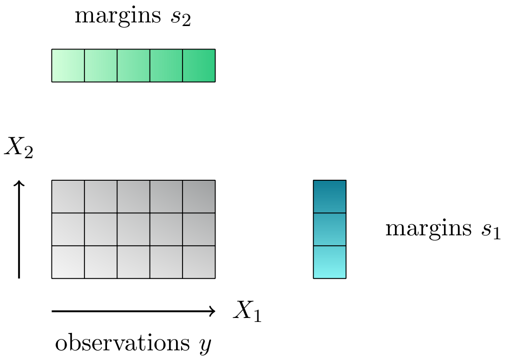
Institute for Health
Metrics and Evaluation

2025 International Conference on Continuous Optimization

Fast optimization approaches for raking

Ariane Ducellier

What is raking?



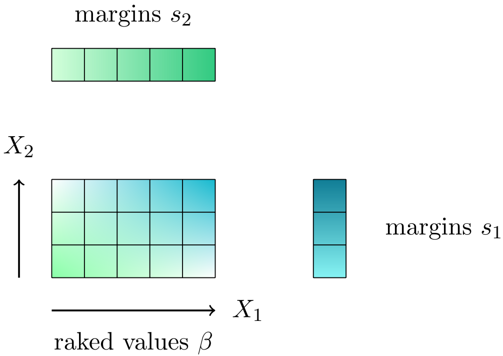
Two categorical variables X_1 and X_2 taking I and J possible values.

When summing the rows and columns of the table y, the observations y_{ij} do not add up to the values in the margins s_1 and s_2.

$$\sum_{i=1}^{I} y_{ij} \neq s_{1j} \quad j = 1, \cdots, J$$

$$\sum_{j=1}^{J} y_{ij} \neq s_{2i} \quad i = 1, \cdots, I$$

What is raking?



After raking, the raked values β_{ij} in the updated table β sum correctly to the values in the margins s_1 and s_2.

$$\sum_{i=1}^{I} \beta_{ij} = s_{1j} \quad j = 1, \cdots, J$$

$$\sum_{j=1}^{J} \beta_{ij} = s_{2i} \quad i = 1, \cdots, I$$

Note: For the problem to have a solution, we need the margins to be consistent:

$$\sum_{j=1}^{J} s_{1j} = \sum_{i=1}^{I} s_{2i}$$

Global health example

- The observation table may be the number of deaths from each cause i and each sub-region j. The margins are the number of deaths from all causes for each sub-region j and the number of deaths from each cause i for the entire region.

- For some reason (e.g. errors in data collection, the table is the output of a model that does not include the constraints on the margins), the partial sums on the observations do not match the margins.

- We trust more the margins than the observations.

Raking as an optimization problem

$y \in \mathbb{R}^p$ is the vectorized observation table.

$s \in \mathbb{R}^k$ are the known margins, i.e. the known partial sums on the table.

$A \in \mathbb{R}^{k \times p}$ summarizes how to compute the partial sums.

$\beta \in \mathbb{R}^p$ are the unknown raked values.

$w \in \mathbb{R}^p$ are raking weights chosen by the user.

f^w is a separable, derivable, positive, strictly convex function chosen by the user.

$$\min_{\beta \in \mathbb{R}^p} f^w(\beta; y) \quad \text{s.t.} \quad A\beta = s \quad \text{with} \quad f^w(\beta; y) = \sum_{i=1}^{p} w_i f_i(\beta_i, y_i)$$

Raking as an optimization problem

Examples:

1D problem: $A = \mathbb{1}_p^T$

2D problem: $A = \begin{pmatrix} I_J \otimes \mathbb{1}_I^T \\ \mathbb{1}_J^T \otimes I_I \end{pmatrix}$

Note: We need to ensure that all the constraints are consistent and we trim the redundant constraints such that rank $\left(A \in \mathbb{R}^{k \times p}\right) = k \leq p$.

Dual formulation

$$\mathcal{P}: \quad \min_{\beta \in \mathbb{R}^p} f^w(\beta, y) \quad \text{s.t} \quad A\beta = s$$

$$\mathcal{L}: \quad f^w(\beta, y) + \lambda^T (A\beta - s)$$

$$\mathcal{D}: \quad \min_{\lambda \in \mathbb{R}^k} f^{w*}(-A^T\lambda) + \lambda^T s$$

As $k \leq p$, we decrease the dimension of the problem by using the dual formulation instead of the primal formulation.

Dual formulation

Solve for λ:

$$s - A\nabla_z f^{w*}\left(-A^T\lambda\right) = 0$$

Newton's method: At each iteration, we get $\lambda^{(n+1)} = \lambda^{(n)} - \gamma\Delta\lambda^{(n)}$ with:

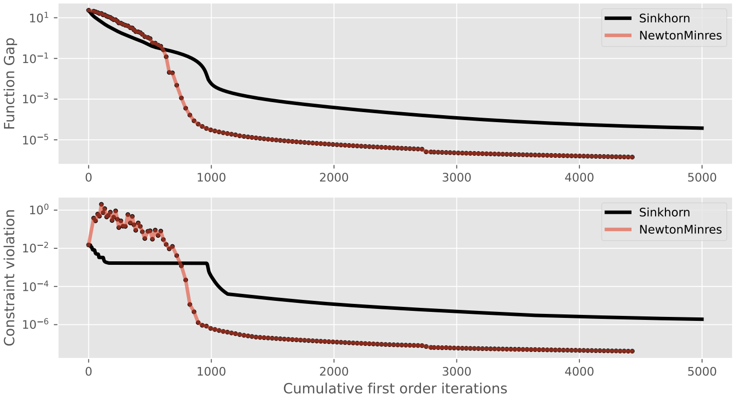
$$A\nabla_z^2 f^{w*}\left(-A^T\lambda^{(n)}\right) A^T\Delta\lambda^{(n)} = s - A\nabla_z f^{w*}\left(-A^T\lambda^{(n)}\right)$$

Final solution: $\beta^* = \nabla_z f^{w*}\left(-A^T\lambda^*\right)$

Note: A has rank k and f^w is separable so $A\nabla_z^2 f^{w*}\left(-A^T\lambda\right) A^T$ is invertible.

Comparison with Sinkhorn algorithm

We can solve the 2D problem with Sinkhorn algorithm → We get a comparable computation time with the dual formulation and Newton's method.



Common distance functions

	Distance $f_i(\beta_i; y_i)$	Solution	Note
χ^2	$\frac{1}{2 y_i} (\beta_i - y_i)^2$	$\beta^* = y \odot \left(1 - \frac{1}{w} \odot A^T \lambda^*\right)$	Solved in 1 iteration.
Entropic	$\beta_i \log\left(\frac{\beta_i}{y_i}\right) - \beta_i + y_i$	$\beta^* = y \odot \exp\left(-\frac{1}{w} \odot A^T \lambda^*\right)$	The raked values have the same sign as the initial observations.
Logit	$(\beta_i - l_i) \log \frac{\beta_i - l_i}{y_i - l_i} + (h_i - \beta_i) \log \frac{h_i - \beta_i}{h_i - y_i}$	$\beta^* = \frac{l \odot (h-y) + h \odot (y-l) \odot e^{-\frac{1}{w} \odot A^T \lambda^*}}{(h-y) + (y-l) \odot e^{-\frac{1}{w} \odot A^T \lambda^*}}$	The raked values stay between l_i and h_i when we rake prevalence observations.

Feasibility of the problem

$f_i\left(\beta_i, y_i\right)$ is only defined when $y_i \neq 0$.

Let $P \in \mathbb{R}^{\tilde{p} \times p}$ be a permutation matrix that selects the $\tilde{p} < p$ entries of y that are non-zeros.

Let $Q \in \mathbb{R}^{(p-\tilde{p}) \times p}$ be a permutation matrix that selects the $p - \tilde{p}$ entries of y that are zeros.

Let us denote:

$$\tilde{y} = Py \quad \text{and} \quad \tilde{A} = AP^T$$

We have:

$$\left(P^T P + Q^T Q\right) y = y \quad \text{and} \quad \left(P^T P + Q^T Q\right) \beta = \beta$$

Feasibility of the problem

The problem becomes:

$$\min_{\beta\in\mathbb{R}^p} f^w\left(P\beta, Py\right) \quad \text{s.t.} \quad A\beta = s \quad \text{and} \quad Q\beta = 0$$

$$\min_{\beta\in\mathbb{R}^p} f^w\left(\tilde{\beta}, \tilde{y}\right) \quad \text{s.t.} \quad A\left(P^TP + Q^TQ\right)\beta = s \quad Q\beta = 0$$

$$\min_{\beta\in\mathbb{R}^p} f^w\left(\tilde{\beta}, \tilde{y}\right) \quad \text{s.t.} \quad \left(AP^T\right)\left(P\beta\right) = s \quad Q\beta = 0$$

We get the reduced problem:

$$\mathcal{P}: \quad \min_{\tilde{\beta}\in\mathbb{R}^{\tilde{p}}} f^w\left(\tilde{\beta}, \tilde{y}\right) \quad \text{s.t.} \quad \tilde{A}\tilde{\beta} = s$$

The problem has a solution if $\tilde{A} = AP^T \in \mathbb{R}^{k\times\tilde{p}}$ has rank k or if s is in the column space of $\tilde{A}$.

Feasibility of the problem

Example:

$$\begin{pmatrix}1 & 2\end{pmatrix}$$

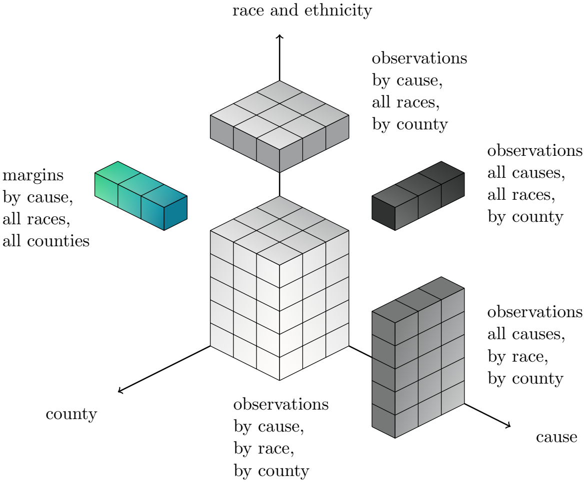
$$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix} \quad \begin{pmatrix}2\\1\end{pmatrix}$$

$$y = \begin{pmatrix}1\\0\\0\\1\end{pmatrix} \quad s = \begin{pmatrix}2\\1\\1\end{pmatrix} \quad A = \begin{pmatrix}1 & 0 & 1 & 0\\0 & 1 & 0 & 1\\1 & 1 & 0 & 0\end{pmatrix}$$

$$AP^T = \begin{pmatrix}1 & 0 & 1 & 0\\0 & 1 & 0 & 1\\1 & 1 & 0 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\0 & 0\\0 & 0\\0 & 1\end{pmatrix} = \begin{pmatrix}1 & 0\\0 & 1\\1 & 0\end{pmatrix}$$ has rank 2

and $\begin{pmatrix}2\\1\\1\end{pmatrix}$ is not in the column space of $\begin{pmatrix}1 & 0\\0 & 1\\1 & 0\end{pmatrix}$ so the problem does not have a solution.

Aggregated observations



Aggregated observations

$$\min_{\beta \in \mathbb{R}^p} f^w(B\beta, y) \quad \text{s.t.} \quad A\beta = s \quad \text{with} \quad B \in \mathbb{R}^{q\times p} \text{ an aggregation matrix}$$

With auxiliary variable $\zeta \in \mathbb{R}^q$ and additional constraint $\zeta := B\beta$, the problem becomes:

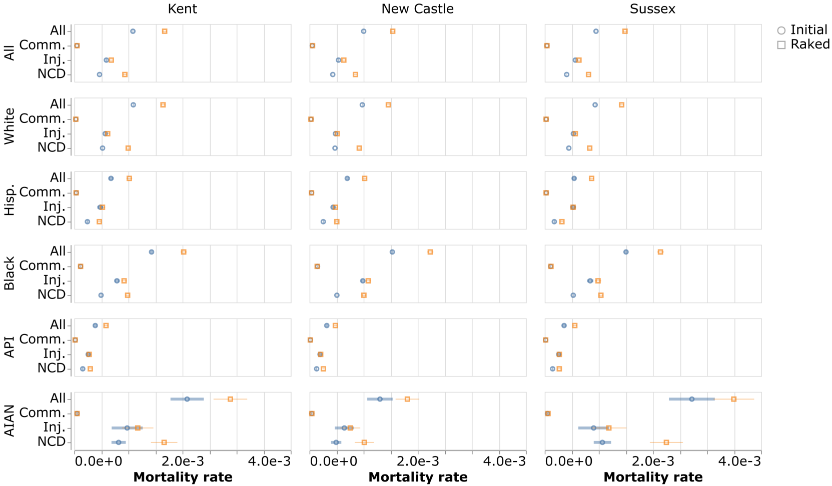
$$\mathcal{P}: \quad \min_{\beta\in\mathbb{R}^p, \zeta\in\mathbb{R}^q} f^w(\zeta, y) \quad \text{s.t.} \quad \begin{pmatrix} A & 0 \\ B & -I \end{pmatrix} \begin{pmatrix} \beta \\ \zeta \end{pmatrix} = \begin{pmatrix} s \\ 0 \end{pmatrix}$$

$$\mathcal{L}: \quad f^w(\zeta, y) + \lambda_A^T(A\beta - s) + \lambda_B^T(B\beta - \zeta)$$

$$\mathcal{D}: \quad \min_{\lambda_A\in\mathbb{R}^k, \lambda_B\in\mathbb{R}^q} \lambda_A^T s + f^{w*}(\lambda_B, y) + \delta_0\left(-A^T\lambda_A - B^T\lambda_B\right)$$

$$\text{OPT}: \quad \zeta^* = \nabla_z f^{w*}(\lambda_B^*, y), \quad \begin{pmatrix} A \\ B \end{pmatrix} \beta^* = \begin{pmatrix} s \\ \zeta^* \end{pmatrix}$$

Aggregated observations



Variance propagation

Given:

- $\Sigma_y \in \mathbb{R}^{p \times p}$, the covariance matrix of the observations vector y,
- $\Sigma_s \in \mathbb{R}^{k \times k}$, the covariance matrix of the margins vector s and
- $\Sigma_{ys} \in \mathbb{R}^{p \times k}$, the covariance matrix of y and s,

find:

- $\Sigma_{\beta^*} \in \mathbb{R}^{p \times p}$, the covariance matrix of the estimated raked values β^*.

Variance propagation

The primal problem:

$$\min_{\beta \in \mathbb{R}^p} \max_{\lambda \in \mathbb{R}^k} f^w(\beta, y) + \lambda^T (A\beta - s)$$

can also be written:

$$F(\beta, \lambda; y, s) = \begin{bmatrix} \nabla_\beta f^w(\beta, y) + A^T \lambda \\ A\beta - s \end{bmatrix} = 0$$

and has solution:

$$\beta^* = \phi(y, s) \text{ with } \phi : \mathbb{R}^{p+k} \to \mathbb{R}^p$$

Variance propagation

We get:

$$\Sigma_{\beta^*} = \phi'_{ys}\left(y, s\right) \Sigma \phi'^{T}_{ys}\left(y, s\right)$$

with:

$$\phi'_{ys}\left(y, s\right) = \begin{pmatrix} \frac{\partial \beta^*}{\partial y} & \frac{\partial \beta^*}{\partial s} \end{pmatrix} = \begin{pmatrix} \frac{\partial \phi_1}{\partial y_1}\left(y, s\right) & \dots & \frac{\partial \phi_1}{\partial y_p}\left(y, s\right) & \frac{\partial \phi_1}{\partial s_1}\left(y, s\right) & \dots & \frac{\partial \phi_1}{\partial s_k}\left(y; s\right) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial \phi_p}{\partial y_1}\left(y, s\right) & \dots & \frac{\partial \phi_p}{\partial y_p}\left(y, s\right) & \frac{\partial \phi_p}{\partial s_1}\left(y, s\right) & \dots & \frac{\partial \phi_p}{\partial s_k}\left(y, s\right) \end{pmatrix}$$

and:

$$\Sigma = \begin{pmatrix} \Sigma_y & \Sigma_{ys} \\ \Sigma_{ys}^T & \Sigma_s \end{pmatrix}$$

Variance propagation

Implicit Function Theorem: When differentiating the primal problem $F(y, s; \phi(y, s)) = 0$ at the solution (β^*, λ^*), we get:

$$\left[D_{\beta,\lambda} F(y, s; \beta^*, \lambda^*)\right] \left[D_{y,s} \phi(y, s)\right] + \left[D_{y,s} F(y, s; \beta^*, \lambda^*)\right] = 0$$

We have:

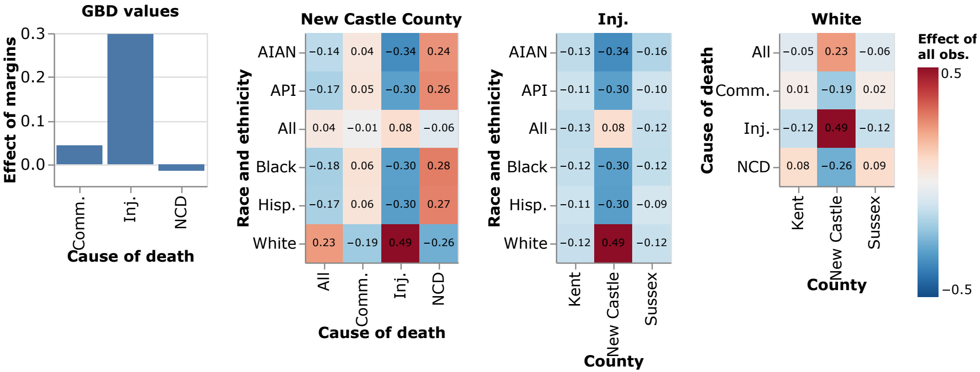
$$D_{\beta,\lambda} F = \begin{pmatrix} \nabla^2_{\beta} f^w(\beta^*, y) & A^T \\ A & 0_{k\times k} \end{pmatrix} \text{ and } D_{y,s} F = \begin{pmatrix} \nabla^2_{\beta y} f^w(\beta^*; y) & 0_{p\times k} \\ 0_{k\times p} & -I_{k\times k} \end{pmatrix}$$

thus we can compute:

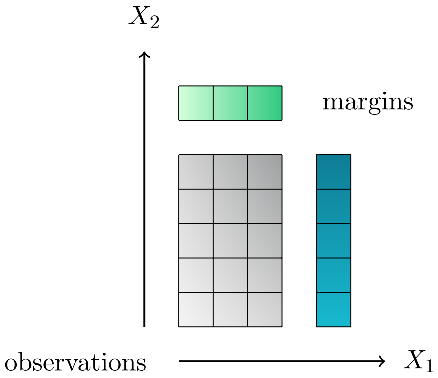
$$D_{y,s}\phi = \begin{pmatrix} \frac{\partial \beta^*}{\partial y} & \frac{\partial \beta^*}{\partial s} \\ \frac{\partial \lambda^*}{\partial y} & \frac{\partial \lambda^*}{\partial s} \end{pmatrix} \text{ and } \phi'_{ys}(y, s) = \begin{pmatrix} \frac{\partial \beta^*}{\partial y} & \frac{\partial \beta^*}{\partial s} \end{pmatrix}$$

Variance propagation

Sensitivity of the variance of the raked value with the observations and margins.



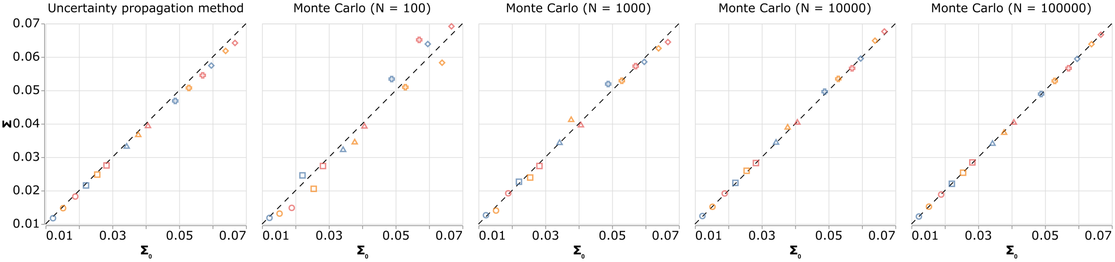
Variance propagation



- Generate 3×5 values following $\mathcal{U}nif\,([2;3])$
- Compute the margins $s_1 = \beta_0^T 1$ and $s_2 = \beta_0 1$
  $\rightarrow$ We get a balanced table β_0
- Add random noise:
  $y_0 = \beta_0 + \mathcal{N}\,(\mu = 0, \sigma = 0.1)$
- Choose covariance matrix Σ:
  - Off-diagonal elements equal to 0.01
  - Diagonal elements equal to $\Sigma_{k,k} = 0.1 \times k$ for $k = 1, 15$
- The observations y follow a MVN distribution with expectancy y_0 and covariance Σ.

Variance propagation

Comparison between the variance of the estimator β^* obtained with the variance propagation method or obtained by raking each sample and computing the sample covariance of the results.



Questions?

PyPI: https://pypi.org/project/raking/

GitHub: https://github.com/ihmeuw-msca/raking