# Saddlepoint Monte Carlo and its Application to Exact Ecological Inference of Voting Patterns

We have a nice model for $X$ but only observe $Y = AX$.

Robin J. Ryder

Department of Mathematics, Imperial College London

26 March 2026

Statistique au sommet de Rochebrune

Joint work with Théo Voldoire (Harvard) and Nicolas Chopin (ENSAE)

# Species under study: *Homo politicus*

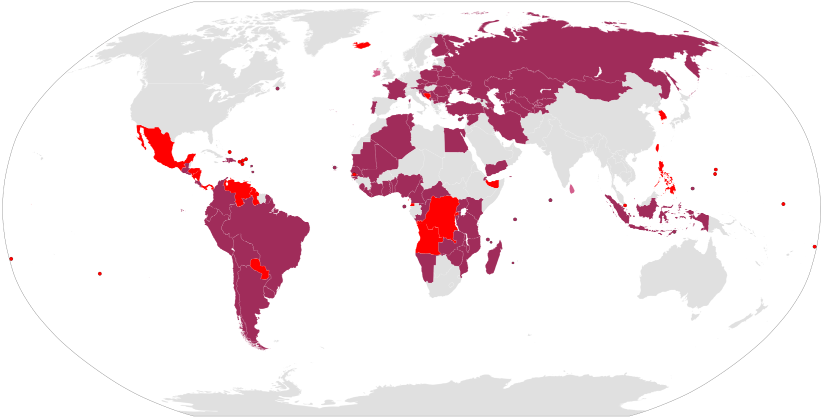
# Species under study: *Homo politicus*

# Background: French elections

- Most French elections have two rounds.
- Round 1: many candidates.
- The top 2 candidates in round 1 qualify for Round 2 (runoff).
- 2022 presidential election:
    - 12 candidates in round 1.
    - Round 1 results: Macron 28%, Le Pen 23%, Mélenchon 22%, Zemmour 7%...
    - Round 2 results: Macron 59%, Le Pen 41%.
- At each round, voters can also abstain. We think of abstention as an extra candidate.
- Voter transfer between the rounds is a question of high interest, and difficult to poll.

# Two-round voting

This system and its variants (eg with more than 2 candidates in the runoff) are widely used across the world: election of head of state in 85 countries; election of legislative chambers in at least 13 countries and 4 US states.



Direct election of head on state: (Map by User:Rankedchoicevoter, Wikimedia Commons.)

Two-round First-past-the-post Instant runoff Indirect election or unelected.

# Data

- About 60 000 polling stations across France
- Typical polling station comprises 1 000 voters
- We observe the marginal results at each round in each polling station.

Example data for a single polling station:

Round 1

| Macron | Le Pen | Mélenchon | Zemmour | Pécresse | ... | Abstention |
|---|---|---|---|---|---|---|
| 317 | 135 | 195 | 110 | 49 | ... | 358 |

Round 2

| Macron | Le Pen | Abstention |
|---|---|---|
| 554 | 279 | 407 |

# Matrix form

For each polling station, we can represent the votes as a matrix:

| | Macron 2 | Le Pen 2 | Abstention 2 | Total Round 1 |
|---|---|---|---|---|
| Macron | ? | ? | ? | 317 |
| Le Pen | ? | ? | ? | 135 |
| Mélenchon | ? | ? | ? | 195 |
| Zemmour | ? | ? | ? | 110 |
| $\vdots$ | ? | ? | ? | $\vdots$ |
| Total Round 2 | 554 | 279 | 407 | 1240 |

We observe row sums and column sums of a collection of latent matrices.
We wish to model the values of individual cells of the matrices.

# Observations

| | Macron 2 | Le Pen 2 | Abstention 2 | Total Round 1 |
|---|---|---|---|---|
| Macron | ? | ? | ? | 317 |
| Le Pen | ? | ? | ? | 135 |
| Mélenchon | ? | ? | ? | 195 |
| Zemmour | ? | ? | ? | 110 |
| ... | ? | ? | ? | ... |
| Total Round 2 | 554 | 279 | 407 | 1240 |

$k = 1$

| | Macron 2 | Le Pen 2 | Abstention 2 | Total Round 1 |
|---|---|---|---|---|
| Macron | ? | ? | ? | 220 |
| Le Pen | ? | ? | ? | 120 |
| Mélenchon | ? | ? | ? | 174 |
| Zemmour | ? | ? | ? | 57 |
| ... | ? | ? | ? | ... |
| Total Round 2 | 444 | 193 | 356 | 993 |

$k = 2$

| | Macron 2 | Le Pen 2 | Abstention 2 | Total Round 1 |
|---|---|---|---|---|
| Macron | ? | ? | ? | 306 |
| Le Pen | ? | ? | ? | 46 |
| Mélenchon | ? | ? | ? | 52 |
| Zemmour | ? | ? | ? | 57 |
| ... | ? | ? | ? | ... |
| Total Round 2 | 467 | 110 | 227 | 804 |

$k = 3$

| | Macron 2 | Le Pen 2 | Abstention 2 | Total Round 1 |
|---|---|---|---|---|
| Macron | ? | ? | ? | 257 |
| Le Pen | ? | ? | ? | 75 |
| Mélenchon | ? | ? | ? | 149 |
| Zemmour | ? | ? | ? | 72 |
| ... | ? | ? | ? | ... |
| Total Round 2 | 465 | 185 | 351 | 1001 |

$k = 4$

| | Macron 2 | Le Pen 2 | Abstention 2 | Total Round 1 |
|---|---|---|---|---|
| Macron | ? | ? | ? | 431 |
| Le Pen | ? | ? | ? | 54 |
| Mélenchon | ? | ? | ? | 107 |
| Zemmour | ? | ? | ? | 175 |
| ... | ? | ? | ? | ... |
| Total Round 2 | 722 | 223 | 300 | 1245 |

$k = 5$

$\ddots$

| | Macron 2 | Le Pen 2 | Abstention 2 | Total Round 1 |
|---|---|---|---|---|
| Macron | ? | ? | ? | 238 |
| Le Pen | ? | ? | ? | 140 |
| Mélenchon | ? | ? | ? | 107 |
| Zemmour | ? | ? | ? | 175 |
| ... | ? | ? | ? | ... |
| Total Round 2 | 457 | 258 | 461 | 1176 |

$k = K = 60\,000$

# Parametrization

We thus observe the margins (row and column sums) of matrices $\mathbf{X}^1, \dots, \mathbf{X}^K$, with $\mathbf{X}^k = (X_{ij}^k)_{i,j}$. Natural models might be of the form

$$(\mathbf{X}^k) \sim \mathcal{M}(n^k, \mathbf{p}^k)$$

for some probabilities $\mathbf{p}^k = (p_{ij}^k)_{i,j}$ at polling station $k$ ($1 \leq k \leq K$, $K = 60\,000$), with for example

# Parametrization

We thus observe the margins (row and column sums) of matrices $\mathbf{X}^1, \dots, \mathbf{X}^K$, with $\mathbf{X}^k = (X_{ij}^k)_{i,j}$. Natural models might be of the form

$$(\mathbf{X}^k) \sim \mathcal{M}(n^k, \mathbf{p}^k)$$

for some probabilities $\mathbf{p}^k = (p_{ij}^k)_{i,j}$ at polling station $k$ ($1 \leq k \leq K$, $K = 60\,000$), with for example

1. constant probabilities: $\forall k, p_{ij}^k = p_{ij}$

# Parametrization

We thus observe the margins (row and column sums) of matrices $\mathbf{X}^1, \dots, \mathbf{X}^K$, with $\mathbf{X}^k = (X_{ij}^k)_{i,j}$. Natural models might be of the form

$$(\mathbf{X}^k) \sim \mathcal{M}(n^k, \mathbf{p}^k)$$

for some probabilities $\mathbf{p}^k = (p_{ij}^k)_{i,j}$ at polling station $k$ ($1 \leq k \leq K$, $K = 60\,000$), with for example

1. constant probabilities: $\forall k, p_{ij}^k = p_{ij}$
2. constant transition rates: $p_{ij}^k = p_i^k \cdot p_{j|i}$

# Parametrization

We thus observe the margins (row and column sums) of matrices $\mathbf{X}^1, \dots, \mathbf{X}^K$, with $\mathbf{X}^k = (X_{ij}^k)_{i,j}$. Natural models might be of the form

$$(\mathbf{X}^k) \sim \mathcal{M}(n^k, \mathbf{p}^k)$$

for some probabilities $\mathbf{p}^k = (p_{ij}^k)_{i,j}$ at polling station $k$ ($1 \leq k \leq K$, $K = 60\,000$), with for example

1. constant probabilities: $\forall k, p_{ij}^k = p_{ij}$
2. constant transition rates: $p_{ij}^k = p_i^k \cdot p_{j|i}$
3. influence of covariates $Z$: $\text{logit}(p_{ij}^k) = Z_k^\top \beta_{ij}$

and we want to perform inference on the $(\mathbf{p}^k)_k$ based on the margins (row and column sums).

Such questions fall in the realm of Ecological Inference (EI).

# Existing methods in Ecological inference

- Ecological inference is a well-studied problem, common in political science, sociology, epidemiology...
- Attempts to resolve the *ecological fallacy* / *aggregation bias*.

For large $K$ and large(ish) dimensions of $X$, existing methods include:

- Approximating the Multinomial distribution by a Normal distribution
- Changing the model to something more tractable
- Asymptotic approximations, e.g. via maximum entropy estimation
- Data augmentation

See King (1997), Wakefield (2001), Chen et al. (2005), Imai et al. (2008)...

# Existing methods in Ecological inference

- Ecological inference is a well-studied problem, common in political science, sociology, epidemiology...
- Attempts to resolve the *ecological fallacy* / *aggregation bias*.

For large $K$ and large(ish) dimensions of $X$, existing methods include:

- Biased Approximating the Multinomial distribution by a Normal distribution
- Not flexible Changing the model to something more tractable
- Biased Asymptotic approximations, e.g. via maximum entropy estimation
- Doesn't scale Data augmentation

See King (1997), Wakefield (2001), Chen et al. (2005), Imai et al. (2008)...

# Our goal

We aim for Ecological inference methodology which is

- Exact
- Scalable for large $K$ and large dimensions of $X$
- Efficient even for more flexible model specifications

# Vectorize

| | | | | |
|---|---|---|---|---|
| $X_{11}$ | $X_{12}$ | $\dots$ | $X_{1c}$ | $Y_{1\cdot} = \sum_i X_{1i}$ |
| $X_{21}$ | $X_{22}$ | $\dots$ | $X_{2c}$ | $Y_{2\cdot} = \sum_i X_{2i}$ |
| $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $X_{r1}$ | $X_{r2}$ | $\dots$ | $X_{rc}$ | $Y_{r\cdot} = \sum_i X_{ri}$ |
| $Y_{\cdot 1} = \sum_j X_{j1}$ | $Y_{\cdot 2} = \sum_j X_{j2}$ | $\dots$ | $Y_{\cdot c} = \sum_j X_{jc}$ | $n$ |

Vectorize as

$$\mathbf{X} = \begin{pmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1c} \\ X_{21} \\ X_{22} \\ \vdots \\ X_{2c} \\ \vdots \\ X_{r1} \\ X_{r2} \\ \vdots \\ X_{rc} \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} Y_{1\cdot} \\ Y_{2\cdot} \\ \vdots \\ Y_{r\cdot} \\ Y_{\cdot 1} \\ Y_{\cdot 2} \\ \vdots \\ Y_{\cdot c} \end{pmatrix} = A\mathbf{X} \qquad A = \begin{pmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots & 1 \\ 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 & \dots & 0 & 0 & \dots & 1 \end{pmatrix}$$

# General setting

Notation:

- $X$, random vector in $\mathbb{R}^{d_X}$, $X \sim \mathcal{L}_\theta$, $\theta \in \mathbb{R}^p$
- $A$, $d_Y \times d_X$ matrix, non-invertible ($d_Y \ll d_X$)

Problem:

- No access to sample $(X^k)_{1 \leq k \leq K}$, but only $(Y^k) = (AX^k)$.
- Problem: $\mathbb{P}(AX = y)$ hard to compute.

Use cases:

- Ecological Inference and more generally privacy by aggregation
- Surjective/non-invertible/ill-posed inverse problems
- Convolution of random variables

# Baseline MCMC framework

- Goal: design a MCMC kernel so that $(\theta_t)$ has stationary distribution $\pi(\theta|AX = y) \propto \pi_{\text{prior}}(\theta) f_{AX}^{\theta}(y)$.
- Metropolis-Rosenbluth-Teller-Hastings with kernel $q$: sample $\tilde{\theta} \sim q(\cdot|\theta_t)$, then

$$\theta_{t+1} = \begin{cases} \tilde{\theta} \text{ with probability } \alpha(\theta_t, \tilde{\theta}) \\ \theta_t \text{ else} \end{cases}, \qquad (1)$$

with $\alpha(\theta_n, \tilde{\theta}) = \frac{\pi(\tilde{\theta}|AX=b)q(\theta_t|\tilde{\theta})}{\pi(\theta_t|AX=b)q(\tilde{\theta}|\theta_t)}$.

- Problem: no access to $f_{AX}^{\theta}(y)$.

# Pseudo-marginal likelihood methods

Drop dependency on $\theta$ from now on: $f_{AX}(y)$.

- Solution: replace $f_{AX}(y)$ with an unbiased estimator

$$\mathbb{E}[\hat{\delta}_{AX=y}] = f_{AX}(y). \tag{2}$$

- Leaves invariant the *true* posterior distribution $\pi(\theta|y)$ (Andrieu & Roberts, 2009).
- New goal: construct $\hat{\delta}_{AX=y}$ of sufficiently small relative variance.

# Characteristic function basics

- Essential assumption: *conditional on* $\theta$, $X$ admits a closed-form characteristic function $\varphi_X(t) = \mathbb{E}\left[e^{itX}\right]$.
- Property: $\varphi_{AX}(z) = \varphi_X(A^\top z)$ ($\rightarrow$ dimension reduction)
- Fourier inversion formula:

$$f_{AX}(y) = \frac{1}{(2\pi)^{d_Y}} \int_{[-\pi,\pi]^{d_Y}} e^{-iz^\top b} \varphi_{AX}(z) dz \tag{3}$$

# Characteristic function basics

- Essential assumption: *conditional on* $\theta$, $X$ admits a closed-form characteristic function $\varphi_X(t) = \mathbb{E}\left[e^{itX}\right]$.
- Property: $\varphi_{AX}(z) = \varphi_X(A^\top z)$ ($\rightarrow$ dimension reduction)
- Fourier inversion formula:

$$f_{AX}(y) = \frac{1}{(2\pi)^{d_Y}} \int_{[-\pi,\pi]^{d_Y}} e^{-iz^\top b} \varphi_{AX}(z) dz \tag{3}$$

Baseline estimator:

$$\hat{\delta}^{\text{Unif}}_{\theta,b} = \frac{1}{N_{\text{Sim}}} \sum_{n=1}^{N_{\text{Sim}}} e^{-iZ_n} \varphi_{AX}(Z_n), \quad Z_n \sim \mathcal{U}([-\pi,\pi]^{d_Y}), \tag{4}$$

Problem: high variance.

# Two remarks

Looking again at:

$$f_{AX}(y) = \frac{1}{(2\pi)^{d_Y}} \int_{[-\pi,\pi]^{d_Y}} e^{-iz^\top b} \varphi_{AX}(z) dz$$

- Aggregation effect (CLT): we expect $AX \approx Y'$, with $Y' \sim \mathcal{N}(A\mu_X, A\Sigma_X A^\top)$, $\mu_X = \mathbb{E}[X]$, $\Sigma_X = \mathrm{Var}[X]$.

# Two remarks

Looking again at:

$$f_{AX}(y) = \frac{1}{(2\pi)^{d_Y}} \int_{[-\pi,\pi]^{d_Y}} e^{-iz^\top b} \varphi_{AX}(z) dz$$

- Aggregation effect (CLT): we expect $AX \approx Y'$, with $Y' \sim \mathcal{N}(A\mu_X, A\Sigma_X A^\top)$, $\mu_X = \mathbb{E}[X]$, $\Sigma_X = \mathrm{Var}[X]$.
- The characteristic function of $\mathcal{N}(0, \Sigma)$ is (up to a constant) the pdf of $\mathcal{N}(0, \Sigma^{-1})$.

# Derivation

$$
\begin{aligned}
f_{AX}(y) &= \int_{\mathbb{R}^{d_Y}} \frac{1_{[-\pi,\pi]^{d_Y}}(z)}{(2\pi)^{d_Y}} e^{-iz^\top y} \varphi_{AX}(z) dz \\
&= \int_{\mathbb{R}^{d_Y}} \frac{1_{[-\pi,\pi]^{d_Y}}(z)}{(2\pi)^{d_Y}} e^{-iz^\top y} \frac{\varphi_{AX}(z)}{\varphi_{AX'}(z)} \varphi_{AX'}(z) dz \\
&= \int_{\mathbb{R}^{d_Y}} \frac{1_{[-\pi,\pi]^{d_Y}}(z)}{(2\pi)^{d_Y}} e^{iz^\top (A\mu_X - y)} \frac{\varphi_{AX}(z)}{\varphi_{AX'}(z)} e^{-\frac{1}{2} z^\top (A\Sigma_X A^\top) z} dz \\
&= \mathbb{E}_{Z \sim \mathcal{N}(0, (A\Sigma_X A^\top)^{-1})} \left[ \eta_{\theta,y}(Z) \right],
\end{aligned}
$$

with $\eta_{\theta,y}(z) = \frac{1_{[-\pi,\pi]^{d_Y}}(z)}{(2\pi)^{d_Y}} e^{iz^\top (A\mu_X - y)} \frac{\varphi_{AX}(z)}{\varphi_{AX'}(z)}$.

# New estimator (Gaussian proposal)

Leading to:

$$\hat{\delta}_{\theta,y}^{\text{Norm}} = \frac{1}{N_{\text{Sim}}} \sum_{k=1}^{N_{\text{Sim}}} \eta_{\theta,y}(Z_k), \tag{5}$$

with $Z_1, ..., Z_{N_{\text{Sim}}} \sim \mathcal{N}(0, (A\Sigma_X A^\top)^{-1})$.

This still has large variance.

# Tilting basics

- Parametric change of measure:

$$f_{X_\rho}(x) = \frac{e^{\rho^\top x}}{M_X(\rho)} f_X(x), \quad \rho \in \mathbb{R}^{d_X}$$

with $M_X$ the moment generating function of $X$.

# Tilting basics

- Parametric change of measure:

$$f_{X_\rho}(x) = \frac{e^{\rho^\top x}}{M_X(\rho)} f_X(x), \quad \rho \in \mathbb{R}^{d_X}$$

  with $M_X$ the moment generating function of $X$.

- If $X \sim \mathcal{M}(n, p)$, $X_\rho \sim \mathcal{M}(n, p_\rho)$, with

$$p_\rho[j] \propto p[j] \exp\left\{\rho[j]\right\}.$$

# Tilting basics

- Parametric change of measure:

$$f_{X_\rho}(x) = \frac{e^{\rho^\top x}}{M_X(\rho)} f_X(x), \quad \rho \in \mathbb{R}^{d_X}$$

with $M_X$ the moment generating function of $X$.

- If $X \sim \mathcal{M}(n, p)$, $X_\rho \sim \mathcal{M}(n, p_\rho)$, with

$$p_\rho[j] \propto p[j] \exp\{\rho[j]\}.$$

- More generally, if dist $X$ is an exponential family, then dist. of $X_\rho$ will be in the same family.

# Tilting basics

- Parametric change of measure:

$$f_{X_\rho}(x) = \frac{e^{\rho^\top x}}{M_X(\rho)} f_X(x), \quad \rho \in \mathbb{R}^{d_X}$$

  with $M_X$ the moment generating function of $X$.
- If $X \sim \mathcal{M}(n, p)$, $X_\rho \sim \mathcal{M}(n, p_\rho)$, with

$$p_\rho[j] \propto p[j] \exp\{\rho[j]\}.$$

- More generally, if dist $X$ is an exponential family, then dist. of $X_\rho$ will be in the same family.
- Problem: we work with $AX$, not $X$.

# Derivation

Developing the tilted form

$$f_{AX_\rho}(y) = \sum_x f_{X_\rho}(x)\mathbf{1}_{Ax=y} = \sum_x \frac{e^{\rho^\top x}}{M_X(\rho)} f_X(x)\mathbf{1}_{Ax=y}. \quad (6)$$

If $\rho = A^\top \nu$ (for som $\nu$), then $\rho^\top x = \nu^\top y$, leading to

$$f_{AX_\rho}(y) = \frac{e^{\nu^\top y}}{M_{AX}(\nu)} f_{AX}(y). \quad (7)$$

A tilting on $X$ leads to a tilting on $AX$!

# Generalised estimator

Since, for any $\nu$,

$$f_{AX}(y) = \frac{M_X(A^\top \nu)}{\exp(\nu^\top y)} f_{AX_\rho}(y), \quad \rho = A^\top \nu,$$

We can consider the generalised estimator:

$$\hat{f}^{\nu}_{AX}(y) := \frac{M_X(A^\top \nu)}{\exp(\nu^\top y)} \hat{f}_{AX_\rho}(y), \quad \rho = A^\top \nu,$$

where $\hat{f}_{AX_\rho}(y)$ is our basic estimator (using a Gaussian proposal) under the distribution of $X_\rho$.

# How do we choose $\nu$?

Idea: choose $\nu$ to miminise variance. Heuristic:

$$A\nabla\kappa_X(A^\top \nu) = y, \tag{8}$$

where $\kappa_X$ is the cumulant function of $X$:

- contrast with standard saddlepoint equation $\nabla\kappa_X(\rho) = y$.
- Often implies $A\mu_X = y$.
- In other words: tilt the distribution so that its expectation fits the data.

# A supporting result

If you do tilting, with $\nu$ as in the previous slide, then

$$\mathrm{Var}\left[\frac{\hat{f}^{\nu}_{AX}(y)}{f_{AX}(y)}\right] \to 0, \quad \text{as } n \to \infty$$

where $n$ is the data sample size.

$\Rightarrow$ Relative error gets small when $n$ gets large.

That result does not seem to hold without tilting (or with tilting and a different choice of $\nu$).

# In practice

For realistic applications:

- The baseline estimator $\hat{\delta}^{\text{Unif}}_{\theta,b_k}$ may require to sample 100 million points per observed $b_k$.
- The tilted gaussian estimator $\hat{\delta}^{\text{Norm, tilt}}_{\theta,b_k}$ may only require 10 points.

# In detail: Relative s.d. of likelihood

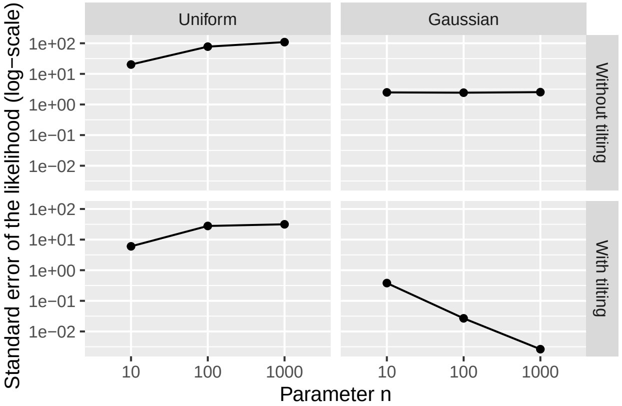



Figure: Comparison of the relative standard error of $\hat{\delta}_{\theta,b}^{\text{Unif}}$ (top left), $\hat{\delta}_{\theta,b}^{\text{Norm}}$ (top right), $\hat{\delta}_{\theta,b}^{\text{Unif},tilt}$ (bottom left), $\hat{\delta}_{\theta,b}^{\text{Norm, tilt}}$ (bottom right), for three settings with $X \sim \mathcal{M}(n, p)$ with $n = 10, 100, 1000$, $p$ uniform, $K = 100$ units, $S = 1000$ experiments, and $N_{\text{Sim}} = 10$ simulations by experiment.

# Mayor of Paris election, 2026

Main candidates:

- Grégoire (PS), 38% first round, 51% second round, elected
- Dati (LR), 25% first round, 42% second round
- Chikirou (LFI), 12% first round, 8% second round
- Bournazel (RE), 11% first round, merged with Dati for the second round
- Knafo (REC), 10% first round, withdrew in favour of Dati

Main questions of interest: behaviour of Chikirou, Bournazel, and Knafo voters.





# Mayor of Paris election, 2026

| | Grégoire | Dati | Chikirou | Abstention |
|---|---|---|---|---|
| Grégoire | 0.99 | 0.00 | 0.00 | 0.01 |
| Dati | 0.00 | 0.98 | 0.00 | 0.02 |
| Knafo | 0.00 | 0.88 | 0.00 | 0.12 |
| Bournazel | 0.44 | 0.53 | 0.00 | 0.04 |
| Chikirou | 0.44 | 0.00 | 0.53 | 0.03 |
| Abstention | 0.05 | 0.04 | 0.01 | 0.89 |

# Aim: understand leader selection in *Homo politicus*


**Jean-Luc Mélenchon**


**Emmanuel Macron**


**Marine Le Pen**

Mélenchon was eliminated in the first round of the 2022 presidential election.
How did people Mélenchon electors vote in the second round? Hypotheses:

- Hypothesis 1 (left vs. right): for Macron, because of higher proximity with the centre than with far-right.
- Hypothesis 2 (populist vs. centre): for Le Pen or no one, because of rejection of the political system.

# Particularities of the data

- "Big" row-wise: $K = 64425$ voting units, accounting for 49 million voters.
- Bigger than traditional EI analyses: 4x3, merging small candidates together
- With one GPU (running on jax), training done in 4-5 minutes for 49 million voters. With a CPU, 1–2 hours.
- We also get a stable estimator of the marginal likelihood so we can easily compute Bayes factors.

# Results: model with a covariate

Contrary to MCMC, we have access to the marginal likelihood, so we can compute Bayes factors to compare models:

$$\log_{10} BF_{3 \text{ vs. } 2} = 62058$$

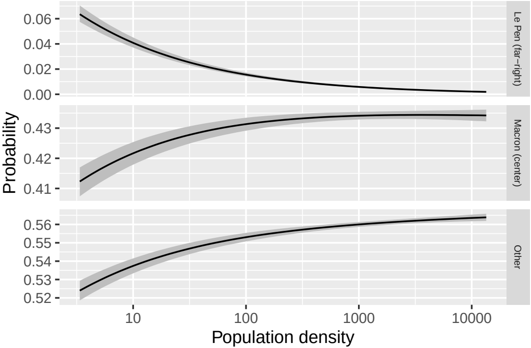
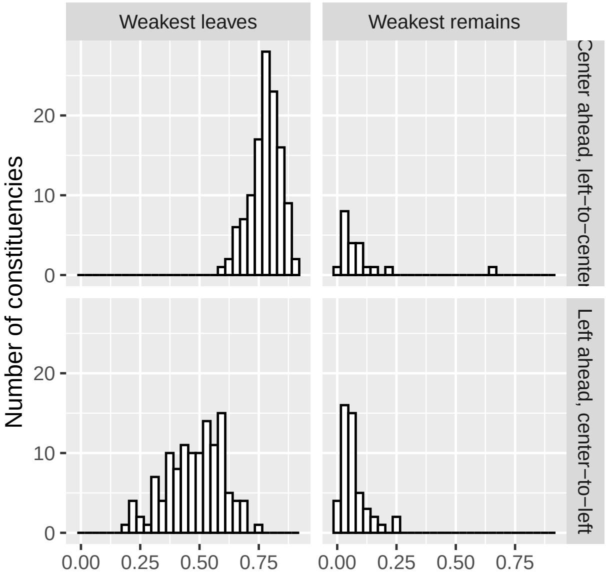



Figure: Predicted probabilities for second round behavior after voting Mélenchon (left/far-left) by population density of the city in which the voting station is located.

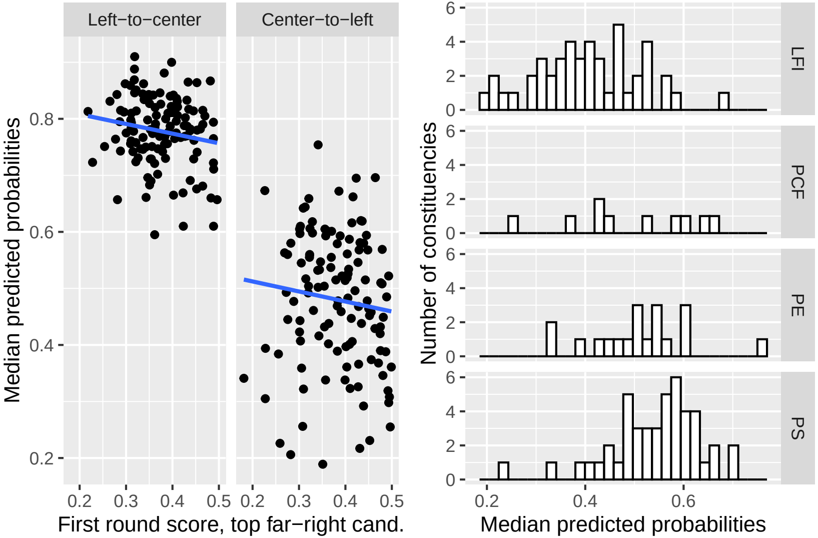
# 2024 parliamentary elections

- Snap election in June 2024.
- Three major blocks: far-right; centre; (far-)left
- Separate two-round election in each of 577 constituencies
- In round 2, there may be 2–4 candidates qualified
- Front républicain: strategic behaviour to avoid far-right win
  - Strategic withdrawal in round 2
  - With 2 candidates: do voters from eliminated/withdrawn candidates vote for a disliked candidate to avoid a far-right win?
  - With 3 candidates: do some voters switch to another candidate to avoid a far-right win?
  - Do these strategies vary based on covariates, such as candidate party or strength of far-right?

# 2024 parliamentary elections: results



# 2024 parliamentary elections: covariates



# Other applications: elections

- One-round elections (UK or US elections, first "round": religion or race)
- Three-round elections (first round could be e.g. employment status)
- Quantiles (e.g. income, WIP)

# Other applications (outside elections)

- Aggregated data (privacy)
- Contingency tables (or Latin squares)
- Network tomography
- Inverse problems?

# Recap

Main steps of our approach:

1. CF of $Y = AX$ obtained from CF of $X$: dimension reduction.

# Recap

Main steps of our approach:

1. CF of $Y = AX$ obtained from CF of $X$: dimension reduction.

2. Aggregation $\Rightarrow$ CLT $\Rightarrow$ Gaussian proposal.

# Recap

Main steps of our approach:

1. CF of $Y = AX$ obtained from CF of $X$: dimension reduction.
2. Aggregation $\Rightarrow$ CLT $\Rightarrow$ Gaussian proposal.
3. Exponential tilting to further reduce the variance.

# Recap

Main steps of our approach:

1. CF of $Y = AX$ obtained from CF of $X$: dimension reduction.
2. Aggregation $\Rightarrow$ CLT $\Rightarrow$ Gaussian proposal.
3. Exponential tilting to further reduce the variance.
4. Unbiased estimate plugged into a Metropolis sampler: pseudo-marginal sampling. Avoid Gibbs sampling.

Together, these steps allow for Ecological inference, at scale ($\sim$ 50M voters, 60K polling stations), with flexible model specifications.

# Questions

Voldoire, Chopin, Rateau & Ryder, *Saddlepoint Monte Carlo and its Application to Exact Ecological Inference*, arXiv:2410.18243. Under revision.

`https://github.com/theovoldoire/saddlepointMC-open/`

# Questions

Voldoire, Chopin, Rateau & Ryder, *Saddlepoint Monte Carlo and its Application to Exact Ecological Inference*, arXiv:2410.18243. Under revision.

`https://github.com/theovoldoire/saddlepointMC-open/`

Stubbs, Mayor of Talkeetna, Alaska from 1997 to 2017