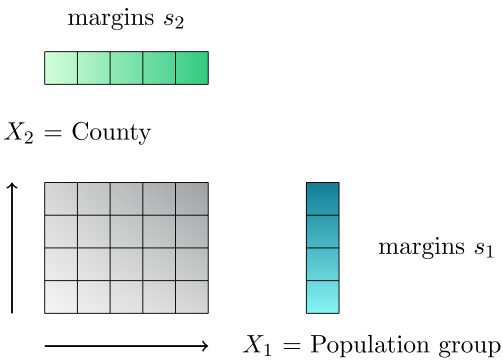
Institute for Health
Metrics and Evaluation

2025 SIAM Annual Meeting

Raking Methods and Applications to Health Metrics

Ariane Ducellier, Peng Zheng, Aleksandr Aravkin

What is raking?



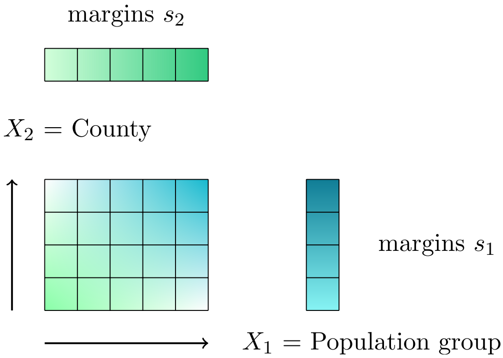
Two categorical variables X_1 and X_2 taking I and J possible values.

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 obesity cases for each population group i and each county j. The margins are the number of obesity cases for the entire population for each county j and the number of cases for each population group i for the entire state.
- 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.

$$\min_{\beta \in \mathbb{R}^p} f^w(\beta; y) \quad \text{s.t.} \quad A\beta = s \quad \text{with} \quad f^w(\beta; y) = \sum_{i=1}^{p} w_i f_i(\beta_i, y_i) \quad \text{and e.g.} \quad A = \begin{pmatrix} I_J \otimes \mathbb{1}_I^T \\ \mathbb{1}_J^T \otimes I_I \end{pmatrix}$$

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

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.

We solve the dual problem using Newton's method.

Common distance functions

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

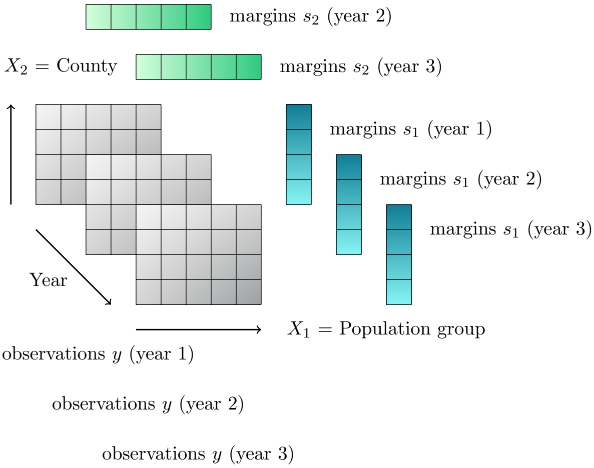
Prior ordinal constraints

We are given observations and margins for n different years $\rightarrow$ We can solve n independent raking problems:

$$\min_{\beta_1 \in \mathbb{R}^p} f^w(\beta_1, y_1) \quad \text{s.t} \quad A_1 \beta_1 = s_1$$

$$\vdots$$

$$\min_{\beta_n \in \mathbb{R}^p} f^w(\beta_n, y_n) \quad \text{s.t} \quad A_n \beta_n = s_n$$



Prior ordinal constraints

We want the raking process to preserve the sign of the trend observed between year 1 and year 2, year 2 and year 3, and so on and so forth until year n.

$(n-1)\,p$ additional constraints must then be added:

$$\begin{aligned}
&\left(\beta_{1,i}-\beta_{2,i}\right)\left(y_{1,i}-y_{2,i}\right) \geq 0 \quad \forall i = 1, \cdots, p \\
&\vdots \\
&\left(\beta_{n-1,i}-\beta_{n,i}\right)\left(y_{n-1,i}-y_{n,i}\right) \geq 0 \quad \forall i = 1, \cdots, p
\end{aligned}$$

The raking problem becomes:

$$\min_{\beta_1, \cdots, \beta_n} f^w\left(\beta_1, y_1\right) + \cdots + f^w\left(\beta_n, y_n\right)$$

$$\text{s.t.} \quad \begin{cases}
A_1 \beta_1 = s_1, \\
\cdots \\
A_n \beta_n = s_n, \\
-\left(\beta_1 - \beta_2\right) \odot \left(y_1 - y_2\right) \leq 0 \\
\cdots \\
-\left(\beta_{n-1} - \beta_n\right) \odot \left(y_{n-1} - y_n\right) \leq 0
\end{cases}$$

Prior ordinal constraints

We end up with a minimization problem with the same form as before:

Inequality constraints

$$\min_{\beta \in \mathbb{R}^{np}} f^w(\beta, y) \quad \text{s.t.} \quad \begin{cases} A\beta = s, \\ C\beta \le c \end{cases}$$

The feasible set may be empty.

Penalty

$$\min_{\beta \in \mathbb{R}^{np}} f^w(\beta, y) + L(c - C\beta, \alpha) \quad \text{s.t.} \quad A\beta = s$$

α is a penalty parameter and L can be the logistic loss:

$$L^{\text{logit}}(x) = \sum_{i=1}^{m} \log(1 + \exp(-x_i))$$

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:

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

can also be written:

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

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:

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

Variance propagation

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

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

Knowing $D_{\beta,\lambda} F$ and $D_{y,s}\phi$, we can compute:

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

We have:

$$D_{\beta,\lambda} F = \begin{pmatrix} \nabla^2_{\beta} f^w(\beta^*, y) + C^T \nabla_{x^2} L(c - C\beta^*, \alpha) C & A^T \\ A & 0_{k \times k} \end{pmatrix}$$

Variance propagation

We denote:

$$\left[A*b\right]_{ij}=\sum_{j=1}^{n}A_{ijk}b_j \quad \text{for} \quad A\in\mathbb{R}^{m\times n\times p} \quad , \quad \beta\in\mathbb{R}^{n} \quad \text{and} \quad A*b\in\mathbb{R}^{m\times p}$$

$$\left[\nabla_y A\right]_{ijk}=\frac{\partial A_{ij}\left(y\right)}{\partial y_k}$$

We get:

$$D_{y,s}F=\begin{pmatrix}\nabla^2_{\beta y}f^w\left(\beta^*;y\right)-\left[\nabla_y C^T\right]\nabla_x L\left(c-C\beta^*,\alpha\right)+C^T\nabla^2_x L\left(c-C\beta^*,\alpha\right)\left[\nabla_y C*\beta\right] & 0_{p\times k}\\ 0_{k\times p} & -I_{k\times k}\end{pmatrix}$$

Some figures

Questions?

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

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