Raking

1. General definition of the raking problem

We have a population $\mathcal{U} = \{1, \cdots, n, \cdots, N\}$. We are interested in the variable X which takes the values x_n and we want to estimate the sum:

$$\mathcal{X} = \sum_{n=1}^{N} x_n$$

To do this, we select a sample $S \subset \mathcal{U}$ and we suppose that we know the values of the $x_n, n \in S$. If we denote π_n the probability of having $n \in S$ and $d_n = \frac{1}{\pi_n}$, an estimator for $\mathcal{X}$ is:

$$\hat{\mathcal{X}} = \sum_{n \in S} \frac{x_n}{\pi_n} = \sum_{n \in S} d_n x_n$$

Note. In practice, we will often have $S = \mathcal{U}$, $\pi_n = 1$ and $d_n = 1$, but we write the formulation with S instead of $\mathcal{U}$ to keep the generality and be able to use the same formulation of the problem if one value is missing for instance.

For the raking problem, we are going to introduce M auxiliary variables and we define the vectors $\boldsymbol{y}_n = (y_{n,1}, \cdots, y_{n,m}, \cdots, y_{n,M})^T$ for $n = 1, \cdots, N$. We suppose that the vector:

$$\sum_{n=1}^{N} \boldsymbol{y}_n = \boldsymbol{y} = (y_1, \cdots, y_m, \cdots, y_M)^T$$

is known, but we do not necessarily know the $\boldsymbol{y}_n$ for all $n = 1, \cdots, N$. We want to find the raking estimator:

$$\hat{\mathcal{X}}_{raking} = \sum_{n \in S} \alpha_n x_n$$

with the constraints:

$$\sum_{n \in S} \alpha_n \boldsymbol{y}_n = \sum_{n=1}^{N} \boldsymbol{y}_n = \boldsymbol{y}$$

If we have only one categorical variable and the $x_n, n = 1, \cdots, N$ can be written as $x_i, i = 1, \cdots, I$ with $n = i$ and $N = I$, there is only $M = 1$ auxiliary variable. We can write the vector $\boldsymbol{y}_n = x_i$ and the constraint as:

$$\sum_{i \in S} \alpha_i x_i = y_1 = \mu$$

If we have two categorical variables and the $x_n, n = 1, \cdots, N$ can be written as $x_{i,j}, i = 1, \cdots, I, j = 1, \cdots, J$ with $n = (j-1)I + i$ and $N = I * J$, there are $M = I + J$ auxiliary variables. We can write the vector $\boldsymbol{y}_n = (y_{n,1}, \cdots, y_{n,I}, y_{n,I+1} \cdots, y_{n,I+J})^T$ and the constraints as:

$$\sum_{n \in S} \alpha_n y_{n,i} = y_i = \mu_{i.} \text{ for } i = 1, \cdots, I$$

$$\sum_{n \in S} \alpha_n y_{n,I+j} = y_{I+j} = \mu_{.j} \text{ for } j = 1, \cdots, J$$

The first I constraints can be written as:

$$y_{n,i} = x_{i,j} \text{ if } n \in \{i, I+i, \cdots, (J-1)I + i\}$$

$$= 0 \text{ elsewhere}$$

The next J constraints can be written as:

$$y_{n,I+j} = x_{i,j} \text{ if } n \in \{(j-1)I + 1, (j-1)I + i, \cdots, (j-1)I + I\}$$

$$= 0 \text{ elsewhere}$$

If we denote $\alpha_{i,j} = \alpha_n = \alpha_{(j-1)I+i}$, the first set of I constraints become:

$$\sum_{\substack{j=1 \\ (i,j) \in S}}^{J} \alpha_{(j-1)I+i} y_{(j-1)I+i,i} = \sum_{\substack{j=1 \\ (i,j) \in S}}^{J} \alpha_{i,j} x_{i,j} = y_i = \mu_{i.} \text{ for } i = 1, \cdots, I$$

and the second set of J constraints become:

$$\sum_{\substack{i=1 \\ (i,j) \in S}}^{I} \alpha_{(j-1)I+i} y_{(j-1)I+i,I+j} = \sum_{\substack{i=1 \\ (i,j) \in S}}^{I} \alpha_{i,j} x_{i,j} = y_{I+j} = \mu_{.j} \text{ for } j = 1, \cdots, J$$

Our objective is to find the weights α_n close to the d_n which respect the constraints:

$$\sum_{n \in S} \alpha_n \boldsymbol{y}_n = \sum_{n=1}^{N} \boldsymbol{y}_n = \boldsymbol{y}$$

For this, we need to define a distance $G(\alpha_n, d_n)$. We suppose that $G(., d_n)$ is positive, derivable and strictly convex such that $G(d_n, d_n) = 0$. We obtain the weights $\alpha_n, n \in S$ by minimizing:

$$\sum_{n \in S} \frac{G(\alpha_n, d_n)}{q_n}$$

where the q_n are coefficients which allow us to give more importance to some of the observations (e.g. they could be a function of the uncertainties σ_n).

We can write the Lagrangian:

$$\mathcal{L}(\alpha_n, n \in S, \lambda_m, m = 1, \cdots, M) = \sum_{n \in S} \frac{G(\alpha_n, d_n)}{q_n} + \boldsymbol{\lambda}^T \left(\sum_{n \in S} \alpha_n \boldsymbol{y}_n - \boldsymbol{y} \right)$$

If we write that the partial derivatives with respect to the α_n are equal to 0, we get:

$$\frac{\partial \mathcal{L}(\alpha_n, n \in S, \lambda_m, m = 1, \cdots, M)}{\partial \alpha_n} = \frac{g(\alpha_n, d_n)}{q_n} + \boldsymbol{\lambda}^T \boldsymbol{y}_n = 0$$

where:

$$g(\alpha_n, d_n) = \frac{\partial G(\alpha_n, d_n)}{\partial \alpha_n}$$

If we denote $d_n F(.)$ the inverse function of $g(., d_n)$ (that is $d_n F(u) = v$ if $g(v, d_n) = u$), we get the weights:

$$\alpha_n = d_n F\left(-q_n \boldsymbol{\lambda}^T \boldsymbol{y}_n\right)$$

Using the constraints, we get:

$$\sum_{n \in S} d_n F\left(-q_n \boldsymbol{\lambda}^T \boldsymbol{y}_n\right) \boldsymbol{y}_n = \boldsymbol{y}$$

We need to solve this last equation for $\boldsymbol{\lambda}$ to get the final values of the raking weights.

2. 1-dimensional raking

In this section, instead of computing the raking weights α_i, we compute the raked values μ_i. That will make it easier to add a condition $0 < \mu_i < 1$ on the raked values if we are interested in prevalence. Our objective is then to find the μ_i that minimize:

$$\sum_{i=1}^{I} \frac{G\left(\mu_i, x_i\right)}{q_i}$$

under the constraint:

$$\boldsymbol{v}^T \boldsymbol{\mu} = \sum_{i=1}^{I} v_i \mu_i = \mu$$

2.1. Chi-square distance

We take:

$$G\left(\mu, x\right) = \frac{1}{2x}\left(\mu - x\right)^2 \text{ and } g\left(\mu, x\right) = \frac{\partial G\left(\mu, x\right)}{\partial \mu} = \frac{\mu}{x} - 1$$

2.1.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}\left(\boldsymbol{\mu}, \lambda\right) = \sum_{i=1}^{I} \frac{1}{2q_i x_i}\left(\mu_i - x_i\right)^2 + \lambda\left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu\right)$$

If we write that the partial derivatives with respect to the μ_i are equal to 0, we get:

$$\frac{\partial \mathcal{L}\left(\boldsymbol{\mu}, \lambda\right)}{\partial \mu_i} = \frac{1}{q_i}\left(\frac{\mu_i}{x_i} - 1\right) + \lambda v_i = 0$$

First method: 1 unknown λ

We get the raked values:

$$\mu_i = x_i\left(1 - q_i v_i \lambda\right)$$

Using the constraint, we get:

$$\sum_{i=1}^{I} v_i x_i\left(1 - q_i v_i \lambda\right) = \mu$$

which gives us directly:

$$\lambda = \frac{\sum_{i=1}^{I} v_i x_i - \mu}{\sum_{i=1}^{I} q_i v_i^2 x_i}$$

Second method: $I + 1$ **unknowns** μ **and** λ

We can also write the problem as a system of $I + 1$ linear equations:

$$\frac{\mu_1}{q_1 x_1} - \frac{1}{q_1} + v_1 \lambda = 0$$

$$\dots$$

$$\frac{\mu_I}{q_I x_I} - \frac{1}{q_I} + v_I \lambda = 0$$

$$\boldsymbol{v}^T \boldsymbol{\mu} = \mu$$

which can be re-written as:

$$
\begin{pmatrix}
\frac{1}{q_1 x_1} & 0 & \dots & 0 & v_1 \\
0 & \ddots & \ddots & \vdots & \vdots \\
\vdots & \ddots & \ddots & 0 & \vdots \\
0 & \dots & 0 & \frac{1}{q_I x_I} & v_I \\
v_1 & \dots & \dots & v_I & 0
\end{pmatrix}
\begin{pmatrix}
\mu_1 \\
\vdots \\
\vdots \\
\mu_I \\
\lambda
\end{pmatrix}
=
\begin{pmatrix}
\frac{1}{q_1} \\
\vdots \\
\vdots \\
\frac{1}{q_I} \\
\mu
\end{pmatrix}
$$

2.1.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$\min_{\boldsymbol{\mu}} \sup_{\lambda} \sum_{i=1}^{I} \frac{(\mu_i - x_i)^2}{2 q_i x_i} + \lambda \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu \right) \geq$$

$$\sup_{\lambda} -\lambda \mu - \sup_{\boldsymbol{\mu}} \sum_{i=1}^{I} -\mu_i v_i \lambda - \frac{(\mu_i - x_i)^2}{2 q_i x_i} =$$

$$\sup_{\lambda} -\lambda \mu - \sum_{i=1}^{I} f_i^* \left(-v_i \lambda \right)$$

where:

$$f^* (z) = \sup_{\mu} \mu z - \frac{(\mu - x)^2}{2 q x}$$

Taking the derivative of f^* with respect to μ, we get:

5

$$\frac{\partial f^* (z)}{\partial \mu} = 0 = z - \frac{(\mu - x)}{qx}$$

which gives us:

$$\mu = x (1 + qz)$$

and:

$$f^* (z) = xz \left(1 + \frac{qz}{2}\right)$$

Its derivative is:

$$f^{*(1)} (z) = x (1 + qz)$$

The dual can then be written as a minimization problem:

$$\min_{\lambda} \lambda \mu + \sum_{i=1}^{I} f_i^* (-v_i \lambda)$$

Taking the derivative with respect to λ, we have:

$$\mu = \sum_{i=1}^{I} v_i f_i^{*(1)} (-v_i \lambda) = \sum_{i=1}^{I} v_i x_i (1 - q_i v_i \lambda)$$

which immediately gives us:

$$\mu_i = f_i^{*(1)} (-v_i \lambda) = x_i (1 - q_i v_i \lambda)$$

that is we get back to the solution of the primal problem.

2.2. $\mathcal{L}_2$ distance

We take:

$$G (\mu, x) = \frac{1}{2} (\mu - x)^2 \text{ and } g (\mu, x) = \frac{\partial G (\mu, x)}{\partial \mu} = \mu - x$$

2.2.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}\left(\boldsymbol{\mu}, \lambda\right) = \sum_{i=1}^{I} \frac{1}{2q_i} \left(\mu_i - x_i\right)^2 + \lambda \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu\right)$$

If we write that the partial derivatives with respect to the μ_i are equal to 0, we get:

$$\frac{\partial \mathcal{L}\left(\boldsymbol{\mu}, \lambda\right)}{\partial \mu_i} = \frac{1}{q_i} \left(\mu_i - x_i\right) + \lambda v_i = 0$$

First method: 1 unknown λ

We get the raked values:

$$\mu_i = x_i - q_i v_i \lambda$$

Using the constraint, we get:

$$\sum_{i=1}^{I} v_i \left(x_i - q_i v_i \lambda\right) = \mu$$

which gives us directly:

$$\lambda = \frac{\sum_{i=1}^{I} v_i x_i - \mu}{\sum_{i=1}^{I} q_i v_i^2}$$

Second method: $I + 1$ unknowns μ and λ

We can also write the problem as a system of $I + 1$ linear equations:

$$\frac{\mu_1 - x_1}{q_1} + v_1 \lambda = 0$$

$$\dots$$

$$\frac{\mu_I - x_I}{q_I} + v_I \lambda = 0$$

$$\boldsymbol{v}^T \boldsymbol{\mu} = \mu$$

which can be re-written as:

$$
\begin{pmatrix}
\frac{1}{q_1} & 0 & \dots & 0 & v_1 \\
0 & \ddots & \ddots & \vdots & \vdots \\
\vdots & \ddots & \ddots & 0 & \vdots \\
0 & \dots & 0 & \frac{1}{q_I} & v_I \\
v_1 & \dots & \dots & v_I & 0
\end{pmatrix}
\begin{pmatrix}
\mu_1 \\ \vdots \\ \vdots \\ \mu_I \\ \lambda
\end{pmatrix}
=
\begin{pmatrix}
\frac{x_1}{q_1} \\ \vdots \\ \vdots \\ \frac{x_I}{q_I} \\ \mu
\end{pmatrix}
$$

2.2.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$
\min_{\boldsymbol{\mu}} \sup_{\lambda} \sum_{i=1}^{I} \frac{(\mu_i - x_i)^2}{2q_i} + \lambda \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu \right) \geq
$$

$$
\sup_{\lambda} -\lambda\mu - \sup_{\boldsymbol{\mu}} \sum_{i=1}^{I} -\mu_i v_i \lambda - \frac{(\mu_i - x_i)^2}{2q_i} =
$$

$$
\sup_{\lambda} -\lambda\mu - \sum_{i=1}^{I} f_i^* \left(-v_i \lambda \right)
$$

where:

$$
f^*(z) = \sup_{\mu} \mu z - \frac{(\mu - x)^2}{2q}
$$

Taking the derivative of f^* with respect to μ, we get:

$$
\frac{\partial f^*(z)}{\partial \mu} = 0 = z - \frac{(\mu - x)}{q}
$$

which gives us:

$$
\mu = x + qz
$$

and:

$$
f^*(z) = z \left(x + \frac{qz}{2} \right)
$$

Its derivative is:

$$
f^{*(1)}(z) = x + qz
$$

The dual can then be written as a minimization problem:

$$\min_{\lambda} \lambda\mu + \sum_{i=1}^{I} f_i^* \left(-v_i\lambda\right)$$

Taking the derivative with respect to λ, we have:

$$\mu = \sum_{i=1}^{I} v_i f_i^{*(1)}\left(-v_i\lambda\right) = \sum_{i=1}^{I} v_i \left(x_i - q_i v_i \lambda\right)$$

which immediately gives us:

$$\mu_i = f_i^{*(1)}\left(-v_i\lambda\right) = x_i - q_i v_i \lambda$$

that is we get back to the solution of the primal problem.

2.3. Entropic distance

We take:

$$G\left(\mu, x\right) = \mu\log\left(\frac{\mu}{x}\right) - \mu + x \text{ and } g\left(\mu, x\right) = \frac{\partial G\left(\mu, x\right)}{\partial\mu} = \log\left(\frac{\mu}{x}\right)$$

2.3.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}\left(\boldsymbol{\mu}, \lambda\right) = \sum_{i=1}^{I} \frac{1}{q_i}\left(\mu_i\log\left(\frac{\mu_i}{x_i}\right) - \mu_i + x_i\right) + \lambda\left(\boldsymbol{v}^T\boldsymbol{\mu} - \mu\right)$$

If we write that the partial derivatives with respect to the μ_i are equal to 0, we get:

$$\frac{\partial\mathcal{L}\left(\boldsymbol{\mu}, \lambda\right)}{\partial\mu_i} = \frac{1}{q_i}\log\left(\frac{\mu_i}{x_i}\right) + \lambda v_i = 0$$

First method: 1 unknown λ

We get the raked values:

$$\mu_i = x_i \exp\left(-q_i v_i \lambda\right)$$

Using the constraint, we get:

$$\sum_{i=1}^{I} v_i x_i \exp\left(-q_i v_i \lambda\right) = \mu$$

We use Newton's method to solve this last equation for λ. We start at $\lambda^0 = 0$ and we iterate:

$$\lambda^{k+1} = \lambda^k - \frac{\mu - \sum_{i=1}^{I} v_i x_i \exp\left(-q_i v_i \lambda^k\right)}{\sum_{i=1}^{I} q_i v_i^2 x_i \exp\left(-q_i v_i \lambda^k\right)}$$

In the special case where $v_i = 1$ and $q_i = 1$, we get:

$$\mu_i = x_i \frac{\mu}{\sum_{i=1}^{n} x_i}$$

Second method: $I + 1$ unknowns μ and λ

We can also write the problem as a system of $I + 1$ linear equations:

$$\frac{1}{q_1} \log\left(\frac{\mu_1}{x_1}\right) + v_1 \lambda = 0$$

$$\dots$$

$$\frac{1}{q_I} \log\left(\frac{\mu_I}{x_I}\right) + v_I \lambda = 0$$

$$\boldsymbol{v}^T \boldsymbol{\mu} - \mu = 0$$

Using Newton's method, we have:

$$\begin{pmatrix} \mu_1^{k+1} \\ \vdots \\ \mu_I^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \mu_1^k \\ \vdots \\ \mu_I^k \\ \lambda^k \end{pmatrix} - \left[\nabla \boldsymbol{f}\left(\boldsymbol{\mu}^k, \lambda^k\right)\right]^{-1} \boldsymbol{f}\left(\boldsymbol{\mu}^k, \lambda^k\right)$$

where:

$$\boldsymbol{f}(\boldsymbol{\mu}, \lambda) = \begin{pmatrix} \frac{1}{q_1} \log\left(\frac{\mu_1}{x_1}\right) + v_1 \lambda \\ \vdots \\ \frac{1}{q_I} \log\left(\frac{\mu_I}{x_I}\right) + v_I \lambda \\ \boldsymbol{v}^T \boldsymbol{\mu} - \mu \end{pmatrix}$$

We have:

$$\frac{\partial}{\partial \mu_j}\left(\frac{1}{q_i}\log\left(\frac{\mu_i}{x_i}\right)+v_i\lambda\right)=\begin{cases}0 & \text{if } i\neq j\\ \frac{1}{q_i\mu_i} & \text{if } i=j\end{cases}$$

$$\frac{\partial}{\partial \lambda}\left(\frac{1}{q_i}\log\left(\frac{\mu_i}{x_i}\right)+v_i\lambda\right)=v_i$$

$$\frac{\left(\partial \boldsymbol{v}^T\boldsymbol{\mu}-\mu\right)}{\partial \mu_i}=v_i$$

$$\frac{\left(\partial \boldsymbol{v}^T\boldsymbol{\mu}-\mu\right)}{\partial \lambda}=0$$

thus the iterations become:

$$\begin{pmatrix}\mu_1^{k+1}\\ \vdots\\ \mu_I^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix}\mu_1^k\\ \vdots\\ \mu_I^k\\ \lambda^k\end{pmatrix}-\begin{pmatrix}\frac{1}{q_1\mu_1^k} & 0 & \cdots & 0 & v_1\\ 0 & \ddots & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & 0 & \vdots\\ 0 & \cdots & 0 & \frac{1}{q_I\mu_I^k} & v_I\\ v_1 & \cdots & \cdots & v_I & 0\end{pmatrix}^{-1}\begin{pmatrix}\frac{1}{q_1}\log\left(\frac{\mu_1^k}{x_1}\right)+v_1\lambda^k\\ \vdots\\ \frac{1}{q_I}\log\left(\frac{\mu_I^k}{x_I}\right)+v_I\lambda^k\\ \boldsymbol{v}^T\boldsymbol{\mu}^k-\mu\end{pmatrix}$$

2.3.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$\min_{\boldsymbol{\mu}}\sup_{\lambda}\sum_{i=1}^{I}\frac{1}{q_i}\left(\mu_i\log\left(\frac{\mu_i}{x_i}\right)-\mu_i+x_i\right)+\lambda\left(\boldsymbol{v}^T\boldsymbol{\mu}-\mu\right)\geq$$

$$\sup_{\lambda}-\lambda\mu-\sup_{\boldsymbol{\mu}}\sum_{i=1}^{I}-\mu_i v_i\lambda-\frac{1}{q_i}\left(\mu_i\log\left(\frac{\mu_i}{x_i}\right)-\mu_i+x_i\right)=$$

$$\sup_{\lambda}-\lambda\mu-\sum_{i=1}^{I}f_i^*\left(-v_i\lambda\right)$$

where:

$$f^*\left(z\right)=\sup_{\mu}\mu z-\frac{1}{q}\left(\mu\log\left(\frac{\mu}{x}\right)-\mu+x\right)$$

Taking the derivative of f^* with respect to μ, we get:

$$\frac{\partial f^*\left(z\right)}{\partial \mu} = 0 = z - \frac{1}{q}\log\left(\frac{\mu}{x}\right)$$

which gives us:

$$\mu = x\exp\left(qz\right)$$

and:

$$f^*\left(z\right) = \frac{x}{q}\exp\left(qz\right)$$

Its derivative is:

$$f^{*(1)}\left(z\right) = x\exp\left(qz\right)$$

The dual can then be written as a minimization problem:

$$\min_{\lambda} \lambda\mu + \sum_{i=1}^{I} f_i^*\left(-v_i\lambda\right)$$

Taking the derivative with respect to λ, we have:

$$\mu = \sum_{i=1}^{I} v_i f_i^{*(1)}\left(-v_i\lambda\right) = \sum_{i=1}^{I} v_i x_i\exp\left(-q_i v_i\lambda\right)$$

which immediately gives us:

$$\mu_i = f_i^{*(1)}\left(-v_i\lambda\right) = x_i\exp\left(-q_i v_i\lambda\right)$$

that is we get back to the solution of the primal problem.

2.4. A more general distance

We take:

$$G\left(\mu,x\right) = \frac{1}{\alpha}\left[\frac{x}{\alpha+1}\left(\frac{\mu}{x}\right)^{\alpha+1} - \mu + c\right] \text{ and } g\left(\mu,x\right) = \frac{\partial G\left(\mu,x\right)}{\partial \mu} = \frac{1}{\alpha}\left[\left(\frac{\mu}{x}\right)^{\alpha} - 1\right]$$

2.4.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}(\boldsymbol{\mu}, \lambda) = \sum_{i=1}^{I} \frac{1}{\alpha q_i} \left[\frac{x_i}{\alpha + 1} \left(\frac{\mu_i}{x_i} \right)^{\alpha+1} - \mu_i + c_i \right] + \lambda \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu \right)$$

If we write that the partial derivatives with respect to the μ_i are equal to 0, we get:

$$\frac{\partial \mathcal{L}(\boldsymbol{\mu}, \lambda)}{\partial \mu_i} = \frac{1}{\alpha q_i} \left[\left(\frac{\mu_i}{x_i} \right)^{\alpha} - 1 \right] + \lambda v_i = 0$$

First method: 1 unknown λ

We get the raked values:

$$\mu_i = x_i \left(1 - \alpha q_i v_i \lambda \right)^{\frac{1}{\alpha}}$$

Using the constraint, we get:

$$\sum_{i=1}^{I} v_i x_i \left(1 - \alpha q_i v_i \lambda \right)^{\frac{1}{\alpha}} = \mu$$

We use Newton's method to solve this last equation for λ. We start at $\lambda^0 = 0$ and we iterate:

$$\lambda^{k+1} = \lambda^k - \frac{\mu - \sum_{i=1}^{I} v_i x_i \left(1 - \alpha q_i v_i \lambda^k \right)^{\frac{1}{\alpha}}}{\sum_{i=1}^{I} q_i v_i^2 x_i \left(1 - \alpha q_i v_i \lambda^k \right)^{\frac{1}{\alpha} - 1}}$$

This adds a condition on λ for $\alpha < -1$ or $\alpha > \frac{1}{2}$:

$$\alpha q_i v_i \lambda < 1 \text{ for all } i$$

Second method: $I + 1$ unknowns μ and λ

We can also write the problem as a system of $I + 1$ linear equations:

$$\frac{1}{q_1 \alpha} \left[\left(\frac{\mu_1}{x_1} \right)^{\alpha} - 1 \right] + v_1 \lambda = 0$$

$$\cdots$$

$$\frac{1}{q_I \alpha} \left[\left(\frac{\mu_I}{x_I} \right)^{\alpha} - 1 \right] + v_I \lambda = 0$$

$$\boldsymbol{v}^T \boldsymbol{\mu} - \mu = 0$$

Using Newton's method, we have:

$$\begin{pmatrix} \mu_1^{k+1} \\ \vdots \\ \mu_I^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \mu_1^{k} \\ \vdots \\ \mu_I^{k} \\ \lambda^{k} \end{pmatrix} - \left[\nabla \boldsymbol{f} \left(\boldsymbol{\mu}^k, \lambda^k \right) \right]^{-1} \boldsymbol{f} \left(\boldsymbol{\mu}^k, \lambda^k \right)$$

where:

$$\boldsymbol{f}(\boldsymbol{\mu}, \lambda) = \begin{pmatrix} \frac{1}{q_1 \alpha} \left[\left(\frac{\mu_1}{x_1} \right)^{\alpha} - 1 \right] + v_1 \lambda \\ \vdots \\ \frac{1}{q_I \alpha} \left[\left(\frac{\mu_I}{x_I} \right)^{\alpha} - 1 \right] + v_I \lambda \\ \boldsymbol{v}^T \boldsymbol{\mu} - \mu \end{pmatrix}$$

We have:

$$\frac{\partial}{\partial \mu_j} \left(\frac{1}{q_i \alpha} \left[\left(\frac{\mu_i}{x_i} \right)^{\alpha} - 1 \right] + v_i \lambda \right) = \begin{cases} 0 & \text{if } i \neq j \\ \frac{1}{q_i x_i} \left(\frac{\mu_i}{x_i} \right)^{\alpha - 1} & \text{if } i = j \end{cases}$$

$$\frac{\partial}{\partial \lambda} \left(\frac{1}{q_i \alpha} \left[\left(\frac{\mu_i}{x_i} \right)^{\alpha} - 1 \right] + v_i \lambda \right) = v_i$$

$$\frac{\left(\partial \boldsymbol{v}^T \boldsymbol{\mu} - \mu \right)}{\partial \mu_i} = v_i$$

$$\frac{\left(\partial \boldsymbol{v}^T \boldsymbol{\mu} - \mu \right)}{\partial \lambda} = 0$$

thus the iterations become:

$$
\begin{pmatrix} \mu_1^{k+1} \\ \vdots \\ \mu_I^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \mu_1^{k} \\ \vdots \\ \mu_I^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \frac{1}{q_1 x_1}\left(\frac{\mu_1^k}{x_1}\right)^{\alpha-1} & 0 & \cdots & 0 & v_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \frac{1}{q_I x_I}\left(\frac{x_I}{x_I}\right)^{\alpha-1} & v_I \\ v_1 & \cdots & \cdots & v_I & 0 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{q_1 \alpha}\left[\left(\frac{\mu_1^k}{x_1}\right)^{\alpha}-1\right]+v_1\lambda^k \\ \vdots \\ \frac{1}{q_I \alpha}\left[\left(\frac{\mu_I^k}{x_I}\right)^{\alpha}-1\right]+v_I\lambda^k \\ \boldsymbol{v}^T\boldsymbol{\mu}^k - \mu \end{pmatrix}
$$

2.4.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$
\min_{\boldsymbol{\mu}} \sup_{\lambda} \sum_{i=1}^{I} \frac{1}{\alpha q_i}\left[\frac{x_i}{\alpha+1}\left(\frac{\mu_i}{x_i}\right)^{\alpha+1}-\mu_i+c_i\right]+\lambda\left(\boldsymbol{v}^T\boldsymbol{\mu}-\mu\right) \geq
$$

$$
\sup_{\lambda} -\lambda\mu - \sup_{\boldsymbol{\mu}} \sum_{i=1}^{I} -\mu_i v_i \lambda - \frac{1}{\alpha q_i}\left[\frac{x_i}{\alpha+1}\left(\frac{\mu_i}{x_i}\right)^{\alpha+1}-\mu_i+c_i\right]=
$$

$$
\sup_{\lambda} -\lambda\mu - \sum_{i=1}^{I} f_i^*\left(-v_i\lambda\right)
$$

where:

$$
f^*(z) = \sup_{\mu} \mu z - \frac{1}{\alpha q}\left[\frac{x}{\alpha+1}\left(\frac{\mu}{x}\right)^{\alpha+1}-\mu+c\right]
$$

Taking the derivative of f^* with respect to μ, we get:

$$
\frac{\partial f^*(z)}{\partial \mu}=0=z-\frac{1}{\alpha q}\left[\left(\frac{\mu}{x}\right)^{\alpha}-1\right]
$$

which gives us:

$$
\mu = x\left(1+\alpha q z\right)^{\frac{1}{\alpha}}
$$

and:

$$
f^*(z) = \frac{x}{q\left(\alpha+1\right)}\left(1+\alpha q z\right)^{\frac{1}{\alpha}+1}
$$

Its derivative is:

$$f^{*(1)}(z) = x\left(1 + \alpha qz\right)^{\frac{1}{\alpha}}$$

The dual can then be written as a minimization problem:

$$\min_{\lambda} \lambda\mu + \sum_{i=1}^{I} f_i^*\left(-v_i\lambda\right)$$

Taking the derivative with respect to λ, we have:

$$\mu = \sum_{i=1}^{I} v_i f_i^{*(1)}\left(-v_i\lambda\right) = \sum_{i=1}^{I} v_i x_i \left(1 - \alpha q_i v_i \lambda\right)^{\frac{1}{\alpha}}$$

which immediately gives us:

$$\mu_i = f_i^{*(1)}\left(-v_i\lambda\right) = x_i \left(1 - q_i v_i \lambda\right)^{\frac{1}{\alpha}}$$

that is we get back to the solution of the primal problem.

2.5. Logit distance

We want to make sure that all the raked values μ are bounded by:

$$l_i < \mu_i < h_i$$

where $\boldsymbol{l}$ and $\boldsymbol{h}$ are known vectors of length I. For example, we could use $\boldsymbol{l} = \boldsymbol{0}$ for the lower bound and the total population number for the upper bound. We assume that all the data respect the condition $l_i < x_i < h_i$. The distance function becomes:

$$G\left(\mu, x\right) = \left(\mu - l\right)\log\frac{\mu - l}{x - l} + \left(h - \mu\right)\log\frac{h - \mu}{h - x}$$

Its derivative is:

$$g\left(\mu, x\right) = \log\frac{\mu - l}{x - l} - \log\frac{h - \mu}{h - x}$$

2.5.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}\left(\boldsymbol{\mu}, \lambda\right) = \sum_{i=1}^{I} \frac{\mu_i - l_i}{q_i} \log \frac{\mu_i - l_i}{x_i - l_i} + \frac{h_i - \mu_i}{q_i} \log \frac{h_i - \mu_i}{h_i - x_i} + \lambda\left(\boldsymbol{v}^T\boldsymbol{\mu} - \mu\right)$$

If we write that the partial derivatives with respect to the μ_i are equal to 0, we get:

$$\frac{\partial \mathcal{L}\left(\boldsymbol{\mu}, \lambda\right)}{\partial \mu_i} = \frac{1}{q_i}\left(\log \frac{\mu_i - l_i}{x_i - l_i} - \log \frac{h_i - \mu_i}{h_i - x_i}\right) + \lambda v_i = 0$$

First method: 1 unknown λ

We get the raked values:

$$\mu_i = \frac{l_i\left(h_i - x_i\right) + h_i\left(x_i - l_i\right) e^{-q_i v_i \lambda}}{\left(h_i - x_i\right) + \left(x_i - l_i\right) e^{-q_i v_i \lambda}}$$

Using the constraint, we get:

$$\sum_{i=1}^{I} v_i \frac{l_i\left(h_i - x_i\right) + h_i\left(x_i - l_i\right) e^{-q_i v_i \lambda}}{\left(h_i - x_i\right) + \left(x_i - l_i\right) e^{-q_i v_i \lambda}} = \mu$$

We use Newton's method to solve this last equation for λ. We start at $\lambda^0 = 0$ and we iterate:

$$\lambda^{k+1} = \lambda^k - \frac{\mu - \sum_{i=1}^{I} v_i \frac{l_i(h_i - x_i) + h_i(x_i - l_i) e^{-q_i v_i \lambda}}{(h_i - x_i) + (x_i - l_i) e^{-q_i v_i \lambda}}}{\sum_{i=1}^{I} q_i v_i \frac{(h_i - l_i)(x_i - l_i)(h_i - x_i) e^{-q_i v_i \lambda}}{\left[(h_i - x_i) + (x_i - l_i) e^{-q_i v_i \lambda}\right]^2}}$$

Second method: $I + 1$ unknowns μ and λ

We can also write the problem as a system of $I + 1$ linear equations:

$$\frac{1}{q_1}\left(\log \frac{\mu_1 - l_1}{x_1 - l_1} - \log \frac{h_1 - \mu_1}{h_1 - x_1}\right) + v_1 \lambda = 0$$

$$\dots$$

$$\frac{1}{q_I}\left(\log \frac{\mu_I - l_I}{x_I - l_I} - \log \frac{h_I - \mu_I}{h_I - x_I}\right) + v_I \lambda = 0$$

$$\boldsymbol{v}^T\boldsymbol{\mu} - \mu = 0$$

Using Newton's method, we have:

$$
\begin{pmatrix} \mu_1^{k+1} \\ \vdots \\ \mu_I^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \mu_1^{k} \\ \vdots \\ \mu_I^{k} \\ \lambda^{k} \end{pmatrix} - \left[\nabla \boldsymbol{f} \left(\boldsymbol{\mu}^k, \lambda^k \right) \right]^{-1} \boldsymbol{f} \left(\boldsymbol{\mu}^k, \lambda^k \right)
$$

where:

$$
\boldsymbol{f} \left(\boldsymbol{\mu}, \lambda \right) = \begin{pmatrix} \frac{1}{q_1} \left(\log \frac{\mu_1 - l_1}{x_1 - l_1} - \log \frac{h_1 - \mu_1}{h_1 - x_1} \right) + v_1 \lambda \\ \vdots \\ \frac{1}{q_I} \left(\log \frac{\mu_I - l_I}{x_I - l_I} - \log \frac{h_I - \mu_I}{h_I - x_I} \right) + v_I \lambda \\ \boldsymbol{v}^T \boldsymbol{\mu} - \mu \end{pmatrix}
$$

We have:

$$
\frac{\partial}{\partial \mu_j} \left(\frac{1}{q_i} \left(\log \frac{\mu_i - l_i}{x_i - l_i} - \log \frac{h_i - \mu_i}{h_i - x_i} \right) + v_i \lambda \right) = \begin{cases} 0 & \text{if } i \neq j \\ \frac{1}{q_i} \left(\frac{1}{\mu_i - l_i} + \frac{1}{h_i - \mu_i} \right) & \text{if } i = j \end{cases}
$$

$$
\frac{\partial}{\partial \lambda} \left(\frac{1}{q_i} \left(\log \frac{\mu_i - l_i}{x_i - l_i} - \log \frac{h_i - \mu_i}{h_i - x_i} \right) + v_i \lambda \right) = v_i
$$

$$
\frac{\partial \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu \right)}{\partial \mu_i} = v_i
$$

$$
\frac{\partial \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu \right)}{\partial \lambda} = 0
$$

thus the iterations become:

$$
\begin{pmatrix} \mu_1^{k+1} \\ \vdots \\ \mu_I^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \mu_1^{k} \\ \vdots \\ \mu_I^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \frac{1}{q_1} \left(\frac{1}{\mu_1^k - l_1} + \frac{1}{h_1 - \mu_1^k} \right) & 0 & \cdots & 0 & v_1 \\ 0 & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \frac{1}{q_I} \left(\frac{1}{\mu_I^k - l_I} + \frac{1}{h_I - \mu_I^k} \right) & v_I \\ v_1 & \cdots & \cdots & v_I & 0 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{q_1} \left(\log \frac{\mu_1^k - l_1}{x_1 - l_1} - \log \frac{h_1 - \mu_1^k}{h_1 - x_1} \right) + v_1 \lambda^k \\ \vdots \\ \frac{1}{q_I} \left(\log \frac{\mu_I^k - l_I}{x_I - l_I} - \log \frac{h_I - \mu_I^k}{h_I - x_I} \right) + v_I \lambda^k \\ \boldsymbol{v}^T \boldsymbol{\mu}^k - \mu \end{pmatrix}
$$

2.5.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$\min_{\boldsymbol{\mu}} \sup_{\lambda} \sum_{i=1}^{I} \frac{1}{q_i} \left[(\mu_i - l_i) \log \frac{\mu_i - l_i}{x_i - l_i} + (h_i - \mu_i) \log \frac{h_i - \mu_i}{h_i - x_i} \right] + \lambda \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu \right) \geq$$

$$\sup_{\lambda} -\lambda\mu - \sup_{\boldsymbol{\mu}} \sum_{i=1}^{I} -\mu_i v_i \lambda - \frac{1}{q_i} \left[(\mu_i - l_i) \log \frac{\mu_i - l_i}{x_i - l_i} + (h_i - \mu_i) \log \frac{h_i - \mu_i}{h_i - x_i} \right] =$$

$$\sup_{\lambda} -\lambda\mu - \sum_{i=1}^{I} f_i^* \left(-v_i \lambda \right)$$

where:

$$f^* (z) = \sup_{\mu} \mu z - \frac{1}{q} \left[(\mu - l) \log \frac{\mu - l}{x - l} + (h - \mu) \log \frac{h - \mu}{h - x} \right]$$

Taking the derivative of f^* with respect to μ, we get:

$$\frac{\partial f^* (z)}{\partial \mu} = 0 = z - \frac{1}{q} \left[\log \frac{\mu - l}{x - l} - \log \frac{h - \mu}{h - x} \right]$$

which gives us:

$$\mu = \frac{l (h - x) + h (x - l) e^{qz}}{(h - x) + (x - l) e^{qz}}$$

and:

$$f^* (z) = zF (qz) - \frac{G (F (qz), x)}{q}$$

where:

$$F (u) = \frac{l (h - x) + h (x - l) e^{u}}{(h - x) + (x - l) e^{u}}$$

Its derivative is:

$$f^{*(1)} (z) = F (qz) = \frac{l (h - x) + h (x - l) e^{qz}}{(h - x) + (x - l) e^{qz}}$$

The dual can then be written as a minimization problem:

$$\min_{\lambda} \lambda\mu + \sum_{i=1}^{I} f_i^* \left(-v_i \lambda \right)$$

Taking the derivative with respect to λ, we have:

$$\mu = \sum_{i=1}^{I} v_i f_i^{*(1)} \left(-v_i \lambda\right) = \sum_{i=1}^{I} v_i \frac{l\left(h_i - x_i\right) + h_i\left(x_i - l_i\right) e^{-q_i v_i \lambda}}{\left(h_i - x_i\right) + \left(x_i - l_i\right) e^{-q_i v_i \lambda}}$$

which immediately gives us:

$$\mu_i = f_i^{*(1)} \left(-v_i \lambda\right) = \frac{l\left(h_i - x_i\right) + h_i\left(x_i - l_i\right) e^{-q_i v_i \lambda}}{\left(h_i - x_i\right) + \left(x_i - l_i\right) e^{-q_i v_i \lambda}}$$

that is we get back to the solution of the primal problem.

2.6. General distance

2.6.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}\left(\boldsymbol{\mu}, \lambda\right) = \sum_{i=1}^{I} \frac{G\left(\mu_i, x_i\right)}{q_i} + \lambda \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu\right)$$

If we write that the partial derivatives with respect to the μ_i are equal to 0, we get:

$$\frac{\partial \mathcal{L}\left(\boldsymbol{\mu}, \lambda\right)}{\partial \mu_i} = \frac{g\left(\mu_i, x_i\right)}{q_i} + \lambda v_i = 0$$

where:

$$g\left(\mu_i, x_i\right) = \frac{\partial G\left(\mu_i, x_i\right)}{\partial \mu_i}$$

First method: 1 unknown λ

If we denote $F\left(., x_i\right)$ the inverse function of $g\left(., x_i\right)$ (that is $F\left(u, x_i\right) = v$ if $g_i\left(v, x_i\right) = u$), we get the raked values:

$$\mu_i = F\left(-q_i v_i \lambda, x_i\right)$$

Using the constraint, we get:

$$\sum_{i=1}^{I} v_i F\left(-q_i v_i \lambda, x_i\right) = \mu$$

We use Newton's method to solve this last equation for λ. We start at $\lambda^0 = 0$ and we iterate:

$$\lambda^{k+1} = \lambda^k - \frac{\mu - \sum_{i=1}^{I} v_i F\left(-q_i v_i \lambda^k, x_i\right)}{\sum_{i=1}^{I} q_i v_i^2 F'\left(-q_i v_i \lambda^k, x_i\right)}$$

Second method: $I + 1$ unknowns μ and λ

We can also write the problem as a system of $I + 1$ linear equations:

$$\frac{g\left(\mu_1, x_1\right)}{q_1} + v_1 \lambda = 0$$

$$\ldots$$

$$\frac{g\left(\mu_I, x_I\right)}{q_I} + v_I \lambda = 0$$

$$\boldsymbol{v}^T \boldsymbol{\mu} - \mu = 0$$

Using Newton's method, we have:

$$\begin{pmatrix} \mu_1^{k+1} \\ \vdots \\ \mu_I^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \mu_1^{k} \\ \vdots \\ \mu_I^{k} \\ \lambda^{k} \end{pmatrix} - \left[\nabla \boldsymbol{f}\left(\boldsymbol{\mu}^k, \lambda^k\right)\right]^{-1} \boldsymbol{f}\left(\boldsymbol{\mu}^k, \lambda^k\right)$$

where:

$$\boldsymbol{f}\left(\boldsymbol{\mu}, \lambda\right) = \begin{pmatrix} \frac{g(\mu_1, x_1)}{q_1} + v_1 \lambda \\ \vdots \\ \frac{g(\mu_I, x_I)}{q_I} + v_I \lambda \\ \boldsymbol{v}^T \boldsymbol{\mu} - \mu \end{pmatrix}$$

We have:

$$\frac{\partial \left(\frac{g(\mu_i, x_i)}{q_i} + v_i \lambda\right)}{\partial \mu_j} = \begin{cases} 0 & \text{if } i \neq j \\ \frac{g'(\mu_i, x_i)}{q_i} & \text{if } i = j \end{cases}$$

$$\frac{\partial \left(\frac{g(\mu_i, x_i)}{q_i} + v_i \lambda\right)}{\partial \lambda} = v_i$$

$$\frac{\partial \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu\right)}{\partial \mu_i} = v_i$$

21

$$\frac{\partial \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu \right)}{\partial \lambda} = 0$$

thus the iterations become:

$$\begin{pmatrix} \mu_1^{k+1} \\ \vdots \\ \mu_I^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \mu_1^k \\ \vdots \\ \mu_I^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \frac{g'\left(\mu_1^k, x_1\right)}{q_1} & 0 & \cdots & 0 & v_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \frac{g'\left(\mu_I^k, x_I\right)}{q_I} & v_I \\ v_1 & \cdots & \cdots & v_I & 0 \end{pmatrix}^{-1} \begin{pmatrix} \frac{g\left(\mu_1^k, x_1\right)}{q_1} + v_1 \lambda^k \\ \vdots \\ \frac{g\left(\mu_I^k, x_I\right)}{q_I} + v_I \lambda^k \\ \boldsymbol{v}^T \boldsymbol{\mu}^k - \mu \end{pmatrix}$$

2.6.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$\min_{\boldsymbol{\mu}} \sup_{\lambda} \sum_{i=1}^{I} \frac{G\left(\mu_i, x_i\right)}{q_i} + \lambda \left(\boldsymbol{v}^T \boldsymbol{\mu} - \mu \right) \geq$$

$$\sup_{\lambda} -\lambda\mu - \sup_{\boldsymbol{\mu}} \sum_{i=1}^{I} -\mu_i v_i \lambda - \frac{G\left(\mu_i, x_i\right)}{q_i} =$$

$$\sup_{\lambda} -\lambda\mu - \sum_{i=1}^{I} f_i^*\left(-v_i \lambda\right)$$

where:

$$f^*\left(z\right) = \sup_{\mu} \mu z - \frac{G\left(\mu, x\right)}{q}$$

Taking the derivative of f^* with respect to μ, we get:

$$\frac{\partial f^*\left(z\right)}{\partial \mu} = 0 = z - \frac{g\left(\mu, x\right)}{q}$$

which gives us:

$$qz = g\left(\mu, x\right)$$

that is:

$$\mu = F\left(qz, x\right)$$

22

and:

$$f^* (z) = zF (qz, x) - \frac{G (F (qz, x), x)}{q}$$

Its derivative is:

$$f^{*(1)} (z) = F (qz, x) + F' (qz, x) (qz - g (F (qz, x), x)) = F (qz, x)$$

because $g (F (qz, x), x) = qz$. The dual can then be written as a minimization problem:

$$\min_{\lambda} \lambda\mu + \sum_{i=1}^{I} f_i^* (-v_i\lambda)$$

Taking the derivative with respect to λ, we have:

$$\mu = \sum_{i=1}^{I} v_i f_i^{*(1)} (-v_i\lambda) = \sum_{i=1}^{I} v_i F (-q_i v_i \lambda)$$

which immediately gives us:

$$\mu_i = f_i^{*(1)} (-v_i\lambda) = F (-q_i v_i \lambda, x_i)$$

that is we get back to the solution of the primal problem.

3. 2-dimensional raking

3.1. General distance

We want to solve the optimization problem:

$$\min_{M} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{G (\mu_{i,j}, x_{i,j})}{q_{i,j}}$$

under the constraints:

$$Mv_1 = \mu_1 \text{ and } M^T v_2 = \mu_2$$

3.1.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}\left(\mu_{i,j}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\right) = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{G_{i,j}\left(\mu_{i,j}, x_{i,j}\right)}{q_{i,j}} + \boldsymbol{\lambda}_1^T\left(\boldsymbol{M}\boldsymbol{v}_1 - \boldsymbol{\mu}_1\right) + \boldsymbol{\lambda}_2^T\left(\boldsymbol{M}^T\boldsymbol{v}_2 - \boldsymbol{\mu}_2\right)$$

If we write that the partial derivatives with respect to the $\mu_{i,j}$ are equal to 0, we get:

$$\frac{\partial \mathcal{L}\left(\mu_{i,j}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\right)}{\partial \mu_i} = \frac{g_{i,j}\left(\mu_{i,j}, x_{i,j}\right)}{q_{i,j}} + \lambda_{1,i}v_{1,j} + \lambda_{2,j}v_{2,i} = 0$$

where:

$$g_{i,j}\left(\mu_{i,j}, x_{i,j}\right) = \frac{\partial G_{i,j}\left(\mu_{i,j}, x_{i,j}\right)}{\partial \mu_{i,j}}$$

First method: $I + J$ unknowns λ_1 and λ_2

If we denote $F_{i,j}\left(.\right)$ the inverse function of $g_{i,j}\left(., x_{i,j}\right)$ (that is $F_{i,j}\left(u\right) = v$ if $g_{i,j}\left(v, x_{i,j}\right) = u$), we get the raked values:

$$\mu_{i,j} = F_{i,j}\left(-q_{i,j}\left(\lambda_{1,i}v_{1,j} + \lambda_{2,j}v_{2,i}\right)\right)$$

Using the constraint, we get:

$$F\left(-\boldsymbol{Q} \odot \left(\boldsymbol{\lambda}_1\boldsymbol{v}_1^T + \boldsymbol{v}_2\boldsymbol{\lambda}_2^T\right)\right)\boldsymbol{v}_1 = \boldsymbol{\mu}_1$$
$$F\left(-\boldsymbol{Q} \odot \left(\boldsymbol{\lambda}_1\boldsymbol{v}_1^T + \boldsymbol{v}_2\boldsymbol{\lambda}_2^T\right)\right)^T\boldsymbol{v}_2 = \boldsymbol{\mu}_2$$

We use alternative iterations on $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ to solve this last linear system for $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$. We start at $\boldsymbol{\lambda}_1^0 = \boldsymbol{0}$ and $\boldsymbol{\lambda}_1^0 = \boldsymbol{0}$ and we iterate:

$$\begin{pmatrix} \lambda_{1,1}^{k+1} \\ \vdots \\ \lambda_{1,I}^{k+1} \end{pmatrix} = \begin{pmatrix} \lambda_{1,1}^{k} \\ \vdots \\ \lambda_{1,I}^{k} \end{pmatrix} - \left[\nabla_{\boldsymbol{\lambda}_1}\boldsymbol{f_1}\left(\boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k\right)\right]^{-1}\boldsymbol{f_1}\left(\boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k\right)$$

$$\begin{pmatrix} \lambda_{2,1}^{k+1} \\ \vdots \\ \lambda_{2,J}^{k+1} \end{pmatrix} = \begin{pmatrix} \lambda_{2,1}^{k} \\ \vdots \\ \lambda_{2,J}^{k} \end{pmatrix} - \left[\nabla_{\boldsymbol{\lambda}_2}\boldsymbol{f_2}\left(\boldsymbol{\lambda}_1^{k+1}, \boldsymbol{\lambda}_2^k\right)\right]^{-1}\boldsymbol{f_2}\left(\boldsymbol{\lambda}_1^{k+1}, \boldsymbol{\lambda}_2^k\right)$$

where:

$$\boldsymbol{f_1}\left(\boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k\right) = \boldsymbol{\mu}_1 - F\left(-\boldsymbol{Q} \odot \left(\boldsymbol{\lambda}_1^k \boldsymbol{v}_1^T + \boldsymbol{v}_2 \boldsymbol{\lambda}_2^{kT}\right)\right) \boldsymbol{v}_1$$

$$\boldsymbol{f_2}\left(\boldsymbol{\lambda}_1^{k+1}, \boldsymbol{\lambda}_2^k\right) = \boldsymbol{\mu}_2 - F\left(-\boldsymbol{Q} \odot \left(\boldsymbol{\lambda}_1^{k+1} \boldsymbol{v}_1^T + \boldsymbol{v}_2 \boldsymbol{\lambda}_2^{kT}\right)\right)^T \boldsymbol{v}_2$$

We have:

$$\nabla_{\boldsymbol{\lambda}_1} \boldsymbol{f_1}\left(\boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k\right) = \mathrm{diag}\left(\left[\boldsymbol{Q} \odot F'\left(-\boldsymbol{Q} \odot \left(\boldsymbol{\lambda}_1^k \boldsymbol{v}_1^T + \boldsymbol{v}_2 \boldsymbol{\lambda}_2^{kT}\right)\right)\right]\left[\boldsymbol{v}_1^{\odot 2}\right]\right)$$

$$\nabla_{\boldsymbol{\lambda}_2} \boldsymbol{f_2}\left(\boldsymbol{\lambda}_1^{k+1}, \boldsymbol{\lambda}_2^k\right) = \mathrm{diag}\left(\left[\boldsymbol{Q} \odot F'\left(-\boldsymbol{Q} \odot \left(\boldsymbol{\lambda}_1^{k+1} \boldsymbol{v}_1^T + \boldsymbol{v}_2 \boldsymbol{\lambda}_2^{kT}\right)\right)\right]^T\left[\boldsymbol{v}_2^{\odot 2}\right]\right)$$

When $G\left(\mu, x\right) = \mu \log\left(\frac{\mu}{x}\right) - \mu + x$, we have $F\left(u, x\right) = xe^u$ and $F'\left(u, x\right) = xe^u$. If $\boldsymbol{v}_1 = \boldsymbol{1}$, $\boldsymbol{v}_2 = \boldsymbol{1}$ and $q_{i,j} = 1$ for all i,j, the iterations become:

$$\boldsymbol{f_{1,i}}\left(\boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k\right) = \boldsymbol{\mu}_{1,i} - e^{-\lambda_{1,i}^k} \sum_{j=1}^{J} x_{i,j} e^{-\lambda_{2,j}^k}$$

$$\boldsymbol{f_{2,j}}\left(\boldsymbol{\lambda}_1^{k+1}, \boldsymbol{\lambda}_2^k\right) = \boldsymbol{\mu}_{2,j} - e^{-\lambda_{2,j}^k} \sum_{i=1}^{I} x_{i,j} e^{-\lambda_{1,i}^{k+1}}$$

and:

$$\nabla_{\boldsymbol{\lambda}_1} \boldsymbol{f_{1,i,i}}\left(\boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k\right) = e^{-\lambda_{1,i}^k} \sum_{j=1}^{J} x_{i,j} e^{-\lambda_{2,j}^k}$$

$$\nabla_{\boldsymbol{\lambda}_1} \boldsymbol{f_{2,j,j}}\left(\boldsymbol{\lambda}_1^{k+1}, \boldsymbol{\lambda}_2^k\right) = e^{-\lambda_{2,j}^k} \sum_{i=1}^{I} x_{i,j} e^{-\lambda_{1,i}^{k+1}}$$

Second method: $IJ + I + J$ **unknowns** M**,** λ_1 **and** λ_2

We can also write the problem as a system of $IJ + I + J$ linear equations:

$$g_{1,1}\left(\mu_{1,1}, x_{1,1}\right) + q_{1,1}\left(\lambda_{1,1} v_{1,1} + \lambda_{2,1} v_{2,1}\right) = 0$$

$$\cdots$$

$$g_{I,J}\left(\mu_{I,J}, x_{I,J}\right) + q_{I,J}\left(\lambda_{1,I} v_{1,J} + \lambda_{2,J} v_{2,I}\right) = 0$$

$$\boldsymbol{M}\boldsymbol{v}_1 - \boldsymbol{\mu}_1 = 0$$

$$\boldsymbol{M}^T\boldsymbol{v}_2 - \boldsymbol{\mu}_2 = 0$$

Using Newton's method, we have:

$$
\begin{pmatrix} \mu_{1,1}^{k+1} \\ \vdots \\ \mu_{I,J}^{k+1} \\ \boldsymbol{\lambda}_1^{k+1} \\ \boldsymbol{\lambda}_2^{k+1} \end{pmatrix} = \begin{pmatrix} \mu_{1,1}^{k} \\ \vdots \\ \mu_{I,J}^{k} \\ \boldsymbol{\lambda}_1^{k} \\ \boldsymbol{\lambda}_2^{k} \end{pmatrix} - \left[\nabla \boldsymbol{f} \left(\boldsymbol{M}^k, \boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k \right) \right]^{-1} \boldsymbol{f} \left(\boldsymbol{M}^k, \boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k \right)
$$

where:

$$
\boldsymbol{f} \left(\boldsymbol{M}^k, \boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k \right) = \begin{pmatrix} g_{1,1} \left(\mu_{1,1}, x_{1,1} \right) + q_{1,1} \left(\lambda_{1,1} v_{1,1} + \lambda_{2,1} v_{2,1} \right) \\ \vdots \\ g_{I,J} \left(\mu_{I,J}, x_{I,J} \right) + q_{I,J} \left(\lambda_{1,I} v_{1,J} + \lambda_{2,J} v_{2,I} \right) \\ \boldsymbol{M} \boldsymbol{v}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{M}^T \boldsymbol{v}_2 - \boldsymbol{\mu}_2 \end{pmatrix}
$$

We have:

$$
\frac{\partial \left(g \left(\mu_{i,j}, x_{i,j} \right) + q_{i,j} \left(\lambda_{1,i} v_{1,j} + \lambda_{2,j} v_{2,i} \right) \right)}{\partial \mu_{k,l}} = \begin{cases} 0 & \text{if } i \neq k \text{ or } j \neq l \\ g' \left(\mu_{i,j}, x_{i,j} \right) & \text{if } i = k \text{ and } j = l \end{cases}
$$

$$
\frac{\partial \left(g \left(\mu_{i,j}, x_{i,j} \right) + q_{i,j} \left(\lambda_{1,i} v_{1,j} + \lambda_{2,j} v_{2,i} \right) \right)}{\partial \lambda_{1,k}} = \begin{cases} 0 & \text{if } i \neq k \\ q_{i,j} v_{1,j} & \text{if } i = k \end{cases}
$$

$$
\frac{\partial \left(g \left(\mu_{i,j}, x_{i,j} \right) + q_{i,j} \left(\lambda_{1,i} v_{1,j} + \lambda_{2,j} v_{2,i} \right) \right)}{\partial \lambda_{2,l}} = \begin{cases} 0 & \text{if } j \neq l \\ q_{i,j} v_{2,i} & \text{if } j = l \end{cases}
$$

$$
\frac{\partial \left(\left[\boldsymbol{M} \boldsymbol{v}_1 - \boldsymbol{\mu}_1 \right]_i \right)}{\partial \mu_{k,l}} = \begin{cases} 0 & \text{if } i \neq k \\ v_{1,l} & \text{if } i = k \end{cases}
$$

$$
\frac{\partial \left(\left[\boldsymbol{M}^T \boldsymbol{v}_2 - \boldsymbol{\mu}_2 \right]_j \right)}{\partial \mu_{k,l}} = \begin{cases} 0 & \text{if } j \neq l \\ v_{2,k} & \text{if } j = l \end{cases}
$$

$$
\frac{\partial \left(\boldsymbol{M} \boldsymbol{v}_1 - \boldsymbol{\mu}_1 \right)}{\partial \lambda_{1,i}} = \frac{\partial \left(\boldsymbol{M} \boldsymbol{v}_1 - \boldsymbol{\mu}_1 \right)}{\partial \lambda_{2,j}} = \frac{\partial \left(\boldsymbol{M}^T \boldsymbol{v}_2 - \boldsymbol{\mu}_2 \right)}{\partial \lambda_{1,i}} = \frac{\partial \left(\boldsymbol{M}^T \boldsymbol{v}_2 - \boldsymbol{\mu}_2 \right)}{\partial \lambda_{2,j}} = 0
$$

3.1.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$\min_{\boldsymbol{M}} \sup_{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{G_{i,j}\left(\mu_{i,j}, x_{i,j}\right)}{q_{i,j}} + \boldsymbol{\lambda}_1^T \left(\boldsymbol{M}\boldsymbol{v}_1 - \boldsymbol{\mu}_1\right) + \boldsymbol{\lambda}_2^T \left(\boldsymbol{M}^T \boldsymbol{v}_2 - \boldsymbol{\mu}_2\right) \geq$$

$$\sup_{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2} -\boldsymbol{\lambda}_1^T \boldsymbol{\mu}_1 - \boldsymbol{\lambda}_2^T \boldsymbol{\mu}_2 - \sup_{\boldsymbol{M}} \sum_{i=1}^{I} \sum_{j=1}^{J} \mu_{i,j}\left(-\lambda_{1,i} v_{1,j} - \lambda_{2,j} v_{2,i}\right) - \frac{G_{i,j}\left(\mu_{i,j}, x_{i,j}\right)}{q_{i,j}} =$$

$$\sup_{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2} -\boldsymbol{\lambda}_1^T \boldsymbol{\mu}_1 - \boldsymbol{\lambda}_2^T \boldsymbol{\mu}_2 - \sum_{i=1}^{I} \sum_{j=1}^{J} f_{i,j}^*\left(-\lambda_{1,i} v_{1,j} - \lambda_{2,j} v_{2,i}\right)$$

where:

$$f^*\left(z\right) = \sup_{\mu} \mu z - \frac{G\left(\mu, x\right)}{q}$$

Taking the derivative of f^* with respect to μ, we get:

$$\frac{\partial f^*\left(z\right)}{\partial \mu} = 0 = z - \frac{g\left(\mu, x\right)}{q}$$

which gives us:

$$qz = g\left(\mu, x\right)$$

that is:

$$\mu = F\left(qz\right)$$

and:

$$f^*\left(z\right) = zF\left(qz\right) - \frac{G\left(F\left(qz\right), x\right)}{q}$$

Its derivative is:

$$f^{*(1)}\left(z\right) = F\left(qz\right) + F'\left(qz\right)\left(qz - g\left(F\left(qz\right), x\right)\right) = F\left(qz\right)$$

because $g\left(F\left(qz\right), x\right) = qz$. The dual can then be written as a minimization problem:

$$\min_{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2} \boldsymbol{\lambda}_1^T \boldsymbol{\mu}_1 + \boldsymbol{\lambda}_2^T \boldsymbol{\mu}_2 + \sum_{i=1}^{I} \sum_{j=1}^{J} f_{i,j}^* \left(-\lambda_{1,i} v_{1,j} - \lambda_{2,j} v_{2,i} \right)$$

Taking the derivative with respect to $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, we have:

$$\boldsymbol{\mu}_1 = \left[f^{*(1)} \left(-\boldsymbol{\lambda}_1 \boldsymbol{v}_1^T - \boldsymbol{v}_2 \boldsymbol{\lambda}_2^T \right) \right] \boldsymbol{v}_1 = F \left(-\boldsymbol{Q} \odot \left(\boldsymbol{\lambda}_1 \boldsymbol{v}_1^T + \boldsymbol{v}_2 \boldsymbol{\lambda}_2^T \right) \right) \boldsymbol{v}_1$$

$$\boldsymbol{\mu}_2 = \left[f^{*(1)} \left(-\boldsymbol{\lambda}_1 \boldsymbol{v}_1^T - \boldsymbol{v}_2 \boldsymbol{\lambda}_2^T \right) \right]^T \boldsymbol{v}_2 = F \left(-\boldsymbol{Q} \odot \left(\boldsymbol{\lambda}_1 \boldsymbol{v}_1^T + \boldsymbol{v}_2 \boldsymbol{\lambda}_2^T \right) \right)^T \boldsymbol{v}_2$$

which immediately gives us:

$$\mu_{i,j} = F_{i,j} \left(-q_{i,j} \left(\lambda_{1,i} v_{1,j} + \lambda_{2,j} v_{2,i} \right) \right)$$

that is we get back to the solution of the primal problem.

4. k-dimensional raking

In this section, instead of computing the raking weights α_n, we compute the raked values μ_n. That will make it easier to add a condition $0 < \mu_n < 1$ on the raked values if we are interested in prevalence. Our objective is then to find the μ_n that minimize:

$$\left(\frac{1}{\boldsymbol{q}} \right)^T \left(G\left(\boldsymbol{\mu}, \boldsymbol{x} \right) \right)$$

under a linear constraint on $\boldsymbol{\mu}$:

$$\boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{y}$$

where $\boldsymbol{A}$ is a matrix with M rows and N columns.

4.1. General distance

4.1.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}\left(\boldsymbol{\mu}, \boldsymbol{\lambda} \right) = \left(\frac{1}{\boldsymbol{q}} \right)^T \left(G\left(\boldsymbol{\mu}, \boldsymbol{x} \right) \right) + \boldsymbol{\lambda}^T \left(\boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y} \right)$$

If we write that the partial derivatives with respect to the μ_n are equal to 0, we get:

$$\frac{\partial \mathcal{L}\left(\boldsymbol{\mu}, \boldsymbol{\lambda}\right)}{\partial \mu_n} = 0 = \frac{g\left(\mu_n, x_n\right)}{q_n} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}$$

where $\boldsymbol{A}_{.,n}$ is the n-th column of $\boldsymbol{A}$ and:

$$g\left(\mu_n, x_n\right) = \frac{\partial G\left(\mu_n, x_n\right)}{\partial \mu_n}$$

First method: M unknowns λ

If we denote $F\left(., x_n\right)$ the inverse function of $g\left(., x_n\right)$ (that is $F\left(u, x_n\right) = v$ if $g\left(v, x_n\right) = u$), we get the raked values:

$$\mu_n = F\left(-q_n \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}, x_n\right)$$

Using the constraint, we get:

$$\boldsymbol{A}\left[F\left(-\boldsymbol{q} \odot \left(\boldsymbol{A}^T \boldsymbol{\lambda}\right), \boldsymbol{x}\right)\right] = \boldsymbol{y}$$

We use Newton's method to solve this last equation for $\boldsymbol{\lambda}$. We denote:

$$\boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}\left[F\left(-\boldsymbol{q} \odot \left(\boldsymbol{A}^T \boldsymbol{\lambda}\right), \boldsymbol{x}\right)\right] \text{ and } \hat{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}$$

thus we need to solve:

$$\boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right) - \hat{\boldsymbol{y}} + \boldsymbol{y} = \boldsymbol{0}$$

We start at $\boldsymbol{\lambda}^0 = \boldsymbol{0}$ and we iterate:

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \left(\nabla_{\boldsymbol{\lambda}} \boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right)\right)^{-1} \left(\boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right) - \hat{\boldsymbol{y}} + \boldsymbol{y}\right)$$

with:

$$\nabla_{\boldsymbol{\lambda}} \boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right) = \boldsymbol{A}\text{diag}\left[\boldsymbol{q} \odot F'\left(-\boldsymbol{q} \odot \left(\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right)\right] \boldsymbol{A}^T$$

Second method: $N + M$ unknowns μ and λ

We can also write the problem as a system of $N + M$ linear equations:

$$\frac{g\left(\mu_1, x_1\right)}{q_1} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,1} = 0$$

$$\dots$$

$$\frac{g\left(\mu_N, x_N\right)}{q_N} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,N} = 0$$

$$\boldsymbol{A}_{1,.}^T \boldsymbol{\mu} - y_1 = 0$$

$$\dots$$

$$\boldsymbol{A}_{M,.}^T \boldsymbol{\mu} - y_M = 0$$

Using Newton's method, we have:

$$
\begin{pmatrix} \mu_1^{k+1} \\ \vdots \\ \mu_N^{k+1} \\ \lambda_1^{k+1} \\ \vdots \\ \lambda_M^{k+1} \end{pmatrix} = \begin{pmatrix} \mu_1^{k} \\ \vdots \\ \mu_N^{k} \\ \lambda_1^{k} \\ \vdots \\ \lambda_M^{k} \end{pmatrix} - \left[\nabla \boldsymbol{H}\left(\boldsymbol{\mu}^k, \boldsymbol{\lambda}^k\right)\right]^{-1} \boldsymbol{H}\left(\boldsymbol{\mu}^k, \boldsymbol{\lambda}^k\right)
$$

where:

$$
\boldsymbol{H}\left(\boldsymbol{\mu}, \boldsymbol{\lambda}\right) = \begin{pmatrix} \frac{g(\mu_1, x_1)}{q_1} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,1} \\ \vdots \\ \frac{g(\mu_N, x_N)}{q_N} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,N} \\ \boldsymbol{A}_{1,.}^T \boldsymbol{\mu} - y_1 \\ \vdots \\ \boldsymbol{A}_{M,.}^T \boldsymbol{\mu} - y_M \end{pmatrix}
$$

We have:

$$
\frac{\partial \left(\frac{g(\mu_n, x_n)}{q_n} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n} \right)}{\partial \mu_l} = \begin{cases} 0 & \text{if } n \neq l \\ \frac{1}{q_n} \frac{\partial g(\mu_n, x_n)}{\partial \mu_n} & \text{if } n = l \end{cases}
$$

$$
\frac{\partial \left(\frac{g(\mu_n, x_n)}{q_n} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n} \right)}{\partial \lambda_m} = A_{mn}
$$

$$\frac{\partial \left(\boldsymbol{A}_{m,.}^T \boldsymbol{\mu} - y_m \right)}{\partial \mu_n} = A_{mn}$$

$$\frac{\partial \left(\boldsymbol{A}_{m,.}^T \boldsymbol{\mu} - y_m \right)}{\partial \lambda_l} = 0$$

thus the iterations become:

$$\begin{pmatrix} \mu_1^{k+1} \\ \vdots \\ \mu_N^{k+1} \\ \lambda_1^{k+1} \\ \vdots \\ \lambda_M^{k+1} \end{pmatrix} = \begin{pmatrix} \mu_1^k \\ \vdots \\ \mu_N^k \\ \lambda_1^k \\ \vdots \\ \lambda_M^k \end{pmatrix} - \begin{pmatrix} \mathrm{diag}\left(\frac{1}{q_n} \frac{\partial g(\mu_n, x_n)}{\partial \mu_n} \right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{m \times m} \end{pmatrix}^{-1} \begin{pmatrix} \frac{g(\mu_1, x_1)}{q_1} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,1} \\ \vdots \\ \frac{g(\mu_N, x_N)}{q_N} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,N} \\ \boldsymbol{A}_{1,.}^T \boldsymbol{\mu} - y_1 \\ \vdots \\ \boldsymbol{A}_{M,.}^T \boldsymbol{\mu} - y_M \end{pmatrix}$$

4.1.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$\min_{\boldsymbol{\mu}} \sup_{\boldsymbol{\lambda}} \left(\frac{1}{\boldsymbol{q}} \right) (G(\boldsymbol{\mu}, \boldsymbol{x})) + \boldsymbol{\lambda}^T (\boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y}) \geq$$
$$\sup_{\boldsymbol{\lambda}} -\boldsymbol{\lambda}^T \boldsymbol{y} - \sup_{\boldsymbol{\mu}} -\boldsymbol{\lambda}^T \boldsymbol{A}\boldsymbol{\mu} - \left(\frac{1}{\boldsymbol{q}} \right)^T (G(\boldsymbol{\mu}, \boldsymbol{x})) =$$
$$\sup_{\boldsymbol{\lambda}} -\boldsymbol{\lambda}^T \boldsymbol{y} - \boldsymbol{1}^T \left[f^* \left(-\boldsymbol{A}^T \boldsymbol{\lambda} \right) \right]$$

where:

$$f^*(z) = \sup_{\mu} \mu z - \frac{G(\mu, x)}{q}$$

Taking the derivative of f^* with respect to μ, we get:

$$\frac{\partial f^*(z)}{\partial \mu} = 0 = z - \frac{g(\mu, x)}{q}$$

which gives us:

$$qz = g(\mu, x)$$

that is:

$$\mu = F(qz)$$

and:

$$f^*(z) = zF(qz) - \frac{G(F(qz),x)}{q}$$

Its derivative is:

$$f^{*(1)}(z) = F(qz) + F'(qz)(qz - g(F(qz),x)) = F(qz)$$

because $g(F(qz),x) = qz$. The dual can be written as a minimization problem:

$$\min_{\boldsymbol{\lambda}} \boldsymbol{\lambda}^T \boldsymbol{y} + \mathbf{1}^T \left[f^*\left(-\boldsymbol{A}^T \boldsymbol{\lambda}\right) \right]$$

Taking the derivative with respect to $\boldsymbol{\lambda}$, we have:

$$\boldsymbol{y} = \boldsymbol{A} \left[f^{*(1)}\left(-\boldsymbol{A}^T \boldsymbol{\lambda}\right) \right]$$

which immediately gives us:

$$\mu_n = f_n^{*(1)}\left(-\boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}\right) = F_n\left(-q_n \boldsymbol{\lambda}^T \boldsymbol{A}_{.n}\right)$$

that is we get back to the solution of the primal problem.

4.2. Chi-square distance

We take:

$$G(\mu, x) = \frac{1}{2x}(\mu - x)^2 \text{ and } g(\mu, x) = \frac{\partial G(\mu, x)}{\partial \mu} = \frac{\mu}{x} - 1$$

4.2.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \left(\frac{1}{q}\right)^T \left[\left(\frac{1}{2\boldsymbol{x}}\right) \odot (\boldsymbol{\mu} - \boldsymbol{x})^{\odot 2} \right] + \boldsymbol{\lambda}^T \left(\boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y}\right)$$

If we write that the partial derivatives with respect to the μ_n are equal to 0, we get:

$$\frac{\partial \mathcal{L}\left(\boldsymbol{\mu}, \boldsymbol{\lambda}\right)}{\partial \mu_n} = 0 = \frac{1}{q_n}\left(\frac{\mu_n}{x_n} - 1\right) + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}$$

where $\boldsymbol{A}_{.,n}$ is the n-th column of $\boldsymbol{A}$.

First method: M unknown λ

We get the raked values:

$$\mu_n = x_n \left(1 - q_n \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}\right)$$

Using the constraint, we get:

$$\boldsymbol{A}\left[\boldsymbol{x} \odot \left(\boldsymbol{1} - \boldsymbol{q} \odot \left(\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right)\right] = \boldsymbol{y}$$

We denote:

$$\boldsymbol{\Phi} = \boldsymbol{A}\text{diag}\left[\boldsymbol{x} \odot \boldsymbol{q}\right]\boldsymbol{A}^T \text{ and } \hat{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}$$

thus we need to solve:

$$\boldsymbol{\Phi}\boldsymbol{\lambda} = \hat{\boldsymbol{y}} - \boldsymbol{y}$$

If the matrix $\boldsymbol{\Phi}$ is invertible, we get:

$$\boldsymbol{\lambda} = \boldsymbol{\Phi}^{-1}\left(\hat{\boldsymbol{y}} - \boldsymbol{y}\right)$$

If the matrix $\boldsymbol{\Phi}$ is not invertible, we use its Moore-Penrose pseudoinverse $\boldsymbol{\Phi}^+$ and get the solution:

$$\boldsymbol{\lambda} = \boldsymbol{\Phi}^+\left(\hat{\boldsymbol{y}} - \boldsymbol{y}\right)$$

Second method: $N + M$ unknowns μ and λ

We can also write the problem as a system of $N + M$ linear equations:

$$\frac{\mu_1}{q_1 x_1} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,1} = \frac{1}{q_1}$$

$$\dots$$

$$\frac{\mu_N}{q_N x_n} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,N} = \frac{1}{q_N}$$

$$\boldsymbol{A}_{1,.}^T \boldsymbol{\mu} = y_1$$

$$\dots$$

$$\boldsymbol{A}_{M,.}^T \boldsymbol{\mu} = y_M$$

which can be re-written as:

$$
\begin{pmatrix} \text{diag}\left(\frac{1}{q \odot x}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{M \times M} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \frac{1}{q} \\ \boldsymbol{y} \end{pmatrix}
$$

4.2.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$\min_{\boldsymbol{\mu}} \sup_{\boldsymbol{\lambda}} \left(\frac{1}{q}\right)^T \left[\left(\frac{1}{2x}\right) \odot (\boldsymbol{\mu} - \boldsymbol{x})^{\odot 2}\right] + \boldsymbol{\lambda}^T (\boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y}) \geq$$

$$\sup_{\boldsymbol{\lambda}} -\boldsymbol{\lambda}^T \boldsymbol{y} - \sup_{\boldsymbol{\mu}} -\boldsymbol{\lambda}^T \boldsymbol{A}\boldsymbol{\mu} - \left(\frac{1}{q}\right)^T \left[\left(\frac{1}{2x}\right) \odot (\boldsymbol{\mu} - \boldsymbol{x})^{\odot 2}\right] =$$

$$\sup_{\boldsymbol{\lambda}} -\boldsymbol{\lambda}^T \boldsymbol{y} - \boldsymbol{1}^T \left[f^* \left(-\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right]$$

where:

$$f^*(z) = \sup_{\mu} \mu z - \frac{(\mu - x)^2}{2qx}$$

Taking the derivative of f^* with respect to μ, we get:

$$\frac{\partial f^*(z)}{\partial \mu} = 0 = z - \frac{1}{q}\left(\frac{\mu}{x} - 1\right)$$

which gives us:

$$\mu = x\left(1 + qz\right)$$

and:

$$f^* (z) = xz \left(1 + \frac{qz}{2}\right)$$

Its derivative is:

$$f^{*(1)} (z) = x (1 + qz)$$

The dual can be written as a minimization problem:

$$\min_{\boldsymbol{\lambda}} \boldsymbol{\lambda}^T \boldsymbol{y} + \boldsymbol{1}^T \left[f^* \left(-\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right]$$

Taking the derivative with respect to $\boldsymbol{\lambda}$, we have:

$$\boldsymbol{y} = \boldsymbol{A} \left[f^{*(1)} \left(-\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right]$$

which immediately gives us:

$$\mu_n = f_n^{*(1)} \left(-\boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}\right) = x_n \left(1 - q_n \boldsymbol{\lambda}^T \boldsymbol{A}_{.n}\right)$$

that is we get back to the solution of the primal problem.

4.3. $\mathcal{L}_2$ **distance**

We take:

$$G (\mu, x) = \frac{1}{2} (\mu - x)^2 \text{ and } g (\mu, x) = \frac{\partial G (\mu, x)}{\partial \mu} = \mu - x$$

4.3.1. *Primal problem*

We can write the Lagrangian:

$$\mathcal{L} (\boldsymbol{\mu}, \boldsymbol{\lambda}) = \left(\frac{1}{\boldsymbol{q}}\right)^T \left[\frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{x})^{\odot 2}\right] + \boldsymbol{\lambda}^T (\boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y})$$

If we write that the partial derivatives with respect to the μ_n are equal to 0, we get:

$$\frac{\partial \mathcal{L} (\boldsymbol{\mu}, \boldsymbol{\lambda})}{\partial \mu_n} = 0 = \frac{\mu_n - x_n}{q_n} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}$$

35

where $\boldsymbol{A}_{.,n}$ is the n-th column of $\boldsymbol{A}$.

First method: M unknown λ

We get the raked values:

$$\mu_n = x_n - q_n \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}$$

Using the constraint, we get:

$$\boldsymbol{A} \left[\boldsymbol{x} - \boldsymbol{q} \odot \left(\boldsymbol{A}^T \boldsymbol{\lambda} \right) \right] = \boldsymbol{y}$$

We denote:

$$\boldsymbol{\Phi} = \boldsymbol{A} \mathrm{diag} \left(\boldsymbol{q} \right) \boldsymbol{A}^T \text{ and } \hat{\boldsymbol{y}} = \boldsymbol{A} \boldsymbol{x}$$

thus we need to solve:

$$\boldsymbol{\Phi} \boldsymbol{\lambda} = \hat{\boldsymbol{y}} - \boldsymbol{y}$$

If the matrix $\boldsymbol{\Phi}$ is invertible, we get:

$$\boldsymbol{\lambda} = \boldsymbol{\Phi}^{-1} \left(\hat{\boldsymbol{y}} - \boldsymbol{y} \right)$$

If the matrix $\boldsymbol{\Phi}$ is not invertible, we use its Moore-Penrose pseudoinverse $\boldsymbol{\Phi}^+$ and get the solution:

$$\boldsymbol{\lambda} = \boldsymbol{\Phi}^+ \left(\hat{\boldsymbol{y}} - \boldsymbol{y} \right)$$

Second method: $N + M$ unknowns μ and λ

We can also write the problem as a system of $N + M$ linear equations:

$$\frac{\mu_1}{q_1} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,1} = \frac{x_1}{q_1}$$

$$\dots$$

$$\frac{\mu_N}{q_N} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,N} = \frac{x_N}{q_N}$$

$$\boldsymbol{A}_{1,.}^T \boldsymbol{\mu} = y_1$$

$$\dots$$

$$\boldsymbol{A}_{M,.}^T \boldsymbol{\mu} = y_M$$

which can be re-written as:

$$\begin{pmatrix} \mathrm{diag}\left(\frac{1}{\boldsymbol{q}}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{M \times M} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \frac{\boldsymbol{x}}{\boldsymbol{q}} \\ \boldsymbol{y} \end{pmatrix}$$

4.3.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$\min_{\boldsymbol{\mu}} \sup_{\boldsymbol{\lambda}} \left(\frac{1}{\boldsymbol{q}}\right)^T \left[\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{x})^{\odot 2}\right] + \boldsymbol{\lambda}^T (\boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y}) \geq$$

$$\sup_{\boldsymbol{\lambda}} -\boldsymbol{\lambda}^T \boldsymbol{y} - \sup_{\boldsymbol{\mu}} -\boldsymbol{\lambda}^T \boldsymbol{A}\boldsymbol{\mu} - \left(\frac{1}{\boldsymbol{q}}\right)^T \left[\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{x})^{\odot 2}\right] =$$

$$\sup_{\boldsymbol{\lambda}} -\boldsymbol{\lambda}^T \boldsymbol{y} - \boldsymbol{1}^T \left[f^* \left(-\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right]$$

where:

$$f^*(z) = \sup_{\mu} \mu z - \frac{(\mu - x)^2}{2q}$$

Taking the derivative of f^* with respect to μ, we get:

$$\frac{\partial f^*(z)}{\partial \mu} = 0 = z - \frac{\mu - x}{q}$$

which gives us:

$$\mu = x + qz$$

and:

$$f^*(z) = z\left(x + \frac{qz}{2}\right)$$

Its derivative is:

$$f^{*(1)}(z) = x + qz$$

The dual can be written as a minimization problem:

$$\min_{\boldsymbol{\lambda}} \boldsymbol{\lambda}^T \boldsymbol{y} + \mathbf{1}^T \left[f^*\left(-\boldsymbol{A}^T\boldsymbol{\lambda}\right)\right]$$

Taking the derivative with respect to $\boldsymbol{\lambda}$, we have:

$$\boldsymbol{y} = \boldsymbol{A}\left[f^{*(1)}\left(-\boldsymbol{A}^T\boldsymbol{\lambda}\right)\right]$$

which immediately gives us:

$$\mu_n = f_n^{*(1)}\left(-\boldsymbol{\lambda}^T\boldsymbol{A}_{.,n}\right) = x_n - q_n\boldsymbol{\lambda}^T\boldsymbol{A}_{.n}$$

that is we get back to the solution of the primal problem.

4.4. Entropic distance

We take:

$$G(\mu, x) = \mu \log\left(\frac{\mu}{x}\right) - \mu + x \text{ and } g(\mu, x) = \frac{\partial G(\mu, x)}{\partial \mu} = \log\left(\frac{\mu}{x}\right)$$

4.4.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \left(\frac{1}{q}\right)^T\left(\boldsymbol{\mu} \odot \log\left(\frac{\boldsymbol{\mu}}{\boldsymbol{x}}\right) - \boldsymbol{\mu} + \boldsymbol{x}\right) + \boldsymbol{\lambda}^T\left(\boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y}\right)$$

If we write that the partial derivatives with respect to the μ_n are equal to 0, we get:

$$\frac{\partial \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda})}{\partial \mu_n} = 0 = \frac{1}{q_n}\log\left(\frac{\mu_n}{x_n}\right) + \boldsymbol{\lambda}^T\boldsymbol{A}_{.,n}$$

where $\boldsymbol{A}_{.,n}$ is the n-th column of $\boldsymbol{A}$.

First method: M **unknowns** λ

We get the raked values:

$$\mu_n = x_n \exp\left(-q_n \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}\right)$$

Using the constraint, we get:

$$\boldsymbol{A}\left[\boldsymbol{x} \odot \exp\left(-\boldsymbol{q} \odot \left(\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right)\right] = \boldsymbol{y}$$

We use Newton's method to solve this last equation for $\boldsymbol{\lambda}$. We denote:

$$\boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}\left[\boldsymbol{x} \odot \exp\left(-\boldsymbol{q} \odot \left(\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right)\right] \text{ and } \hat{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}$$

thus we need to solve:

$$\boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right) - \hat{\boldsymbol{y}} + \boldsymbol{y} = \boldsymbol{0}$$

We start at $\boldsymbol{\lambda}^0 = \boldsymbol{0}$ and we iterate:

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \left(\nabla_{\boldsymbol{\lambda}} \boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right)\right)^{-1} \left(\boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right) - \hat{\boldsymbol{y}} + \boldsymbol{y}\right)$$

with:

$$\nabla_{\boldsymbol{\lambda}} \boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right) = \boldsymbol{A}\mathrm{diag}\left[\boldsymbol{x} \odot \boldsymbol{q} \odot \exp\left(-\boldsymbol{q} \odot \left(\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right)\right] \boldsymbol{A}^T$$

Second method: $N + M$ **unknowns** μ **and** λ

We can also write the problem as a system of $N + M$ linear equations:

$$\frac{1}{q_1} \log\left(\frac{\mu_1}{x_1}\right) + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,1} = 0$$

$$\dots$$

$$\frac{1}{q_N} \log\left(\frac{\mu_N}{x_N}\right) + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,N} = 0$$

$$\boldsymbol{A}_{1,.}^T \boldsymbol{\mu} - y_1 = 0$$

$$\dots$$

$$\boldsymbol{A}_{M,.}^T \boldsymbol{\mu} - y_M = 0$$

Using Newton's method, we have:

$$\begin{pmatrix} \boldsymbol{\mu}^{k+1} \\ \boldsymbol{\lambda}^{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}^k \\ \boldsymbol{\lambda}^k \end{pmatrix} - \left[\nabla \boldsymbol{H}\left(\boldsymbol{\mu}^k, \boldsymbol{\lambda}^k\right)\right]^{-1} \boldsymbol{H}\left(\boldsymbol{\mu}^k, \boldsymbol{\lambda}^k\right)$$

where:

$$\boldsymbol{H}\left(\boldsymbol{\mu}, \boldsymbol{\lambda}\right) = \begin{pmatrix} \frac{1}{q} \odot \log\left(\frac{\boldsymbol{\mu}}{\boldsymbol{x}}\right) + \boldsymbol{A}^T \boldsymbol{\lambda} \\ \boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y} \end{pmatrix}$$

We have:

$$\frac{\partial\left(\frac{1}{q_n} \log\left(\frac{\mu_n}{x_n}\right) + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}\right)}{\partial \mu_l} = \begin{cases} 0 & \text{if } n \neq l \\ \frac{1}{q_n \mu_n} & \text{if } n = l \end{cases}$$

$$\frac{\partial\left(\frac{1}{q_n} \log\left(\frac{\mu_n}{x_n}\right) + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}\right)}{\partial \lambda_m} = A_{mn}$$

$$\frac{\partial\left(\boldsymbol{A}_{m,.}^T \boldsymbol{\mu} - y_m\right)}{\partial \mu_n} = A_{mn}$$

$$\frac{\partial\left(\boldsymbol{A}_{m,.}^T \boldsymbol{\mu} - y_m\right)}{\partial \lambda_l} = 0$$

thus the iterations become:

$$\begin{pmatrix} \boldsymbol{\mu}^{k+1} \\ \boldsymbol{\lambda}^{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}^k \\ \boldsymbol{\lambda}^k \end{pmatrix} - \begin{pmatrix} \text{diag}\left(\frac{1}{\boldsymbol{q}\odot\boldsymbol{\mu}}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{m\times m} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\boldsymbol{q}} \odot \log\left(\frac{\boldsymbol{\mu}^k}{\boldsymbol{x}}\right) + \boldsymbol{A}^T \boldsymbol{\lambda}^k \\ \boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y} \end{pmatrix}$$

4.4.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$
\min_{\boldsymbol{\mu}} \sup_{\boldsymbol{\lambda}} \left(\frac{1}{\boldsymbol{q}}\right)^T \left(\boldsymbol{\mu} \odot \log\left(\frac{\boldsymbol{\mu}}{\boldsymbol{x}}\right) - \boldsymbol{\mu} + \boldsymbol{x}\right) + \boldsymbol{\lambda}^T (\boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y}) \geq
$$

$$
\sup_{\boldsymbol{\lambda}} -\boldsymbol{\lambda}^T \boldsymbol{y} - \sup_{\boldsymbol{\mu}} -\boldsymbol{\lambda}^T \boldsymbol{A}\boldsymbol{\mu} - \left(\frac{1}{\boldsymbol{q}}\right)^T \left(\boldsymbol{\mu} \odot \log\left(\frac{\boldsymbol{\mu}}{\boldsymbol{x}}\right) - \boldsymbol{\mu} + \boldsymbol{x}\right) =
$$

$$
\sup_{\boldsymbol{\lambda}} -\boldsymbol{\lambda}^T \boldsymbol{y} - \mathbf{1}^T \left[f^* \left(-\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right]
$$

where:

$$
f^*(z) = \sup_{\mu} \mu z - \frac{1}{q} \left(\mu \log\left(\frac{\mu}{x}\right) - \mu + x\right)
$$

Taking the derivative of f^* with respect to μ, we get:

$$
\frac{\partial f^*(z)}{\partial \mu} = 0 = z - \frac{1}{q} \log\left(\frac{\mu}{x}\right)
$$

which gives us:

$$
\mu = x \exp(qz)
$$

and:

$$
f^*(z) = \frac{x}{q} \left(\exp(qz) - 1\right)
$$

Its derivative is:

$$
f^{*(1)}(z) = x \exp(qz)
$$

The dual can be written as a minimization problem:

$$
\min_{\boldsymbol{\lambda}} \boldsymbol{\lambda}^T \boldsymbol{y} + \mathbf{1}^T \left[f^* \left(-\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right]
$$

Taking the derivative with respect to $\boldsymbol{\lambda}$, we have:

$$y = A \left[f^{*(1)} \left(-A^T \lambda \right) \right]$$

which immediately gives us:

$$\mu_n = f_n^{*(1)} \left(-\lambda^T A_{.,n} \right) = x_n \exp \left(-q_n \lambda^T A_{.n} \right)$$

that is we get back to the solution of the primal problem.

4.5. A more general distance

We take:

$$G \left(\mu, x \right) = \frac{1}{\alpha} \left[\frac{x}{\alpha + 1} \left(\frac{\mu}{x} \right)^{\alpha + 1} - \mu + c \right] \text{ and } g \left(\mu, x \right) = \frac{\partial G \left(\mu, x \right)}{\partial \mu} = \frac{1}{\alpha} \left[\left(\frac{\mu}{x} \right)^{\alpha} - 1 \right]$$

4.5.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L} \left(\mu, \lambda \right) = \frac{1}{\alpha} \left(\frac{1}{q} \right)^T \left[\frac{1}{\alpha + 1} x \odot \left(\frac{\mu}{x} \right)^{\alpha + 1} - \mu + c \right] + \lambda^T \left(A\mu - y \right)$$

If we write that the partial derivatives with respect to the μ_n are equal to 0, we get:

$$\frac{\partial \mathcal{L} \left(\mu, \lambda \right)}{\partial \mu_n} = 0 = \frac{1}{\alpha q_n} \left[\left(\frac{\mu_n}{x_n} \right)^{\alpha} - 1 \right] + \lambda^T A_{.,n}$$

where $A_{.,n}$ is the n-th column of A .

First method: M unknowns λ

We get the raked values:

$$\mu_n = x_n \left(1 - \alpha q_n \lambda^T A_{.,n} \right)^{\frac{1}{\alpha}}$$

Using the constraint, we get:

$$A \left[x \odot \left(1 - \alpha q \odot \left(A^T \lambda \right) \right)^{\frac{1}{\alpha}} \right] = y$$

We use Newton's method to solve this last equation for λ. We denote:

$$\mathbf{\Phi}(\boldsymbol{\lambda}) = \boldsymbol{Ax} - \boldsymbol{A}\left[\boldsymbol{x} \odot \left(1 - \alpha\boldsymbol{q} \odot \left(\boldsymbol{A}^T\boldsymbol{\lambda}\right)\right)^{\frac{1}{\alpha}}\right] \text{ and } \hat{\boldsymbol{y}} = \boldsymbol{Ax}$$

thus we need to solve:

$$\mathbf{\Phi}(\boldsymbol{\lambda}) - \hat{\boldsymbol{y}} + \boldsymbol{y} = \boldsymbol{0}$$

We start at $\boldsymbol{\lambda}^0 = \boldsymbol{0}$ and we iterate:

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \left(\nabla_{\boldsymbol{\lambda}}\mathbf{\Phi}(\boldsymbol{\lambda})\right)^{-1}\left(\mathbf{\Phi}(\boldsymbol{\lambda}) - \hat{\boldsymbol{y}} + \boldsymbol{y}\right)$$

with:

$$\nabla_{\boldsymbol{\lambda}}\mathbf{\Phi}(\boldsymbol{\lambda}) = \boldsymbol{A}\text{diag}\left[\boldsymbol{x} \odot \boldsymbol{q} \odot \left(1 - \boldsymbol{q} \odot \left(\boldsymbol{A}^T\boldsymbol{\lambda}\right)\right)^{\frac{1}{\alpha}-1}\right]\boldsymbol{A}^T$$

Second method: $N + M$ unknowns μ and λ

We can also write the problem as a system of $N + M$ linear equations:

$$\frac{1}{\alpha q_1}\left[\left(\frac{\mu_1}{x_1}\right)^{\alpha} - 1\right] + \boldsymbol{\lambda}^T\boldsymbol{A}_{.,1} = 0$$

$$\dots$$

$$\frac{1}{\alpha q_N}\left[\left(\frac{\mu_N}{x_N}\right)^{\alpha} - 1\right] + \boldsymbol{\lambda}^T\boldsymbol{A}_{.,N} = 0$$

$$\boldsymbol{A}_{1,.}^T\boldsymbol{\mu} - y_1 = 0$$

$$\dots$$

$$\boldsymbol{A}_{M,.}^T\boldsymbol{\mu} - y_M = 0$$

Using Newton's method, we have:

$$\begin{pmatrix}\boldsymbol{\mu}^{k+1} \\ \boldsymbol{\lambda}^{k+1}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\mu}^k \\ \boldsymbol{\lambda}^k\end{pmatrix} - \left[\nabla\boldsymbol{H}\left(\boldsymbol{\mu}^k, \boldsymbol{\lambda}^k\right)\right]^{-1}\boldsymbol{H}\left(\boldsymbol{\mu}^k, \boldsymbol{\lambda}^k\right)$$

where:

$$\boldsymbol{H}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \begin{pmatrix}\frac{1}{\alpha q} \odot \left[\left(\frac{\mu}{x}\right)^{\alpha} - 1\right] + \boldsymbol{A}^T\boldsymbol{\lambda} \\ \boldsymbol{A\mu} - \boldsymbol{y}\end{pmatrix}$$

We have:

$$\frac{\partial\left(\frac{1}{\alpha q_n}\left[\left(\frac{\mu_n}{x_n}\right)^{\alpha}-1\right]+\boldsymbol{\lambda}^T\boldsymbol{A}_{.,n}\right)}{\partial\mu_l} = \begin{cases} 0 & \text{if } n \neq l \\ \frac{1}{q_n x_n}\left(\frac{\mu_n}{x_n}\right)^{\alpha-1} & \text{if } n = l \end{cases}$$

$$\frac{\partial\left(\frac{1}{\alpha q_n}\left[\left(\frac{\mu_n}{x_n}\right)^{\alpha}-1\right]+\boldsymbol{\lambda}^T\boldsymbol{A}_{.,n}\right)}{\partial\lambda_m} = A_{mn}$$

$$\frac{\partial\left(\boldsymbol{A}_{m,.}^T\boldsymbol{\mu}-y_m\right)}{\partial\mu_n} = A_{mn}$$

$$\frac{\partial\left(\boldsymbol{A}_{m,.}^T\boldsymbol{\mu}-y_m\right)}{\partial\lambda_l} = 0$$

thus the iterations become:

$$\begin{pmatrix} \boldsymbol{\mu}^{k+1} \\ \boldsymbol{\lambda}^{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}^{k} \\ \boldsymbol{\lambda}^{k} \end{pmatrix} - \begin{pmatrix} \operatorname{diag}\left(\frac{1}{\boldsymbol{q}\odot\boldsymbol{x}}\odot\left(\frac{\boldsymbol{\mu}}{\boldsymbol{x}}\right)^{\alpha-1}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{m\times m} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\alpha\boldsymbol{q}}\odot\left[\left(\frac{\boldsymbol{\mu}^k}{\boldsymbol{x}}\right)^{\alpha}-1\right]+\boldsymbol{A}^T\boldsymbol{\lambda}^k \\ \boldsymbol{A}\boldsymbol{\mu}-\boldsymbol{y} \end{pmatrix}$$

4.5.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$\min_{\boldsymbol{\mu}}\sup_{\boldsymbol{\lambda}} \frac{1}{\alpha}\left(\frac{1}{\boldsymbol{q}}\right)^T\left[\frac{1}{\alpha+1}\boldsymbol{x}\odot\left(\frac{\boldsymbol{\mu}}{\boldsymbol{x}}\right)^{\alpha+1}-\boldsymbol{\mu}+\boldsymbol{c}\right]+\boldsymbol{\lambda}^T\left(\boldsymbol{A}\boldsymbol{\mu}-\boldsymbol{y}\right) \geq$$

$$\sup_{\boldsymbol{\lambda}} -\boldsymbol{\lambda}^T\boldsymbol{y} - \sup_{\boldsymbol{\mu}} -\boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{\mu} - \frac{1}{\alpha}\left(\frac{1}{\boldsymbol{q}}\right)^T\left[\frac{1}{\alpha+1}\boldsymbol{x}\odot\left(\frac{\boldsymbol{\mu}}{\boldsymbol{x}}\right)^{\alpha+1}-\boldsymbol{\mu}+\boldsymbol{c}\right] =$$

$$\sup_{\boldsymbol{\lambda}} -\boldsymbol{\lambda}^T\boldsymbol{y} - \boldsymbol{1}^T\left[f^*\left(-\boldsymbol{A}^T\boldsymbol{\lambda}\right)\right]$$

where:

$$f^*(z) = \sup_{\mu} \mu z - \frac{1}{\alpha q}\left[\frac{x}{\alpha+1}\left(\frac{\mu}{x}\right)^{\alpha+1}-\mu+c\right]$$

Taking the derivative of f^* with respect to μ, we get:

$$\frac{\partial f^*(z)}{\partial\mu} = 0 = z - \frac{1}{\alpha q}\left[\left(\frac{\mu}{x}\right)^{\alpha}-1\right]$$

which gives us:

$$\mu = x \left(1 + \alpha q z\right)^{\frac{1}{\alpha}}$$

and:

$$f^* \left(z\right) = \frac{x}{q \left(\alpha + 1\right)} \left(1 + \alpha q z\right)^{\frac{1}{\alpha}+1}$$

Its derivative is:

$$f^{*(1)} \left(z\right) = x \left(1 + \alpha q z\right)^{\frac{1}{\alpha}}$$

The dual can be written as a minimization problem:

$$\min_{\boldsymbol{\lambda}} \boldsymbol{\lambda}^T \boldsymbol{y} + \mathbf{1}^T \left[f^* \left(-\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right]$$

Taking the derivative with respect to $\boldsymbol{\lambda}$, we have:

$$\boldsymbol{y} = \boldsymbol{A} \left[f^{*(1)} \left(-\boldsymbol{A}^T \boldsymbol{\lambda}\right)\right]$$

which immediately gives us:

$$\mu_n = f_n^{*(1)} \left(-\boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}\right) = x_n \left(1 - q_n \boldsymbol{\lambda}^T \boldsymbol{A}_{.n}\right)^{\frac{1}{\alpha}}$$

that is we get back to the solution of the primal problem.

4.6. Logit distance

We want to make sure that all the raked values μ are bounded by:

$$l_n < \mu_n < h_n$$

where $\boldsymbol{l}$ and $\boldsymbol{h}$ are known vectors of length I. For example, we could use $\boldsymbol{l} = \boldsymbol{0}$ for the lower bound and the total population number for the upper bound. We assume that all the data respect the condition $l_n < x_n < h_n$. The distance function becomes:

$$G \left(\mu, x\right) = \left(\mu - l\right) \log \frac{\mu - l}{x - l} + \left(h - \mu\right) \log \frac{h - \mu}{h - x}$$

Its derivative is:

$$g(\mu, x) = \log \frac{\mu - l}{x - l} - \log \frac{h - \mu}{h - x}$$

4.6.1. Primal problem

We can write the Lagrangian:

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \left(\frac{1}{\boldsymbol{q}}\right)^T \left((\boldsymbol{\mu} - \boldsymbol{l}) \log \frac{\boldsymbol{\mu} - \boldsymbol{l}}{\boldsymbol{x} - \boldsymbol{l}} + (\boldsymbol{h} - \boldsymbol{\mu}) \log \frac{\boldsymbol{h} - \boldsymbol{\mu}}{\boldsymbol{h} - \boldsymbol{x}}\right) + \boldsymbol{\lambda}^T (\boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y})$$

If we write that the partial derivatives with respect to the μ_n are equal to 0, we get:

$$\frac{\partial \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda})}{\partial \mu_n} = 0 = \frac{1}{q_n} \left(\log \frac{\mu_n - l_n}{x_n - l_n} - \log \frac{h_n - \mu_n}{h_n - x_n}\right) + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}$$

where $\boldsymbol{A}_{.,n}$ is the n-th column of $\boldsymbol{A}$.

First method: M unknowns λ

We get the raked values:

$$\mu_n = \frac{l_n (h_n - x_n) + h_n (x_n - l_n) e^{-q_n \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}}}{(h_n - x_n) + (x_n - l_n) e^{-q_n \boldsymbol{\lambda}^T \boldsymbol{A}_{.,n}}}$$

Using the constraint, we get:

$$\boldsymbol{A} \left[\frac{\boldsymbol{l} \odot (\boldsymbol{h} - \boldsymbol{x}) + \boldsymbol{h} \odot (\boldsymbol{x} - \boldsymbol{l}) e^{-\boldsymbol{q} \odot (\boldsymbol{A}^T \boldsymbol{\lambda})}}{(\boldsymbol{h} - \boldsymbol{x}) + (\boldsymbol{x} - \boldsymbol{l}) e^{-\boldsymbol{q} \odot \boldsymbol{A}^T \boldsymbol{\lambda}}}\right] = \boldsymbol{y}$$

We use Newton's method to solve this last equation for $\boldsymbol{\lambda}$. We denote:

$$\boldsymbol{\Phi}(\boldsymbol{\lambda}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A} \left[\frac{\boldsymbol{l} \odot (\boldsymbol{h} - \boldsymbol{x}) + \boldsymbol{h} \odot (\boldsymbol{x} - \boldsymbol{l}) e^{-\boldsymbol{q} \odot (\boldsymbol{A}^T \boldsymbol{\lambda})}}{(\boldsymbol{h} - \boldsymbol{x}) + (\boldsymbol{x} - \boldsymbol{l}) e^{-\boldsymbol{q} \odot \boldsymbol{A}^T \boldsymbol{\lambda}}}\right] \text{ and } \hat{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}$$

thus we need to solve:

$$\boldsymbol{\Phi}(\boldsymbol{\lambda}) - \hat{\boldsymbol{y}} + \boldsymbol{y} = \boldsymbol{0}$$

We start at $\boldsymbol{\lambda}^0 = \boldsymbol{0}$ and we iterate:

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - (\nabla_{\boldsymbol{\lambda}} \boldsymbol{\Phi}(\boldsymbol{\lambda}))^{-1} (\boldsymbol{\Phi}(\boldsymbol{\lambda}) - \hat{\boldsymbol{y}} + \boldsymbol{y})$$

with:

$$\nabla_{\boldsymbol{\lambda}}\boldsymbol{\Phi}\left(\boldsymbol{\lambda}\right) = \boldsymbol{A}\mathrm{diag}\left[\boldsymbol{q}\odot\frac{\left(\boldsymbol{h}-\boldsymbol{l}\right)\left(\boldsymbol{x}-\boldsymbol{l}\right)\left(\boldsymbol{h}-\boldsymbol{x}\right)e^{-\boldsymbol{q}\odot\left(\boldsymbol{A}^T\boldsymbol{\lambda}\right)}}{\left[\left(\boldsymbol{h}-\boldsymbol{x}\right)+\left(\boldsymbol{x}-\boldsymbol{l}\right)e^{-\boldsymbol{q}\odot\left(\boldsymbol{A}^T\boldsymbol{\lambda}\right)}\right]^2}\right]\boldsymbol{A}^T$$

Second method: $N+M$ unknowns μ and λ

We can also write the problem as a system of $N+M$ linear equations:

$$\frac{1}{q_1}\left(\log\frac{\mu_1-l_1}{x_1-l_1}-\log\frac{h_1-\mu_1}{h_1-x_1}\right)+\boldsymbol{\lambda}^T\boldsymbol{A}_{.,1}=0$$

$$\ldots$$

$$\frac{1}{q_N}\left(\log\frac{\mu_N-l_N}{x_N-l_N}-\log\frac{h_N-\mu_N}{h_N-x_N}\right)+\boldsymbol{\lambda}^T\boldsymbol{A}_{.,N}=0$$

$$\boldsymbol{A}^T_{1,.}\boldsymbol{\mu}-y_1=0$$

$$\ldots$$

$$\boldsymbol{A}^T_{M,.}\boldsymbol{\mu}-y_M=0$$

Using Newton's method, we have:

$$\begin{pmatrix}\mu_1^{k+1}\\\vdots\\\mu_N^{k+1}\\\lambda_1^{k+1}\\\vdots\\\lambda_M^{k+1}\end{pmatrix}=\begin{pmatrix}\mu_1^{k}\\\vdots\\\mu_N^{k}\\\lambda_1^{k}\\\vdots\\\lambda_M^{k}\end{pmatrix}-\left[\nabla\boldsymbol{H}\left(\boldsymbol{\mu}^k,\boldsymbol{\lambda}^k\right)\right]^{-1}\boldsymbol{H}\left(\boldsymbol{\mu}^k,\boldsymbol{\lambda}^k\right)$$

where:

$$\boldsymbol{H}\left(\boldsymbol{\mu},\boldsymbol{\lambda}\right)=\begin{pmatrix}\frac{1}{q_1}\left(\log\frac{\mu_1-l_1}{x_1-l_1}-\log\frac{h_1-\mu_1}{h_1-x_1}\right)+\boldsymbol{\lambda}^T\boldsymbol{A}_{.,1}\\\vdots\\\frac{1}{q_N}\left(\log\frac{\mu_N-l_N}{x_N-l_N}-\log\frac{h_N-\mu_N}{h_N-x_N}\right)+\boldsymbol{\lambda}^T\boldsymbol{A}_{.,N}\\\boldsymbol{A}^T_{1,.}\boldsymbol{\mu}-y_1\\\vdots\\\boldsymbol{A}^T_{M,.}\boldsymbol{\mu}-y_M\end{pmatrix}$$

We have:

$$\frac{\partial\left(\frac{1}{q_n}\left(\log\frac{\mu_n-l_n}{x_n-l_n}-\log\frac{h_n-\mu_n}{h_n-x_n}\right)+\boldsymbol{\lambda}^T\boldsymbol{A}_{.,n}\right)}{\partial\mu_l}=\begin{cases}0 & \text{if } n\neq l\\ \frac{1}{q_n}\left(\frac{1}{\mu_n-l_n}+\frac{1}{h_n-\mu_n}\right) & \text{if } n=l\end{cases}$$

$$\frac{\partial\left(\frac{1}{q_n}\left(\log\frac{\mu_n-l_n}{x_n-l_n}-\log\frac{h_n-\mu_n}{h_n-x_n}\right)+\boldsymbol{\lambda}^T\boldsymbol{A}_{.,n}\right)}{\partial\lambda_m}=A_{mn}$$

$$\frac{\partial\left(\boldsymbol{A}_{m,.}^T\boldsymbol{\mu}-y_m\right)}{\partial\mu_n}=A_{mn}$$

$$\frac{\partial\left(\boldsymbol{A}_{m,.}^T\boldsymbol{\mu}-y_m\right)}{\partial\lambda_l}=0$$

thus the iterations become:

$$\begin{pmatrix}\mu_1^{k+1}\\ \vdots\\ \mu_N^{k+1}\\ \lambda_1^{k+1}\\ \vdots\\ \lambda_M^{k+1}\end{pmatrix}=\begin{pmatrix}\mu_1^{k}\\ \vdots\\ \mu_N^{k}\\ \lambda_1^{k}\\ \vdots\\ \lambda_M^{k}\end{pmatrix}-\begin{pmatrix}\mathrm{diag}\left(\frac{1}{q_n}\left(\frac{1}{\mu_n-l_n}+\frac{1}{h_n-\mu_n}\right)\right) & \boldsymbol{A}^T\\ \boldsymbol{A} & \boldsymbol{0}_{m\times m}\end{pmatrix}^{-1}\begin{pmatrix}\frac{1}{q_1}\left(\log\frac{\mu_1-l_1}{x_1-l_1}-\log\frac{h_1-\mu_1}{h_1-x_1}\right)+\boldsymbol{\lambda}^T\boldsymbol{A}_{.,1}\\ \vdots\\ \frac{1}{q_N}\left(\log\frac{\mu_N-l_N}{x_N-l_N}-\log\frac{h_N-\mu_N}{h_N-x_N}\right)+\boldsymbol{\lambda}^T\boldsymbol{A}_{.,N}\\ \boldsymbol{A}_{1,.}^T\boldsymbol{\mu}-y_1\\ \vdots\\ \boldsymbol{A}_{M,.}^T\boldsymbol{\mu}-y_M\end{pmatrix}$$

4.6.2. Dual problem

Another way of solving this problem is to write the dual problem:

$$\min_{\boldsymbol{\mu}}\sup_{\boldsymbol{\lambda}}\left(\frac{1}{\boldsymbol{q}}\right)^T\left((\boldsymbol{\mu}-\boldsymbol{l})\odot\log\frac{\boldsymbol{\mu}-\boldsymbol{l}}{\boldsymbol{x}-\boldsymbol{l}}+(\boldsymbol{h}-\boldsymbol{\mu})\odot\log\frac{\boldsymbol{h}-\boldsymbol{\mu}}{\boldsymbol{h}-\boldsymbol{x}}\right)+\boldsymbol{\lambda}^T\left(\boldsymbol{A}\boldsymbol{\mu}-\boldsymbol{y}\right)\geq$$
$$\sup_{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^T\boldsymbol{y}-\sup_{\boldsymbol{\mu}}-\boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{\mu}-\left(\frac{1}{\boldsymbol{q}}\right)^T\left((\boldsymbol{\mu}-\boldsymbol{l})\odot\log\frac{\boldsymbol{\mu}-\boldsymbol{l}}{\boldsymbol{x}-\boldsymbol{l}}+(\boldsymbol{h}-\boldsymbol{\mu})\odot\log\frac{\boldsymbol{h}-\boldsymbol{\mu}}{\boldsymbol{h}-\boldsymbol{x}}\right)=$$
$$\sup_{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^T\boldsymbol{y}-\boldsymbol{1}^T\left[f^*\left(-\boldsymbol{A}^T\boldsymbol{\lambda}\right)\right]$$

where:

$$f^*\left(z\right)=\sup_{\mu}\mu z-\frac{1}{q}\left((\mu-l)\log\frac{\mu-l}{x-l}+(h-\mu)\log\frac{h-\mu}{h-x}\right)$$

Taking the derivative of f^* with respect to μ, we get:

$$\frac{\partial f^*(z)}{\partial \mu} = 0 = z - \frac{1}{q}\left[\log\frac{\mu - l}{x - l} - \log\frac{h - \mu}{h - x}\right]$$

which gives us:

$$\mu = \frac{l(h - x) + h(x - l)e^{qz}}{(h - x) + (x - l)e^{qz}}$$

and:

$$f^*(z) = zF(qz) - \frac{G(F(qz), x)}{q}$$

where:

$$F(u) = \frac{l(h - x) + h(x - l)e^u}{(h - x) + (x - l)e^u}$$

Its derivative is:

$$f^{*(1)}(z) = F(qz) = \frac{l(h - x) + h(x - l)e^{qz}}{(h - x) + (x - l)e^{qz}}$$

The dual can be written as a minimization problem:

$$\min_{\boldsymbol{\lambda}} \boldsymbol{\lambda}^T \boldsymbol{y} + \mathbf{1}^T \left[f^*\left(-\boldsymbol{A}^T\boldsymbol{\lambda}\right)\right]$$

Taking the derivative with respect to $\boldsymbol{\lambda}$, we have:

$$\boldsymbol{y} = \boldsymbol{A}\left[f^{*(1)}\left(-\boldsymbol{A}^T\boldsymbol{\lambda}\right)\right]$$

which immediately gives us:

$$\mu_n = f_n^{*(1)}\left(-\boldsymbol{\lambda}^T\boldsymbol{A}_{.,n}\right) = \frac{l_n(h_n - x_n) + h_n(x_n - l_n)e^{-q_n\boldsymbol{\lambda}^T\boldsymbol{A}_{.n}}}{(h_n - x_n) + (x_n - l_n)e^{-q_n\boldsymbol{\lambda}^T\boldsymbol{A}_{.n}}}$$

that is we get back to the solution of the primal problem.

5. Introducing uncertainty

5.1. Using the $\mathcal{L}_2$ distance

5.1.1. 1-dimensional raking

We suppose that the $x_i, i = 1, \cdots, I$ follow a distribution $\mathcal{L}(\bar{\mu}_i, \sigma_i)$ and that μ follow a distribution $\mathcal{L}(\bar{\mu}, \sigma_\mu)$. Our aim is to minimize:

$$\min_{\boldsymbol{\mu}} \sum_{i=1}^{I} \frac{1}{2q_i} (\mu_i - x_i)^2$$

under the constraint:

$$\boldsymbol{v}^T \boldsymbol{\mu} = \mu$$

We get:

$$\lambda = \frac{\sum_{i=1}^{I} v_i x_i - \mu}{\sum_{i=1}^{I} q_i v_i^2}$$

and:

$$\mu_i = x_i - q_i v_i \lambda$$

If the x_i and μ are mutually independent, we can then write the variance:

$$\mathbb{V}(\lambda) = \frac{1}{\left(\sum_{i=1}^{I} q_i v_i^2\right)^2} \left(\sum_{i=1}^{I} v_i^2 \mathbb{V}(x_i) + \mathbb{V}(\mu)\right) = \frac{\sum_{i=1}^{I} v_i^2 \sigma_i^2 + \sigma_\mu^2}{\left(\sum_{i=1}^{I} q_i v_i^2\right)^2}$$

and:

$$\mathbb{V}(\mu_i) = \mathbb{V}(x_i) + q_i^2 v_i^2 \mathbb{V}(\lambda) = \sigma_i^2 + q_i^2 v_i^2 \frac{\sum_{j=1}^{I} v_j^2 \sigma_j^2 + \sigma_\mu^2}{\left(\sum_{j=1}^{I} q_j v_j^2\right)^2}$$

If the x_i and μ are not independent, we denote $\boldsymbol{\Sigma}$ the covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{\sigma}$ the vector containing the covariances $Cov(x_i, \mu)$. We have:

$$\mu_i = x_i - q_i v_i \frac{\sum_j v_j x_j - \mu}{\sum_j q_j v_j^2} = x_i + \frac{q_i v_i}{\sum_j q_j v_j^2} \mu - \frac{q_i v_i}{\sum_j q_j v_j^2} \sum_j v_j x_j$$

from which we can compute the variance:

$$\mathbb{V}\left(\mu_i\right) = \mathbb{V}\left(x_i\right) + \left(\frac{q_i v_i}{\sum_j q_j v_j^2}\right)^2 \mathbb{V}\left(\mu\right) + \left(\frac{q_i v_i}{\sum_j q_j v_j^2}\right)^2 \mathbb{V}\left(\sum_j v_j x_j\right)$$

$$+ 2Cov\left(x_i, \frac{q_i v_i}{\sum_j q_j v_j^2}\mu\right) + 2Cov\left(x_i, \frac{q_i v_i}{\sum_j q_j v_j^2}\sum_j v_j x_j\right) + 2Cov\left(\frac{q_i v_i}{\sum_j q_j v_j^2}\mu, \frac{q_i v_i}{\sum_j q_j v_j^2}\sum_j v_j x_j\right)$$

$$= \Sigma_{i,i} + \left(\frac{q_i v_i}{\sum_j q_j v_j^2}\right)^2 \sigma_\mu^2 + \left(\frac{q_i v_i}{\sum_j q_j v_j^2}\right)^2 \boldsymbol{v}^T \boldsymbol{\Sigma v}$$

$$+ 2\left(\frac{q_i v_i}{\sum_j q_j v_j^2}\right)\boldsymbol{\sigma}_i + 2\left(\frac{q_i v_i}{\sum_j q_j v_j^2}\right)\left(\boldsymbol{\Sigma v}\right)_i + 2\left(\frac{q_i v_i}{\sum_j q_j v_j^2}\right)^2 \boldsymbol{v}^T \boldsymbol{\sigma}$$

5.1.2. k-dimensional raking

We denote $\boldsymbol{\Sigma_x}$ the $N * N$ covariance matrix of the data vector $\boldsymbol{x}$, $\boldsymbol{\Sigma_y}$, the $M * M$ covariance matrix of the margins vector $\boldsymbol{y}$ and $\boldsymbol{\Sigma_{xy}}$ the $N * M$ covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{y}$. Our objective is to find $\boldsymbol{\Sigma_\mu}$, the $N * N$ covariance matrix of the raked values vector $\boldsymbol{\mu}$. Our aim is to minimize:

$$\min_{\boldsymbol{\mu}} \left(\frac{1}{\boldsymbol{q}}\right)^T \left[\frac{1}{2}\left(\boldsymbol{\mu} - \boldsymbol{x}\right)^{\odot 2}\right]$$

under the constraint:

$$\boldsymbol{A\mu} = \boldsymbol{y}$$

We get:

$$\boldsymbol{\lambda} = \boldsymbol{\Phi}^{-1}\left(\hat{\boldsymbol{y}} - \boldsymbol{y}\right) \text{ with } \boldsymbol{\Phi} = \boldsymbol{A}\mathrm{diag}\left(\boldsymbol{q}\right)\boldsymbol{A}^T \text{ and } \hat{\boldsymbol{y}} = \boldsymbol{Ax}$$

and:

$$\boldsymbol{\mu} = \boldsymbol{x} - \boldsymbol{q} \odot \left(\boldsymbol{A}^T \boldsymbol{\lambda}\right) = \boldsymbol{x} - \left[\mathrm{diag}\left(\boldsymbol{q}\right)\boldsymbol{A}^T\right]\boldsymbol{\lambda}$$

Denoting:

$$\boldsymbol{\Phi}_1 = \boldsymbol{I} - \left[\mathrm{diag}\left(\boldsymbol{q}\right)\boldsymbol{A}\right]^T \left[\boldsymbol{A}\mathrm{diag}\left(\boldsymbol{q}\right)\boldsymbol{A}^T\right]^{-1}\boldsymbol{A}$$

$$\boldsymbol{\Phi}_2 = \left[\mathrm{diag}\left(\boldsymbol{q}\right)\boldsymbol{A}\right]^T \left[\boldsymbol{A}\mathrm{diag}\left(\boldsymbol{q}\right)\boldsymbol{A}^T\right]^{-1}$$

We have:

$$\boldsymbol{\mu} = \boldsymbol{\Phi}_1 \boldsymbol{x} + \boldsymbol{\Phi}_2 \boldsymbol{y}$$

which gives us:

$$\boldsymbol{\Sigma_\mu} = \boldsymbol{\Phi}_1 \boldsymbol{\Sigma_x} \boldsymbol{\Phi}_1^T + \boldsymbol{\Phi}_2 \boldsymbol{\Sigma_y} \boldsymbol{\Phi}_2^T + 2\boldsymbol{\Phi}_1 \boldsymbol{\Sigma_{xy}} \boldsymbol{\Phi}_2^T$$

5.2. Using automatic differentiation

5.2.1. Delta method

Let us suppose that:

$$\sqrt{M+N} \left(\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix} - \begin{pmatrix} \boldsymbol{m_y} \\ \boldsymbol{m_x} \end{pmatrix} \right) \xrightarrow{d} \mathcal{N} \left(\boldsymbol{0}, \begin{pmatrix} \boldsymbol{\Sigma_y} & \boldsymbol{\Sigma_{yx}} \\ \boldsymbol{\Sigma_{xy}} & \boldsymbol{\Sigma_x} \end{pmatrix} \right)$$

then if we suppose that:

$$\mu_n = h_n \left(\boldsymbol{y}, \boldsymbol{x} \right)$$

and we denote the gradient:

$$\nabla_{\boldsymbol{y},\boldsymbol{x}} h_n \left(\boldsymbol{y}, \boldsymbol{x} \right) = \begin{pmatrix} \frac{\partial h_n(\boldsymbol{y},\boldsymbol{x})}{\partial y_1} \\ \vdots \\ \frac{\partial h_n(\boldsymbol{y},\boldsymbol{x})}{\partial y_M} \\ \frac{\partial h_n(\boldsymbol{y},\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial h_n(\boldsymbol{y},\boldsymbol{x})}{\partial x_N} \end{pmatrix}$$

we have:

$$\sqrt{M+N} \left(h_n \left(\boldsymbol{y}, \boldsymbol{x} \right) - h_n \left(\boldsymbol{m_y}, \boldsymbol{m_x} \right) \right) \xrightarrow{d} \mathcal{N} \left(\boldsymbol{0}, \nabla_{\boldsymbol{y},\boldsymbol{x}} h_n \left(\boldsymbol{y}, \boldsymbol{x} \right)^T \begin{pmatrix} \boldsymbol{\Sigma_y} & \boldsymbol{\Sigma_{yx}} \\ \boldsymbol{\Sigma_{xy}} & \boldsymbol{\Sigma_x} \end{pmatrix} \nabla_{\boldsymbol{y},\boldsymbol{x}} h_n \left(\boldsymbol{y}, \boldsymbol{x} \right) \right)$$

and:

$$\mathbb{V}\left(\mu_n\right) = \nabla_{\boldsymbol{y},\boldsymbol{x}} h_n\left(\boldsymbol{y},\boldsymbol{x}\right)^T \begin{pmatrix} \Sigma_{\boldsymbol{y}} & \Sigma_{\boldsymbol{yx}} \\ \Sigma_{\boldsymbol{xy}} & \Sigma_{\boldsymbol{x}} \end{pmatrix} \nabla_{\boldsymbol{y},\boldsymbol{x}} h_n\left(\boldsymbol{y},\boldsymbol{x}\right)$$

5.2.2. Computing the gradient

From the solution of the optimization problem, we have:

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \left(\nabla_{\boldsymbol{\lambda}} \boldsymbol{\Phi}\left(\boldsymbol{\lambda}^k\right)\right)^{-1}\left(\boldsymbol{\Phi}\left(\boldsymbol{\lambda}^k\right) - \hat{\boldsymbol{y}} + \boldsymbol{y}\right)$$

with:

$$\boldsymbol{\Phi}\left(\boldsymbol{\lambda}^k\right) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}F\left(-\boldsymbol{q} \odot \left(\boldsymbol{A}^T\boldsymbol{\lambda}^k\right)\right) \text{ and } \hat{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}$$

and:

$$\boldsymbol{\mu} = F\left(-\boldsymbol{q} \odot \left(\boldsymbol{A}^T\boldsymbol{\lambda}\right)\right)$$

We can re-write these two equations as:

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{f_1}\left(\boldsymbol{\lambda}^k, \boldsymbol{x}, \boldsymbol{y}\right)$$

and:

$$\boldsymbol{\mu} = \boldsymbol{f_2}\left(\boldsymbol{\lambda}, \boldsymbol{x}\right)$$

Let us suppose that we have 3 iterations for the Newton's method, we can write $\boldsymbol{\mu}$ as a function of $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\boldsymbol{\mu} = \boldsymbol{f_2}\left(\boldsymbol{f_1}\left(\boldsymbol{f_1}\left(\boldsymbol{f_1}\left(\boldsymbol{x},\boldsymbol{y}\right),\boldsymbol{x},\boldsymbol{y}\right),\boldsymbol{x},\boldsymbol{y}\right),\boldsymbol{x}\right) = \boldsymbol{h}\left(\boldsymbol{y},\boldsymbol{x}\right)$$

We denote $\mu_n = h_n\left(\boldsymbol{y},\boldsymbol{x}\right)$. Using automatic differentiation, we can compute numerically:

$$\nabla_{\boldsymbol{x}}\boldsymbol{\mu} = \nabla_{\boldsymbol{x}}\boldsymbol{h}\left(\boldsymbol{y},\boldsymbol{x}\right) \text{ and } \nabla_{\boldsymbol{y}}\boldsymbol{\mu} = \nabla_{\boldsymbol{y}}\boldsymbol{h}\left(\boldsymbol{y},\boldsymbol{x}\right)$$

In the case of the chi-square distance and 1-dimensional raking, we can easily compute an explicit formula for the gradient. We have:

$$\mu_i = x_i \left(1 - q_i v_i \frac{\sum_j v_j x_j - \mu}{\sum_j q_j v_j^2 x_j} \right)$$

thus we get:

$$\frac{\partial \mu_i}{\partial \mu} = \frac{q_i v_i x_i}{\sum_j q_j v_j^2 x_j}$$

$$\frac{\partial \mu_i}{\partial x_k} = -q_i v_i x_i \frac{v_k \sum_j q_j v_j^2 x_j - q_k v_k^2 \left(\sum_j v_j x_j - \mu \right)}{\left(\sum_j q_j v_j^2 x_j \right)^2} \text{ if } k \neq i$$

$$\frac{\partial \mu_i}{\partial x_i} = 1 - \frac{q_i v_i}{\left(\sum_j q_j v_j^2 x_j \right)^2} \left[\left(\sum_j v_j x_j - \mu + x_i v_i \right) \left(\sum_j q_j v_j^2 x_j \right) - q_j v_j^2 x_i \left(\sum_j v_j x_j - \mu \right) \right]$$

5.3. Implicit function theorem

Assume that S is an open subset of $\mathbb{R}^{2n+2m}$ and that $\boldsymbol{H} : S \to \mathbb{R}^{n+m}$ is a function of class C^1. Assume also that $(\boldsymbol{x}_0, \boldsymbol{y}_0; \boldsymbol{\mu}_0, \boldsymbol{\lambda}_0)$ is a point in S such that:

$$\boldsymbol{H}(\boldsymbol{x}_0, \boldsymbol{y}_0; \boldsymbol{\mu}_0, \boldsymbol{\lambda}_0) = 0 \text{ and } \det D_{\boldsymbol{\mu}, \boldsymbol{\lambda}} \boldsymbol{H}(\boldsymbol{x}_0, \boldsymbol{y}_0; \boldsymbol{\mu}_0, \boldsymbol{\lambda}_0) \neq 0$$

i. Then there exist $r_0, r_1 > 0$ such that for every $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{m+n}$ such that $\|(\boldsymbol{x}, \boldsymbol{y}) - (\boldsymbol{x}_0, \boldsymbol{y}_0)\| < r_0$, there exists a unique $(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^{n+m}$ such that:

$$\|(\boldsymbol{\mu}, \boldsymbol{\lambda}) - (\boldsymbol{\mu}_0, \boldsymbol{\lambda}_0)\| < r_1 \text{ and } \boldsymbol{H}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0 \tag{1}$$

In other words, equation **(1)** implicitly defines a function $(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y})$ for $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{n+m}$ near $(\boldsymbol{x}_0, \boldsymbol{y}_0)$, with $(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y})$ close to $(\boldsymbol{\mu}_0, \boldsymbol{\lambda}_0)$. Note in particular that $\boldsymbol{h}(\boldsymbol{x}_0, \boldsymbol{y}_0) = (\boldsymbol{\mu}_0, \boldsymbol{\lambda}_0)$.

ii. Moreover, the function $\boldsymbol{h} : B(r_0; \boldsymbol{x}, \boldsymbol{y}) \to B(r_1; \boldsymbol{\mu}, \boldsymbol{\lambda}) \subset \mathbb{R}^{m+n}$ from part **(i)** above is of class C^1, and its derivatives may be determined by differentiating the identity:

$$\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y})) = 0$$

(a consequence of the definition of $\boldsymbol{h}$) and solving to find the partial derivatives of $\boldsymbol{h}$.

When differentiating $\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y})) = 0$ at the solution $(\boldsymbol{\mu}_0, \boldsymbol{\lambda}_0)$, we get:

$$[D_{\boldsymbol{\mu},\boldsymbol{\lambda}}\boldsymbol{H}\left(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\mu}_0,\boldsymbol{\lambda}_0\right)]\left[D_{\boldsymbol{x},\boldsymbol{y}}\boldsymbol{h}\left(\boldsymbol{x},\boldsymbol{y}\right)\right]=-\left[D_{\boldsymbol{x},\boldsymbol{y}}\boldsymbol{H}\left(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\mu}_0,\boldsymbol{\lambda}_0\right)\right]$$

The unknown matrix that we need to compute to get the gradient is:

$$D_{\boldsymbol{x},\boldsymbol{y}}\boldsymbol{h}=\begin{pmatrix}\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{x}}&\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{y}}\\\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{x}}&\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{y}}\end{pmatrix}$$

and we can compute easily the two other matrices:

$$D_{\boldsymbol{\mu},\boldsymbol{\lambda}}\boldsymbol{H}=\begin{pmatrix}\frac{\partial\boldsymbol{H}_1}{\partial\boldsymbol{\mu}}&\frac{\partial\boldsymbol{H}_1}{\partial\boldsymbol{\lambda}}\\\frac{\partial\boldsymbol{H}_2}{\partial\boldsymbol{\mu}}&\frac{\partial\boldsymbol{H}_2}{\partial\boldsymbol{\lambda}}\end{pmatrix}=\begin{pmatrix}\mathrm{diag}\left(\frac{1}{q_n}\frac{\partial g(\mu_n,x_n)}{\partial\mu_n}\right)&\boldsymbol{A}^T\\\boldsymbol{A}&\boldsymbol{0}_{m\times m}\end{pmatrix}$$

and:

$$D_{\boldsymbol{x},\boldsymbol{y}}\boldsymbol{H}=\begin{pmatrix}\frac{\partial\boldsymbol{H}_1}{\partial\boldsymbol{x}}&\frac{\partial\boldsymbol{H}_1}{\partial\boldsymbol{y}}\\\frac{\partial\boldsymbol{H}_2}{\partial\boldsymbol{x}}&\frac{\partial\boldsymbol{H}_2}{\partial\boldsymbol{y}}\end{pmatrix}=\begin{pmatrix}\mathrm{diag}\left(\frac{1}{q_n}\frac{\partial g(\mu_n,x_n)}{\partial x_n}\right)&\boldsymbol{0}_{n\times m}\\\boldsymbol{0}_{m\times n}&-\boldsymbol{I}_{m\times m}\end{pmatrix}$$

5.4. Using Schulz's iterative method

From Ben-Israel, 1965. Let $\boldsymbol{\Phi}$ be an arbitrary (nonzero) complex $m\times n$ matrix of rank r and let:

$$\lambda_1\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^*\right)\geq\lambda_2\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^*\right)\geq\cdots\geq\lambda_r\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^*\right)$$

denote the eigenvalues of $\boldsymbol{\Phi}\boldsymbol{\Phi}^*$. If the real scalar α satisfies:

$$0<\alpha<\frac{2}{\lambda_1\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^*\right)}$$

then the sequences defined by:

$$\boldsymbol{\Phi}_0=\alpha\boldsymbol{\Phi}^*$$
$$\boldsymbol{\Phi}_{k+1}=\boldsymbol{\Phi}_k\left(2\boldsymbol{I}-\boldsymbol{\Phi}\boldsymbol{\Phi}_k\right)$$

converges to $\boldsymbol{\Phi}^+$ as $k\to\infty$.

5.4.1. Application to raking problem with the Chi2 distance

We want to solve the system:

$$\frac{\mu_1}{q_1 x_1} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,1} = \frac{1}{q_1}$$

$$\dots$$

$$\frac{\mu_N}{q_N x_N} + \boldsymbol{\lambda}^T \boldsymbol{A}_{.,N} = \frac{1}{q_N}$$

$$\boldsymbol{A}_{1,.}^T \boldsymbol{\mu} - y_1 = 0$$

$$\dots$$

$$\boldsymbol{A}_{M,.}^T \boldsymbol{\mu} - y_M = 0$$

which can be re-written as:

$$\boldsymbol{\Phi} \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \frac{1}{q} \\ y \end{pmatrix}$$

with:

$$\boldsymbol{\Phi} = \begin{pmatrix} \text{diag}\left(\frac{1}{q \odot x}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{M \times M} \end{pmatrix}$$

We can approximate the solution by:

$$\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\lambda} \end{pmatrix} = \boldsymbol{\Phi}^+ \begin{pmatrix} \frac{1}{q} \\ y \end{pmatrix} \text{ with } \boldsymbol{\Phi}^+ = \alpha \boldsymbol{\Phi}^* \left(2\boldsymbol{I}_{NM \times NM} - \alpha \boldsymbol{\Phi} \boldsymbol{\Phi}^*\right)$$

We get:

$$\phi^+ = 2\alpha \begin{pmatrix} \text{diag}\left(\frac{1}{q \odot x}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{M \times M} \end{pmatrix} - \alpha^2 \begin{pmatrix} \text{diag}\left(\frac{1}{q \odot x}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{M \times M} \end{pmatrix} \begin{pmatrix} \text{diag}\left(\frac{1}{q \odot x}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{M \times M} \end{pmatrix} \begin{pmatrix} \text{diag}\left(\frac{1}{q \odot x}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{M \times M} \end{pmatrix}$$

$$= 2\alpha \begin{pmatrix} \text{diag}\left(\frac{1}{q \odot x}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{M \times M} \end{pmatrix}$$

$$- \alpha^2 \begin{pmatrix} \text{diag}\left(\frac{1}{q^{\odot 3} \odot x^{\odot 3}}\right) + \boldsymbol{A}^T \boldsymbol{A} \text{diag}\left(\frac{1}{q \odot x}\right) + \text{diag}\left(\frac{1}{q \odot x}\right) \boldsymbol{A}^T \boldsymbol{A} & \text{diag}\left(\frac{1}{q^{\odot 2} \odot x^{\odot 2}}\right) \boldsymbol{A}^T + \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{A}^T \\ \boldsymbol{A} \text{diag}\left(\frac{1}{q^{\odot 2} \odot x^{\odot 2}}\right) & \boldsymbol{0}_{M \times M} \end{pmatrix}$$

which gives us:

$$\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\lambda} \end{pmatrix} = 2\alpha \begin{pmatrix} \left(\frac{1}{\boldsymbol{q}^{\odot 2}\odot\boldsymbol{x}}\right) + \boldsymbol{A}^T\boldsymbol{y} \\ \boldsymbol{A}\left(\frac{1}{\boldsymbol{q}}\right) \end{pmatrix}$$

$$- \alpha^2 \begin{pmatrix} \left(\frac{1}{\boldsymbol{q}^{\odot 4}\odot\boldsymbol{x}^{\odot 3}}\right) + \boldsymbol{A}^T\boldsymbol{A}\left(\frac{1}{\boldsymbol{q}^{\odot 2}\boldsymbol{x}}\right) + \mathrm{diag}\left(\frac{1}{\boldsymbol{q}\odot\boldsymbol{x}}\right)\boldsymbol{A}^T\boldsymbol{A}\left(\frac{1}{\boldsymbol{q}}\right) + \mathrm{diag}\left(\frac{1}{\boldsymbol{q}^{\odot 2}\odot\boldsymbol{x}^{\odot 2}}\right)\boldsymbol{A}^T\boldsymbol{y} + \boldsymbol{A}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{y} \\ \boldsymbol{A}\left(\frac{1}{\boldsymbol{q}^{\odot 3}\odot\boldsymbol{x}^{\odot 2}}\right) \end{pmatrix}$$

from which we finally get:

$$\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{y}} = 2\alpha\boldsymbol{A}^T - \alpha^2\left(\mathrm{diag}\left(\frac{1}{\boldsymbol{q}^{\odot 2}\odot\boldsymbol{x}^{\odot 2}}\right)\boldsymbol{A}^T + \boldsymbol{A}^T\boldsymbol{A}\boldsymbol{A}^T\right)$$

and:

$$\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{x}} = -2\alpha\mathrm{diag}\left(\frac{1}{\boldsymbol{q}^{\odot 2}\odot\boldsymbol{x}^{\odot 2}}\right)$$
$$- \alpha^2\left(-3\mathrm{diag}\left(\frac{1}{\boldsymbol{q}^{\odot 4}\odot\boldsymbol{x}^{\odot 4}}\right) - \boldsymbol{A}^T\boldsymbol{A}\mathrm{diag}\left(\frac{1}{\boldsymbol{q}^{\odot 2}\odot\boldsymbol{x}^{\odot 2}}\right) - \mathrm{diag}\left(\frac{1}{\boldsymbol{q}\odot\boldsymbol{x}^{\odot 2}}\odot\boldsymbol{A}^T\boldsymbol{A}\left(\frac{1}{\boldsymbol{q}}\right)\right) - 2\mathrm{diag}\left(\frac{1}{\boldsymbol{q}^{\odot 2}\odot\boldsymbol{x}^{\odot 3}}\odot\boldsymbol{A}^T\boldsymbol{y}\right)\right.$$

5.4.2. *Application to raking problem with the entropic distance*

At each step of the Newton's algorithm, we solve:

$$\nabla\boldsymbol{H}\left(\boldsymbol{\mu}^k,\boldsymbol{\lambda}^k\right)\left[\begin{pmatrix} \boldsymbol{\mu}^{k+1} \\ \boldsymbol{\lambda}^{k+1} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}^k \\ \boldsymbol{\lambda}^k \end{pmatrix}\right] = -\boldsymbol{H}\left(\boldsymbol{\mu}^k,\boldsymbol{\lambda}^k\right)$$

At step 0, we choose $\boldsymbol{\mu}^0 = \boldsymbol{x}$ and $\boldsymbol{\lambda}^0 = \boldsymbol{0}_M$, which gives us:

$$\begin{pmatrix} \boldsymbol{\mu}^1 \\ \boldsymbol{\lambda}^1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{0}_M \end{pmatrix} - (\nabla\boldsymbol{H})^+\boldsymbol{H}\left(\boldsymbol{x},\boldsymbol{0}_M\right)$$

$$= \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{0}_M \end{pmatrix} - \alpha(\nabla\boldsymbol{H})^*\left[2\boldsymbol{I}_{NM\times NM} - \alpha(\nabla\boldsymbol{H})(\nabla\boldsymbol{H})^*\right]\boldsymbol{H}\left(\boldsymbol{x},\boldsymbol{0}_M\right)$$

For the entropic distance, we have:

$$\nabla\boldsymbol{H}\left(\boldsymbol{\mu},\boldsymbol{\lambda}\right) = \begin{pmatrix} \mathrm{diag}\left(\frac{1}{\boldsymbol{q}\odot\boldsymbol{\mu}}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{m\times m} \end{pmatrix}$$

and:

$$H\left(\boldsymbol{\mu}, \boldsymbol{\lambda}\right) = \begin{pmatrix} \frac{1}{\boldsymbol{q}} \odot \log\left(\frac{\boldsymbol{\mu}}{\boldsymbol{x}}\right) + \boldsymbol{A}^T \boldsymbol{\lambda} \\ \boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{y} \end{pmatrix}$$

which gives for $\boldsymbol{\mu}^0 = \boldsymbol{x}$ and $\boldsymbol{\lambda}^0 = \boldsymbol{0}_M$:

$$\nabla H\left(\boldsymbol{x}, \boldsymbol{0}_M\right) = \begin{pmatrix} \text{diag}\left(\frac{1}{\boldsymbol{q} \odot \boldsymbol{x}}\right) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}_{m \times m} \end{pmatrix}$$

and:

$$H\left(\boldsymbol{x}, \boldsymbol{0}_M\right) = \begin{pmatrix} \boldsymbol{0}_N \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{y} \end{pmatrix}$$

Finally, we get:

$$\begin{pmatrix} \boldsymbol{\mu}^1 \\ \boldsymbol{\lambda}^1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{0}_M \end{pmatrix} - 2\alpha \begin{pmatrix} \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}^T \boldsymbol{y} \\ \boldsymbol{0}_M \end{pmatrix}$$
$$+ \alpha^2 \begin{pmatrix} \text{diag}\left(\frac{1}{\boldsymbol{q}^{\odot 2} \odot \boldsymbol{x}^{\odot 2}}\right) \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} - \text{diag}\left(\frac{1}{\boldsymbol{q}^{\odot 2} \odot \boldsymbol{x}^{\odot 2}}\right) \boldsymbol{A}^T \boldsymbol{y} - \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{A}^T \boldsymbol{y} \\ \boldsymbol{A}\text{diag}\left(\frac{1}{\boldsymbol{q} \odot \boldsymbol{x}}\right) \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}\text{diag}\left(\frac{1}{\boldsymbol{q} \odot \boldsymbol{x}}\right) \boldsymbol{A}^T \boldsymbol{y} \end{pmatrix}$$

which gives us:

$$\frac{\partial \boldsymbol{\mu}^1}{\partial \boldsymbol{y}} = 2\alpha \boldsymbol{A}^T - \alpha^2 \left(\text{diag}\left(\frac{1}{\boldsymbol{q}^{\odot 2} \odot \boldsymbol{x}^{\odot 2}}\right) \boldsymbol{A}^T + \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{A}^T \right)$$

and:

$$\frac{\partial \boldsymbol{\mu}^1}{\partial \boldsymbol{x}} = \boldsymbol{I}_{N \times N} - 2\alpha \boldsymbol{A}^T \boldsymbol{A} + \alpha^2 \left(\left[(-1)^{i==j} \frac{\left(\boldsymbol{A}^T \boldsymbol{A}\right)_{ij}}{q_i^2 x_i^2} \right] + \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{A}^T \boldsymbol{A} + 2\text{diag}\left(\frac{1}{\boldsymbol{q}^{\odot 2} \odot \boldsymbol{x}^{\odot 3}} \odot \boldsymbol{A}^T \boldsymbol{y}\right) \right)$$

5.5. Other idea

This idea is taking from [?]. The constraint in the raking problem is written as:

$$\boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{y}$$

where $\boldsymbol{A}$ is an $m * n$ matrix. If we denote $\boldsymbol{C}$ a basis of the null space of $\boldsymbol{A}$ and d the dimension of the null space, then for any vector $\boldsymbol{t}$ of $\mathbb{R}^d$, we have:

$$\boldsymbol{ACt} = \boldsymbol{0}$$

If $\hat{\boldsymbol{\mu}}$ is a solution of $\boldsymbol{A\mu} = \boldsymbol{y}$, then $\hat{\boldsymbol{\mu}} + \boldsymbol{Ct}$ is also a solution. If we denote $P = \boldsymbol{C}\left(\boldsymbol{C}^T\boldsymbol{C}\right)^{-1}\boldsymbol{C}^T$ the projection matrix of $\boldsymbol{C}$, then for any vector $\boldsymbol{v}$ of $\mathbb{R}^n$, we have:

$$\boldsymbol{APv} = \boldsymbol{0}$$

thus $\hat{\boldsymbol{\mu}} + \boldsymbol{Pv}$ is also a solution of the constraint.

If we denote:

$$h\left(\boldsymbol{x}\right) = \lambda\mu + \sum_i f_i^*\left(-v_i\lambda\right) = \lambda\mu + \sum_i x_i\left(1 + \alpha q_i v_i \lambda\right)^{\frac{1}{\alpha}}$$

and λ^* the solution of $\min_\lambda h\left(\boldsymbol{x}\right)$, we have:

$$h\left(\boldsymbol{x} + \boldsymbol{dx}\right) = h\left(\boldsymbol{x}\right) + \nabla h\left(\boldsymbol{x}\right)^T \boldsymbol{dx} + \frac{1}{2}\boldsymbol{dx}^T H\left(\boldsymbol{dx}\right)\boldsymbol{dx}$$

6. Application

We can think of two ways of doing the workflow. Let us denote $x_{i,j,k}$ and $\mu_{i,j,k}$ the initial values and the raked values for cause i, race j and county k. We suppose that:

- $i = 0$ represents all causes while $i = 1, \cdots, I$ represent the level-1 causes.
- $j = 0$ represents all races while $j = 1, \cdots, J$ represent the J different races or ethnicities.
- $k = 0$ represents the state value while $k = 1, \cdots, K$ represent the counties values.

$\mu_{0,0,0}$ and $\mu_{i,0,0}, i = 1, \cdots, I$ are known from the GBD. As it is currently implemented by the USHD team, we have four steps for the raking:

- Step 1: Minimize $\sum_{k=1}^{K} G\left(\mu_{0,0,k}, x_{0,0,k}\right)$ under the constraint $\sum_{k=1}^{K} \mu_{0,0,k} = \mu_{0,0,0}$.
- Step 2: Minimize $\sum_{i=1}^{I}\sum_{k=1}^{K} G\left(\mu_{i,0,k}, x_{i,0,k}\right)$ under the constraints $\sum_{k=1}^{K} \mu_{i,0,k} = \mu_{i,0,0}$ for all $i = 1, \cdots, I$ and $\sum_{i=1}^{I} \mu_{i,0,k} = \mu_{0,0,k}$ for all $k = 1, \cdots, K$.
- Step 3: For all $k = 1, \cdots, K$, minimize $\sum_{j=1}^{J} G\left(\mu_{0,j,k}, x_{0,j,k}\right)$ under the constraint $\sum_{j=1}^{J} \mu_{0,j,k} = \mu_{0,0,k}$.
- Step 4: For all $k = 1, \cdots, K$, minimize $\sum_{i=1}^{I}\sum_{j=1}^{J} G\left(\mu_{i,j,k}, x_{i,j,k}\right)$ under the constraints $\sum_{i=1}^{I} \mu_{i,j,k} = \mu_{0,j,k}$ for all $j = 1, \cdots, J$ and $\sum_{j=1}^{J} \mu_{i,j,k} = \mu_{i,0,k}$ for all $i = 1, \cdots, I$.

We can also do it all at once by looking for the raked values minimizing:

$$\sum_{k=1}^{K} G\left(\mu_{0,0,k}, x_{0,0,k}\right) + \sum_{i=1}^{I}\sum_{k=1}^{K} G\left(\mu_{i,0,k}, x_{i,0,k}\right) + \sum_{k=1}^{K}\sum_{j=1}^{J} G\left(\mu_{0,j,k}, x_{0,j,k}\right) + \sum_{k=1}^{K}\sum_{i=1}^{I}\sum_{j=1}^{J} G\left(\mu_{i,j,k}, x_{i,j,k}\right)$$

under the constraints:

$$\sum_{k=1}^{K} \mu_{0,0,k} = \mu_{0,0,0}$$

$$\sum_{k=1}^{K} \mu_{i,0,k} = \mu_{i,0,0} \text{ for } i = 1, \cdots, I$$

$$\sum_{i=1}^{I} \mu_{i,0,k} - \mu_{0,0,k} = 0 \text{ for } k = 1, \cdots, K$$

$$\sum_{j=1}^{J} \mu_{0,j,k} - \mu_{0,0,k} = 0 \text{ for } k = 1, \cdots, K$$

$$\sum_{i=1}^{I} \mu_{i,j,k} - \mu_{0,j,k} = 0 \text{ for } j = 1, \cdots, J \text{ and } k = 1, \cdots, K$$

$$\sum_{j=1}^{J} \mu_{i,j,k} - \mu_{i,0,k} = 0 \text{ for } i = 1, \cdots, I \text{ and } k = 1, \cdots, K$$

6.1. Additional notes

Consider the full 3D problem with single dimensional constraints:

$$\min_{\mu} \sum_{i,j,k} G\left(\mu_{i,j,k}, x_{i,j,k}\right)$$

$$\text{s.t.} \sum_{j,k} \mu_{i,j,k} = \mu_{i,0,0} \text{ for } i = 1, \cdots, I$$

This problem will satisfy consistency with GBD results but may fail to have internal consistency. When raking under a single set of constraints, there is no reason other patterns should be preserved.

Currently the team preserves the patterns that should be consistent. To understand the problem, we first consider a 2D simple case.

Consider

$$\min_{M} \sum_{i,j} G(\mu_{i,j}, x_{i,j}) \quad \text{s.t.} \quad M1 = r_c.$$

A priori there is no reason why the original structure in X should be preserved. To preserve it, we could actually add a constraint that all column sums are proportional to the original column sums (this assumes cross-entropy

for G):

$$\min_{M} \sum_{i,j} G(\mu_{i,j}, x_{i,j})$$

$$\text{s.t.} \quad M1 = r_c$$

$$1^T M = c1^T X.$$

Moreover we can easily solve for the constant, since we have to have

$$1^T M 1 = 1^T r_c = c1^T X 1 \quad \Rightarrow \quad c = \frac{1^T r_c}{1^T X 1}$$

This gives us the problem

$$\min_{M} \sum_{i,j} G(\mu_{i,j}, x_{i,j})$$

$$\text{s.t.} \quad M1 = r$$

$$1^T M = \frac{1^T r_c}{1^T X 1} 1^T X.$$

However this makes sense only if we are assuming that G is the cross-entropy loss, so that 'consistent' really means proportional.

Otherwise, following through on the previous proposal, we would solve the following:

$$\min_{c,M} \sum_{i} G(c, 1^T X_{\cdot,i}) + \sum_{i,j} G(\mu_{i,j}, x_{i,j})$$

$$\text{s.t.} \quad 1^T c = 1^T r$$

$$M1 = r$$

$$1^T M = c$$

This is a simple example of the full problem posed above.

6.2. Current implementation

The current code to do the raking is located here: Current raking functions.

In 1D, the rate is multiplied by the population, then the raking ratio is computed by dividing the total number of deaths by the sum of deaths over all locations. The raked values are obtained by multiplying the initial values with the raking ratio. We verify on a small example that we obtain the same raking ratios using the current raking function from the LSAE team and using our own raking function implementing Equation 1. The Python script is located here.

In 2D, the 1D raking is done alternatively between the two dimensions until convergence. We verify on a small example that we obtain the same raking ratios using the current raking function from the LSAE team and using our own raking function implementing Equations 2 and 3. The Python script is located here.