# Iterated Proportional Fitting Procedure and infinite products of stochastic matrices

Christophe Leuridan

juin 2023

Journées de probabilités

**Initial problem**.

Let $p \geq 2$ and $q \geq 2$ be two integers. We are given:

- two target marginals $a \in (\mathbf{R}_+^*)^p$, $b \in (\mathbf{R}_+^*)^q$ such that $a_1 + \cdots + a_p = b_1 + \cdots + b_q = 1$,
- a matrix $X_0 \in \mathcal{M}_{p,q}(\mathbf{R}_+)$. One may assume that the entries add up to 1.

Call $\Gamma(a, b, X_0)$ the set of all matrices $X \in \mathcal{M}_{p,q}(\mathbf{R}_+)$ with support contained in $\mathrm{Supp}(X_0)$ such that

$$\forall i \in [1, p], \qquad X(i,+) := \sum_{j=1}^{q} X(i,j) \text{ equals } a_i,$$

$$\forall j \in [1, q], \qquad X(+,j) := \sum_{i=1}^{p} X(i,j) \text{ equals } b_j.$$

We look for a matrix $X \in \Gamma(a, b, X_0)$ as close to $X_0$ as possible:

- we want $X$ to be diagonally equivalent to $X_0$, namely $X = D_1 X_0 D_2$ for some diagonal matrices $D_1 \in \mathcal{M}_p(\mathbf{R}_+^*)$ and $D_2 \in \mathcal{M}_q(\mathbf{R}_+^*)$ (biproportional fitting);
- equivalently, we want to minimize the relative entropy

$$D(X||X_0) := \sum_{(i,j)\in\mathrm{Supp}(X_0)} X(i,j) \ln \frac{X(i,j)}{X_0(i,j)}.$$

**Concrete application:** estimation of origin/destination matrix on a public bus line
$X(i,j) =$ proportion of population going from station $i$ to station $j$.
$X_0(i,j) =$ rough estimation.
$a_i =$ proportion of users getting in at station $i$ (well known)
$b_j =$ proportion of users getting out at station $j$ (well known)

**Probabilistic interpretation** We view $a$, $b$ and $X_0$ as probability measures on $[1, p]$, $[1, q]$ and $[1, p] \times [1, q]$. So we look at couplings of $a$ and $b$ with support contained in $\mathrm{Supp}(X_0)$.

- Existence of such couplings?
- Existence and uniqueness of a coupling $X$ minimizing $D(X||X_0)$?
- How to approach it?

Note that set $\Gamma(a, b, X_0)$ is a compact (convex) polyedron in $\mathcal{M}_{p,q}(\mathbf{R})$, possibly empty. Fourier or Farkas' lemma applies.

For $A \subset [1, p]$ and $B \subset [1, q]$, we set

$$a(A) := \sum_{i \in A} a_i, \quad b(B) := \sum_{i \in A} b_j.$$

**Cause of incompatibility:** two (non-empty) subsets $A \subset [1, p]$ and $B \subset [1, q]$ such that $X_0$ is null on $A \times B$ and $a(A) + b(B) > 1$.

$$X_0 = \begin{array}{cc} \left(\begin{array}{ccc} & | & 0 \\ - & & - \\ & | & \end{array}\right) & \begin{array}{l} \}A \\ \\ \}A^c \end{array} \\ \begin{array}{cc} \underbrace{\qquad}_{B^c} & \underbrace{\qquad}_{B} \end{array} & \end{array}$$

**Cause of criticality:** two non-empty subsets $A \subset [1, p]$ and $B \subset [1, q]$ such that $X_0$ is null on $A \times B$ and $a(A) + b(B) = 1$. This forces matrices in $\Gamma(a, b, X_0)$ to be null on $A^c \times B^c$.

### Theorem (Bacharach, Pukelsheim)

*Three situations may occur.*

1. *(Nice case) If there is no cause of incompatibility neither criticality, then* $\Gamma(a, b, X_0)$ *contains some matrix with same support as* $X_0$*.*
2. *(Critical case) If there is no cause of incompatibility but some cause of criticality, then* $\Gamma(a, b, X_0)$ *is not empty and contains only matrices with support strictly included in* $\mathrm{Supp}(X_0)$*.*
3. *(Case of incompatibility) If there is some cause of incompatibility, then,* $\Gamma(a, b, X_0)$ *is empty.*

The IPFP, introduced in 1937 by Kruithof, works as follows.
Call $\Gamma(*,*,*)$ the set of all non-negative $p \times q$ matrices having positive sum on each row or column.
For $X \in \Gamma(*,*,*)$, $i \in [1,p]$ and $j \in [1,q]$, set

$$R_i(X) := a_i^{-1} X(i,+), \quad C_j(X) := b_j^{-1} X(+,j).$$

Define $T_R : \Gamma(*,*,*) \to \Gamma(a,*,*)$, $T_C : \Gamma(*,*,*) \to \Gamma(*,b,*)$ by

$$T_R(X)(i,j) := R_i(X)^{-1} X(i,j), \quad T_C(X) := C_j(X)^{-1} X(i,j).$$

Set $X_{2n+1} = T_R(X_{2n})$, $X_{2n+2} = T_C(X_{2n+1})$ for every $n \geq 0$.
Does the sequence $(X_n)_{n \geq 0}$ thus defined converge?
If yes, the limit $X_\infty$ belongs to $\Gamma(a,b,X_0)$.

## Theorem (Bacharach, Bregman, Sinkhorn, Csiszar, Pretzel...)

*(Nice case: fast convergence) Assume that* $\Gamma(a, b, X_0)$ *contains some matrix with same support as* $X_0$. *Then*

1. *The sequences* $(R_i(X_{2n})_{n\geq 0})$ *and* $(C_j(X_{2n+1})_{n\geq 0})$ *converge to* 1 *at an at least geometric rate.*
2. *The sequence* $(X_n)_{n\geq 0}$ *converges to some matrix* $X_\infty$ *which has the same support as* $X_0$. *The rate of convergence is at least geometric.*
3. *The limit* $X_\infty$ *is the only matrix in* $\Gamma(a, b, X_0)$ *which is diagonally equivalent to* $X_0$, *or equivalently, which minimizes of* $D(Y||X_0)$ *over all* $Y \in \Gamma(a, b, X_0)$.

**Theorem (Bacharach, Bregman, Sinkhorn, Csiszar, Pretzel...)**

*Assume that* $\Gamma(a,b,X_0)$ *is not empty. Then*

1. *The sequences* $(R_i(X_{2n}))_{n\geq 0}$ *and* $((C_j(X_{2n+1}))_{n\geq 0}$ *are* $1+o(n^{-1/2})$.
2. *The sequence* $(X_n)_{n\geq 0}$ *converges to some matrix* $X_\infty$ *whose support contains the support of every matrix in* $\Gamma(a,b,X_0)$.
3. *The limit* $X_\infty$ *is the unique matrix achieving the minimum of* $D(Y||X_0)$ *over all* $Y\in\Gamma(a,b,X_0)$.

Remark: if $\Gamma(a,b,X_0)$ contains only matrices with support strictly included in $\mathrm{Supp}(X_0)$, the additional zeroes in $X_\infty$ are given by the causes of criticality. One gets the same limit with a faster convergence by replacing the corresponding entries in $X_0$ with 0.

### Theorem (Gietl and Reffel, Aas, L.)

*There exist* $r \in [1, \min(p,q)]$ *(*$r \geq 2$ *iff* $\Gamma(a,b,X_0)$ *is empty) and partitions* $\{I_1, \ldots, I_r\}$ *of* $[1,p]$ *and* $\{J_1, \ldots, J_r\}$ *of* $[1,q]$ *such that:*

1. *The ratios* $\lambda_k := b(J_k)/a(I_k)$ *increase with* $k \in [1,r]$*;*
2. $(R_i(X_{2n}))_{n \geq 0}$ *converges to* $\lambda_k$ *whenever* $i \in I_k$*;*
3. $(C_j(X_{2n+1}))_{n \geq 0}$ *converges to* $\lambda_k^{-1}$ *whenever* $j \in J_k$*;*
4. *When* $k < k'$, $X_n(i,j) = 0$ *for every* $n \geq 0$, $i \in I_k$ *and* $j \in J_{k'}$*;*
5. *When* $k < k'$, $X_n(i,j) \to 0$ *for every* $n \geq 0$, $i \in I_k$ *and* $j \in J_{k'}$*;*
6. $(X_{2n})_{n \geq 0}$ *converges to* $\arg\min_{Y \in \Gamma(a',b,X_0)} D(Y||X_0)$*, where* $a'_i = \lambda_k a_i$ *whenever* $i \in I_k$*;*
7. $(X_{2n+1})_{n \geq 0}$ *converges to* $\arg\min_{Y \in \Gamma(a,b',X_0)} D(Y||X_0)$*, where* $b'_j = \lambda_k^{-1} b_j$ *whenever* $j \in J_k$*;*

The partitions $\{I_1, \dots, I_r\}$ and $\{J_1, \dots, J_r\}$ are algorithmically determined by the causes of incompatibility and depend only on $a, b$ and $\mathrm{Supp}(X_0)$.
Replacing $a$ by $a'$ or $b$ by $b'$ transforms causes of incompatibility into causes of criticality.

Let $X \in \Gamma(*, b, *)$. Then for every $j \in [1, q]$,

$$C_j(T_R(X)) = \frac{1}{b_j}\sum_{i=1}^{p} R_i(X)^{-1}X(i,j) = \sum_{i=1}^{p}\frac{X(i,j)}{b_j}R_i(X)^{-1}.$$

In the same way, each $R_i(T_C(T_R(X)))$ is a weighted mean of the quantities $C_j(T_R(X))^{-1}$, therefore a weighted hybrid mean of the quantities $R_k(X)$... In particular,

$$\begin{aligned}[\underline{R}(X), \overline{R}(X)] &\supset [\overline{C}(T_R(X))^{-1}, \underline{C}(T_R(X))^{-1}] \\ &\supset [\underline{R}((T_C(T_R(X))), \overline{R}((T_C(T_R(X)))].\end{aligned}$$

This can be applied to $X_2, X_4, X_6, \ldots$.

### Lemma

*Let* $X \in \Gamma(*, b, *)$. *Call* $R(X)$ *the column vector with entries* $R(X_1), \dots, R(X_p)$. *Then* $R(T_C(T_R(X))) = P(X)R(X)$, *where* $P(X)$ *is the* $p \times p$ *stochastic matrix given by*

$$\forall i, k \in [1, p], \quad P(X)(i,k) = \sum_{j=1}^{q} \frac{T_R(X)(i,j)\,T_R(X)(k,j)}{a_i b_j C_j(T_R(X))}.$$

*Moreover,*

$$\forall i \in [1, p], \quad P(X)(i,i) \geq \frac{\underline{a}}{\overline{b}\ \overline{C}(T_R(X))q}$$

*and*

$$\forall i, k \in [1, p], \quad P(X)(k,i) \leq \frac{\overline{a}}{\underline{a}} P(X)(i,k).$$

Thus, the convergence of $R(X_{2n}) = P(X_{2n-2}) \cdots P(X_2)R(X_2)$ follows from the next theorem, which improves on Lorenz' stabilization theorem.

### Theorem (L.)

*Let* $(M_n)_{n \geq 1}$ *be some sequence of* $d \times d$ *stochastic matrices. Assume that there exists some constants* $\gamma > 0$, *and* $\rho \geq 1$ *such that for every* $n \geq 1$ *and* $i, j$ *in* $[1, d]$, $M_n(i, i) \geq \gamma$ *and* $M_n(i, j) \leq \rho M_n(j, i)$.

*Then the sequence* $(M_n \cdots M_1)_{n \geq 1}$ *has a finite variation, so it converges to some stochastic matrix* $L$. *Moreover, the series* $\sum_n M_n(i, j)$ *and* $\sum_n M_n(j, i)$ *converge whenever the rows of* $L$ *with indexes* $i$ *and* $j$ *are different.*