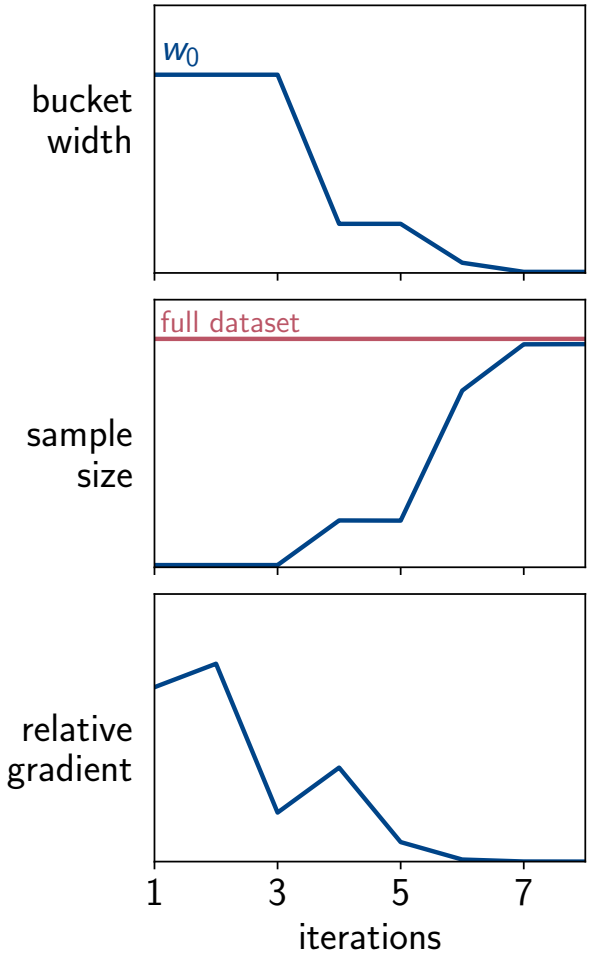
Bucket width update

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

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 $||\nabla_{\text{rel}} \mathcal{L}(\theta_k)|| < \varepsilon$.



Experimental design

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

- Mode choice, 4 alternatives.
- 81'086 observations.

Models [Hillel, 2019]

MNL-S

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

MNL-M

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

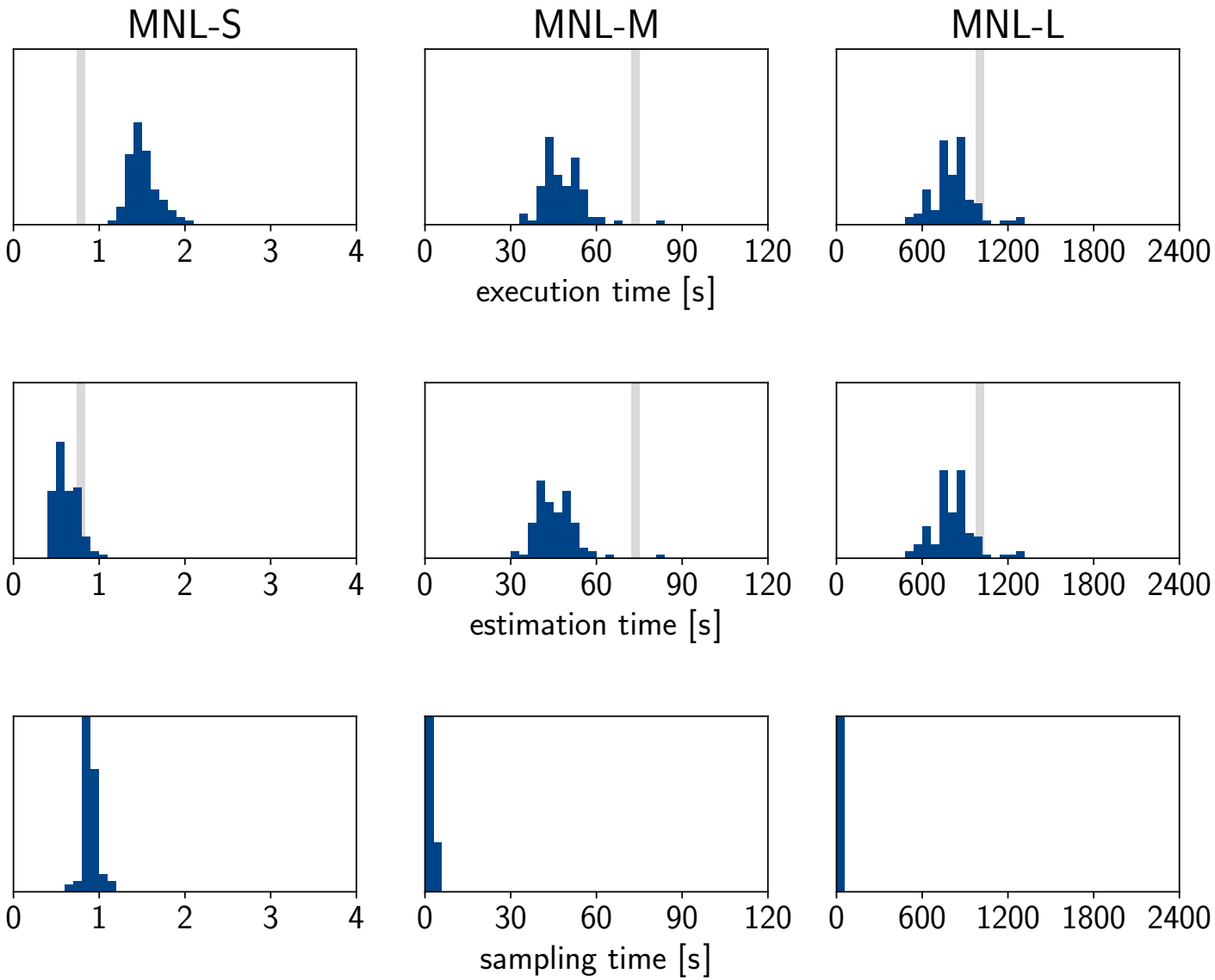
MNL-L

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

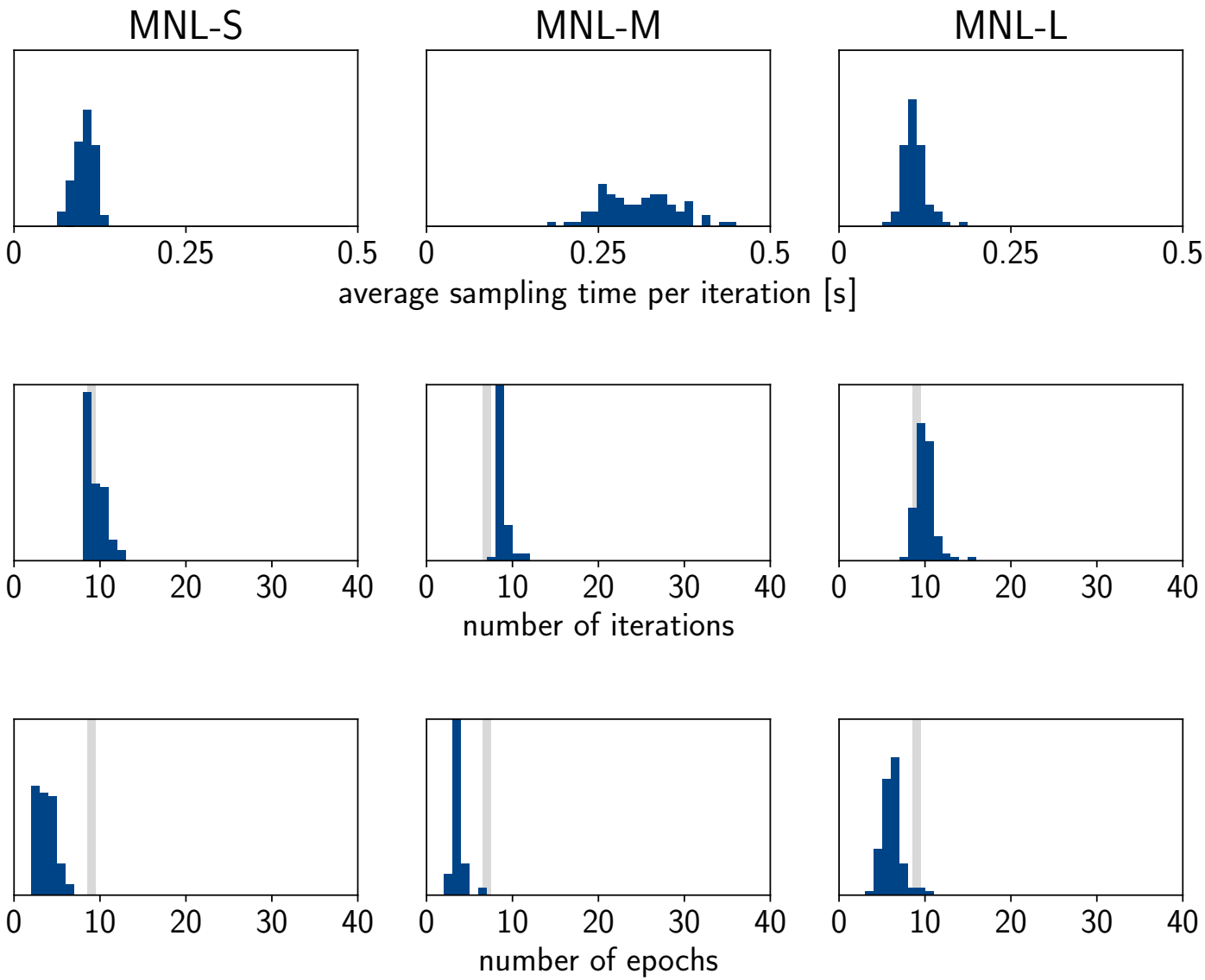
Optimization algorithm

- Newton-TR + STAR (100 repetitions).
- Benchmark: standard Newton-TR.

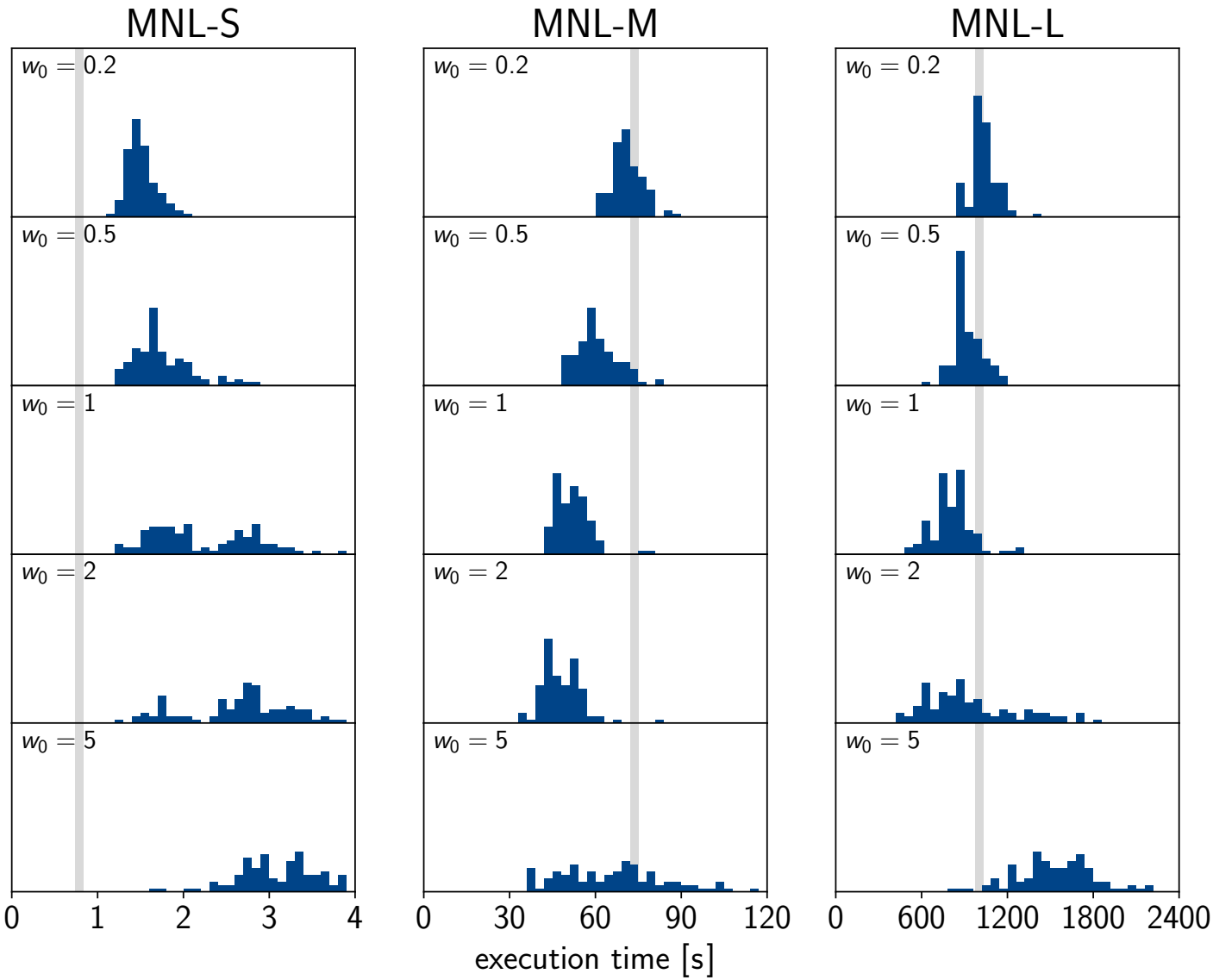
Results: estimation and sampling times



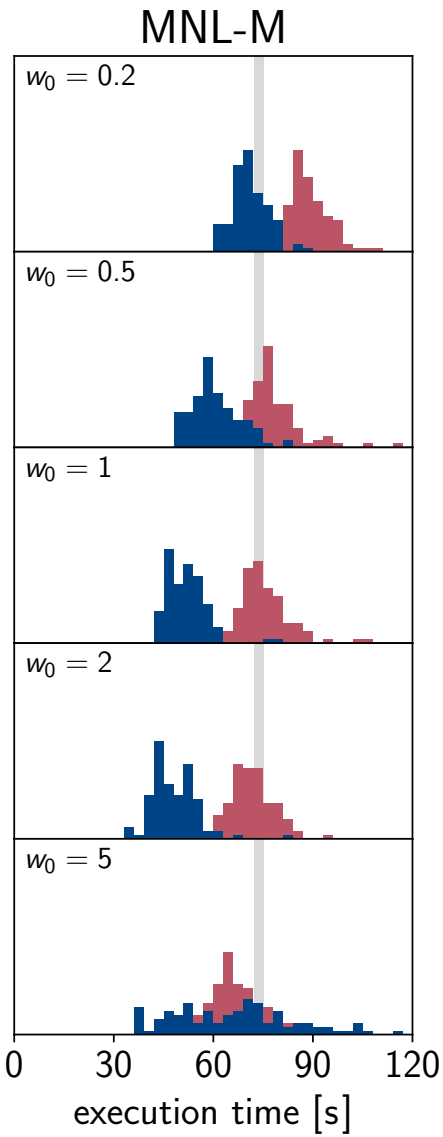
Results: iterations, epochs



Results: effect of w_0



Results: weighted/unweighted subsamples



Conclusion

Summary

- LSH-DR if rough estimates are sufficient.
- STAR when full-dataset estimates are needed.

Next steps

- Bucket width update.
- More advanced DCMs.
- Test with other optimization algorithms.

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 precedents

- 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

15th Workshop on Discrete Choice Models
1–3 June 2023 | EPFL, Switzerland

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

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