IFSBM Module 11

GUSTAVE ROUSSY
CANCER CAMPUS
GRAND PARIS

université PARIS-SACLAY

IFSBM
INSTITUT DE FORMATION SUPÉRIEURE BIOMÉDICALE

Factorisation matricielle non-négative et analyse de survie

Yoann Pradat

18/12/2024

https://github.com/gustaveroussy/IFSBM-bigdata/tree/master/Presentations/Mercredi

Sommaire

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$: ensemble d'observations (=dataset)

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$: ensemble d'observations (=dataset)

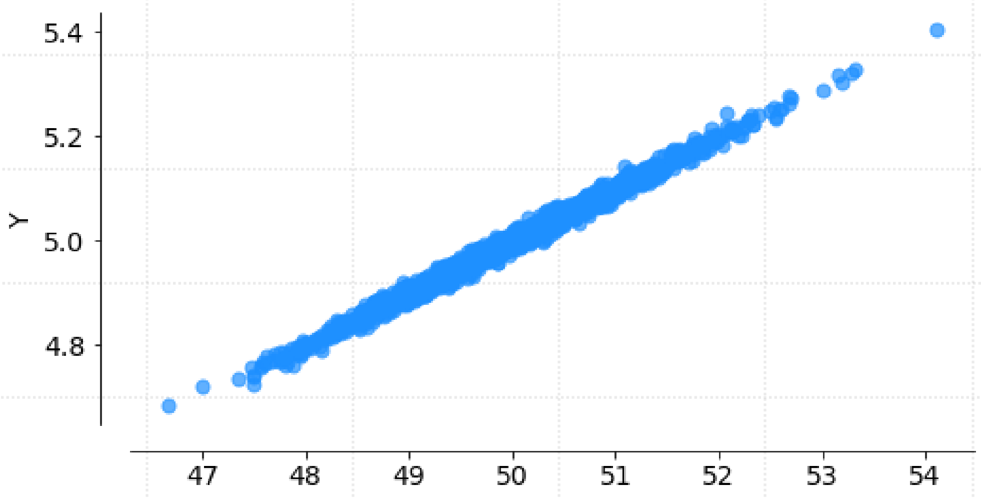
Idée Les observations sont générées par un petit nombre processus K << P

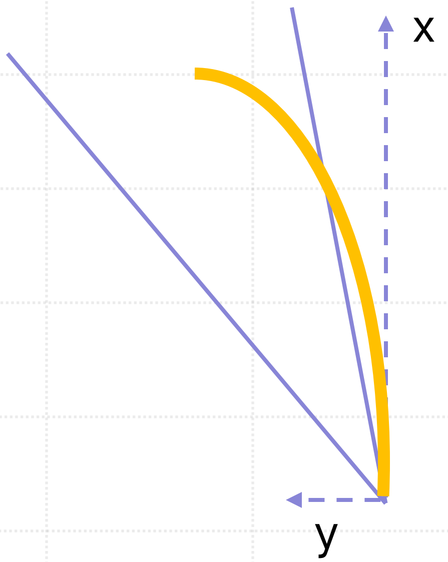
Les observations sont bien approchées par une représentation en basse dimension

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)
$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)
$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$: ensemble d'observations (=dataset)

Idée Les observations sont générées par un petit nombre processus K << P
Les observations sont bien approchées par une représentation en basse dimension





1.1 Factorisations matricielles

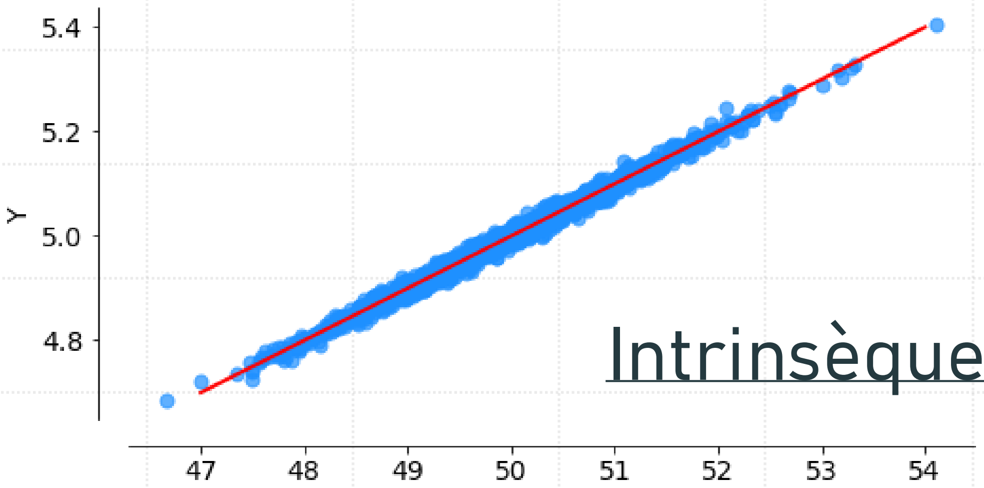
Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

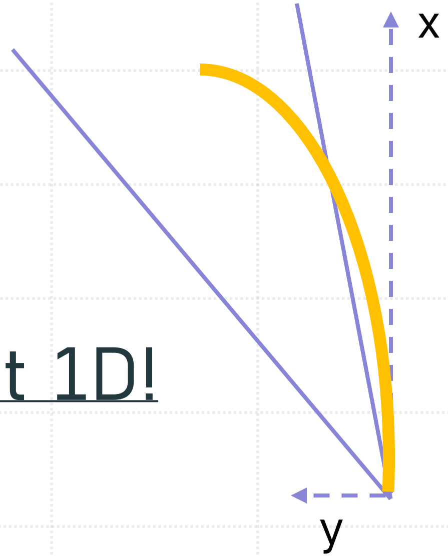
$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$: ensemble d'observations (=dataset)

Idée Les observations sont générées par un petit nombre processus K << P

Les observations sont bien approchées par une représentation en basse dimension



Intrinsèquement 1D!



1.1 Factorisations matricielles

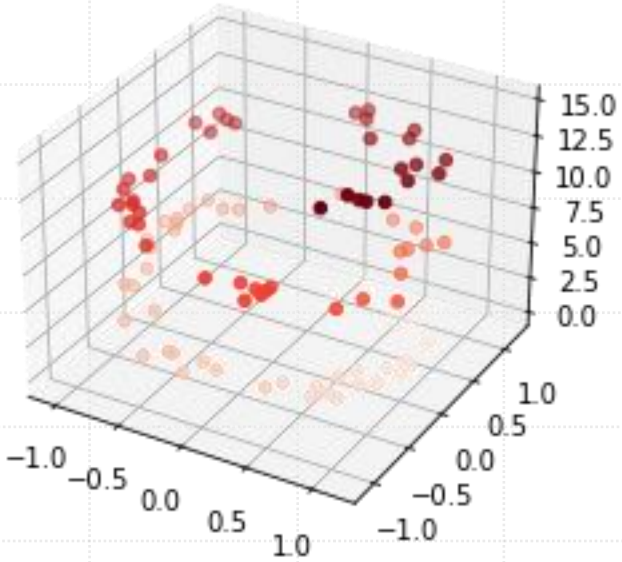
<u>Notations</u> $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$: ensemble d'observations (=dataset)

<u>Idée</u> Les observations sont générées par un petit nombre processus K << P

Les observations sont bien approchées par une représentation en basse dimension



1.1 Factorisations matricielles

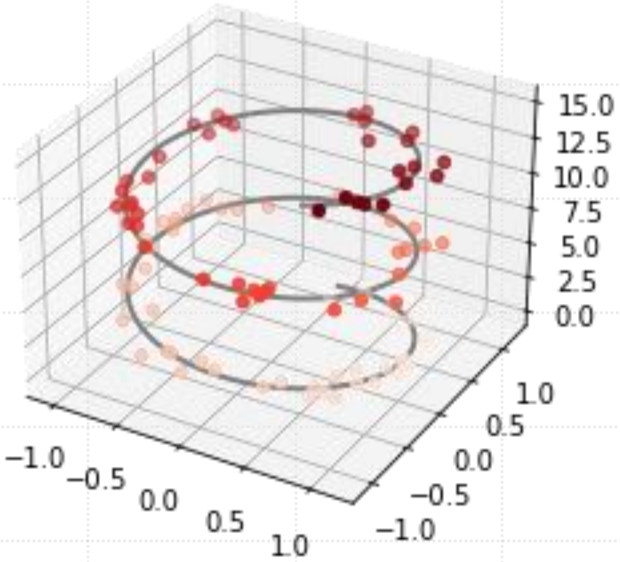
Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$: ensemble d'observations (=dataset)

Idée Les observations sont générées par un petit nombre processus K << P

Les observations sont bien approchées par une représentation en basse dimension



Intrinsèquement 1D!

1.1 Factorisations matricielles

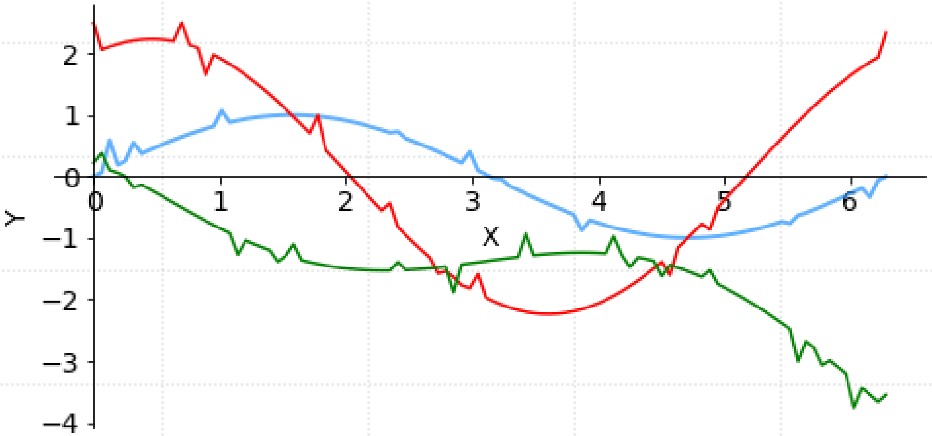
<u>Notations</u> $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

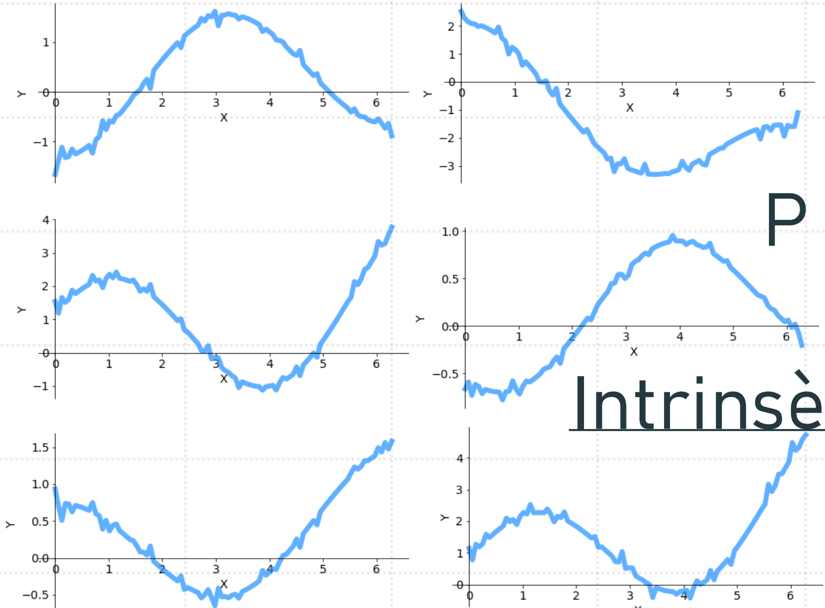
$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$: ensemble d'observations (=dataset)

<u>Idée</u> Les observations sont générées par un petit nombre processus K << P

Les observations sont bien approchées par une représentation en basse dimension



<u>Sources</u>



P = 100 points

<u>Intrinsèquement 3D!</u>

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)
$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)
$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$: ensemble d'observations (=dataset)

Idée Les observations sont générées par un petit nombre processus K << P
Les observations sont bien approchées par une représentation en basse dimension

$$\begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1N} \\ v_{21} & v_{22} & \cdots & v_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ v_{P1} & v_{p2} & \cdots & v_{PN} \end{bmatrix} \approx \begin{bmatrix} w_{11} & \cdots & w_{1K} \\ w_{21} & \cdots & w_{2K} \\ \vdots & \ddots & \vdots \\ w_{P1} & \cdots & w_{PK} \end{bmatrix} \begin{bmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{N1} & \cdots & h_{KN} \end{bmatrix}$$

1.1 Factorisations matricielles

<u>Notations</u> $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$: ensemble d'observations (=dataset)

<u>Idée</u> Les observations sont générées par un petit nombre processus K << P

Les observations sont bien approchées par une représentation en basse dimension

$$\begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1N} \\ v_{21} & v_{22} & \cdots & v_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ v_{P1} & v_{p2} & \cdots & v_{PN} \end{bmatrix} \approx \begin{bmatrix} \boxed{\begin{matrix} w_{11} \\ w_{21} \\ \vdots \\ w_{P1} \end{matrix}} & \begin{matrix} \cdots \\ \cdots \\ \ddots \\ \cdots \end{matrix} & \begin{matrix} w_{1K} \\ w_{2K} \\ \vdots \\ w_{PK} \end{matrix} \end{bmatrix} \begin{bmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{K1} & \cdots & h_{KN} \end{bmatrix}$$

Source 1 ou Axe 1

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$: ensemble d'observations (=dataset)

Idée Les observations sont générées par un petit nombre processus K << P
Les observations sont bien approchées par une représentation en basse dimension

$$\begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1N} \\ v_{21} & v_{22} & \cdots & v_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ v_{P1} & v_{p2} & \cdots & v_{PN} \end{bmatrix} \approx \begin{bmatrix} w_{11} & \cdots & w_{1K} \\ w_{21} & \cdots & w_{2K} \\ \vdots & \ddots & \vdots \\ w_{P1} & \cdots & w_{PK} \end{bmatrix} \begin{bmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{K1} & \cdots & h_{KN} \end{bmatrix}$$

Source 1 ou Axe 1 Source K ou Axe K

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$: ensemble d'observations (=dataset)

Idée Les observations sont générées par un petit nombre processus K << P

Les observations sont bien approchées par une représentation en basse dimension

Poids source 1 ou coord axe 1

$$\begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1N} \\ v_{21} & v_{22} & \cdots & v_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ v_{P1} & v_{p2} & \cdots & v_{PN} \end{bmatrix} \approx \begin{bmatrix} \boxed{\begin{matrix} w_{11} \\ w_{21} \\ \vdots \\ w_{P1} \end{matrix}} & \begin{matrix} \cdots \\ \cdots \\ \ddots \\ \cdots \end{matrix} & \boxed{\begin{matrix} w_{1K} \\ w_{2K} \\ \vdots \\ w_{PK} \end{matrix}} \end{bmatrix} \begin{bmatrix} \boxed{\begin{matrix} h_{11} & \cdots & h_{1N} \end{matrix}} \\ \begin{matrix} \vdots & \ddots & \vdots \end{matrix} \\ \begin{matrix} h_{K1} & \cdots & h_{KN} \end{matrix} \end{bmatrix}$$

Source 1 ou Axe 1 Source K ou Axe K

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$: ensemble d'observations (=dataset)

Idée Les observations sont générées par un petit nombre processus K << P

Les observations sont bien approchées par une représentation en basse dimension

Poids source 1 ou coord axe 1

$$\begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1N} \\ v_{21} & v_{22} & \cdots & v_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ v_{P1} & v_{p2} & \cdots & v_{PN} \end{bmatrix} \approx \begin{bmatrix} w_{11} & \cdots & w_{1K} \\ w_{21} & \cdots & w_{2K} \\ \vdots & \ddots & \vdots \\ w_{P1} & \cdots & w_{PK} \end{bmatrix} \begin{bmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{K1} & \cdots & h_{KN} \end{bmatrix}$$

Poids source K ou coord axe K

Source 1 ou Axe 1 Source K ou Axe K

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$: ensemble d'observations (=dataset)

Idée Les observations sont générées par un petit nombre processus K << P

Les observations sont bien approchées par une représentation en basse dimension

$$\mathbf{v} = v_1\mathbf{e_1} + v_2\mathbf{e_2} + \cdots + v_{P-1}\mathbf{e_{P-1}} + v_P\mathbf{e_P}$$

Data dimension = P

$$\boxed{\mathbf{v}_{\text{approx}} = h_1\mathbf{w_1} + \cdots + h_K\mathbf{w_K}}$$

Approx dimension = K << P

1.1 Factorisations matricielles

<u>Notations</u> $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$: ensemble d'observations (=dataset)

<u>Idée</u> Les observations sont générées par un petit nombre processus K << P

Les observations sont bien approchées par une représentation en basse dimension

$$\mathbf{v} = v_1\mathbf{e_1} + v_2\mathbf{e_2} + \cdots + v_{P-1}\mathbf{e_{P-1}} + v_P\mathbf{e_P}$$

Data dimension = P

$$\boxed{\mathbf{v}_{\text{approx}} = h_1\mathbf{w_1} + \cdots + h_K\mathbf{w_K}}$$

Approx dimension = K << P

<u>Clé:</u> distance $d(v, v_{\text{approx}})$

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$: ensemble d'observations (=dataset)

Formulation mathématique Factorisation matricielle

$\mathbf{V} \in \mathbb{R}^{\mathrm{P} \times \mathrm{N}}$ est "bien approchée" par le produit de $\mathbf{W} \in \mathbb{R}^{\mathrm{P} \times \mathrm{K}}$ et $\mathbf{H} \in \mathbb{R}^{\mathrm{K} \times \mathrm{N}}$

$$\mathbf{V} \approx \tilde{\mathbf{V}} = \mathbf{W}\mathbf{H}$$

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$: ensemble d'observations (=dataset)

Formulation mathématique Factorisation matricielle

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$ est "bien approchée" par le produit de $\mathbf{W} \in \mathbb{R}^{\mathrm{P}\times\mathrm{K}}$ et $\mathbf{H} \in \mathbb{R}^{\mathrm{K}\times\mathrm{N}}$

$$\mathbf{V} \approx \tilde{\mathbf{V}} = \mathbf{WH}$$

"Bien approchée" étant quantifié par $\mathrm{D}(\mathbf{V}||\mathbf{WH}) = \sum_{n=1}^{N} \mathrm{D}(\mathbf{v_n}||\left(\mathbf{WH}\right)_n)$

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$: ensemble d'observations (=dataset)

Formulation mathématique Factorisation matricielle

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$ est "bien approchée" par le produit de $\mathbf{W} \in \mathbb{R}^{\mathrm{P}\times\mathrm{K}}$ et $\mathbf{H} \in \mathbb{R}^{\mathrm{K}\times\mathrm{N}}$

$$\mathbf{V} \approx \tilde{\mathbf{V}} = \mathbf{WH}$$

"Bien approchée" étant quantifié par $\mathrm{D}(\mathbf{V}||\mathbf{WH}) = \sum_{n=1}^{N} \mathrm{D}(\mathbf{v_n}||\left(\mathbf{WH}\right)_n)$

Où $\mathrm{D} : \mathbb{R}^P \times \mathbb{R}^P \to \mathbb{R}$ fonction telle

$$\forall \mathbf{x}, \mathbf{y}, \quad \mathrm{D}(\mathbf{x}||\mathbf{y}) = \sum_{p=1}^{P} \mathrm{d}(x_p||y_p) \qquad \qquad \mathrm{D}(\mathbf{x}||\mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$$

1.1 Factorisations matricielles

<u>Notations</u> $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$: ensemble d'observations (=dataset)

<u>Formulation mathématique</u> Factorisation matricielle

$$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}} \text{ est "bien approchée" par le produit de } \mathbf{W} \in \mathbb{R}^{\mathrm{P}\times\mathrm{K}} \text{ et } \mathbf{H} \in \mathbb{R}^{\mathrm{K}\times\mathrm{N}}$$

$$\mathbf{V} \approx \tilde{\mathbf{V}} = \mathbf{WH}$$

$$\text{"Bien approchée" étant quantifié par } \mathrm{D}(\mathbf{V}||\mathbf{WH}) = \sum_{n=1}^{N} \mathrm{D}(\mathbf{v_n}||(\mathbf{WH})_n)$$

Divergence

$$\text{Où } \mathrm{D} : \mathbb{R}^P \times \mathbb{R}^P \to \mathbb{R} \text{ fonction telle}$$

$$\forall \mathbf{x}, \mathbf{y}, \quad \mathrm{D}(\mathbf{x}||\mathbf{y}) = \sum_{p=1}^{P} \mathrm{d}(x_p||y_p) \qquad \mathrm{D}(\mathbf{x}||\mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$$

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)

$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$: ensemble d'observations (=dataset)

Formulation mathématique Factorisation matricielle

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$ est "bien approchée" par le produit de $\mathbf{W} \in \mathbb{R}^{\mathrm{P}\times\mathrm{K}}$ et $\mathbf{H} \in \mathbb{R}^{\mathrm{K}\times\mathrm{N}}$

$$\mathbf{V} \approx \tilde{\mathbf{V}} = \mathbf{WH}$$

$$\boxed{\hat{\mathbf{W}}, \hat{\mathbf{H}} \quad \underset{\mathbf{W},\mathbf{H}\in\mathcal{W}\times\mathcal{H}}{\operatorname{argmin}} \quad \mathrm{D}(\mathbf{V}||\mathbf{WH})}$$

Divergence

$\mathcal{W} \subset \mathbb{R}^{\mathrm{P}\times\mathrm{K}}$, $\mathcal{H} \subset \mathbb{R}^{\mathrm{K}\times\mathrm{N}}$ des ensembles (convexes) de matrices.

1.1 Factorisations matricielles

Notations $\mathrm{N} \in \mathbb{N}^*$: nombre d'observations (=individus, échantillons)
$\mathrm{P} \in \mathbb{N}^*$: nombre de variables (=covariables, prédicteurs, features)
$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$: ensemble d'observations (=dataset)

Formulation mathématique Factorisation matricielle

$\mathbf{V} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$ est "bien approchée" par le produit de $\mathbf{W} \in \mathbb{R}^{\mathrm{P}\times\mathrm{K}}$ et $\mathbf{H} \in \mathbb{R}^{\mathrm{K}\times\mathrm{N}}$

$$\mathbf{V} \approx \tilde{\mathbf{V}} = \mathbf{W}\mathbf{H}$$

$$\boxed{\hat{\mathbf{W}}, \hat{\mathbf{H}} \quad \underset{\mathbf{W},\mathbf{H} \in \boxed{\mathcal{W}\times\mathcal{H}}}{\operatorname{argmin}} \quad \boxed{\mathrm{D}}(\mathbf{V}||\mathbf{W}\mathbf{H})}$$

Divergence

$\mathcal{W} \subset \mathbb{R}^{\mathrm{P}\times\mathrm{K}}$, $\mathcal{H} \subset \mathbb{R}^{\mathrm{K}\times\mathrm{N}}$ des ensembles (convexes) de matrices.

1.1 Factorisations matricielles

Différentes formes de factorisation

1. Analyse en composante principales (PCA)
    - $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2$
    - $\mathcal{W} = \mathbb{R}^{\mathrm{P}\times\mathrm{K}}$, $\mathcal{H} = \mathbb{R}^{\mathrm{K}\times\mathrm{N}}$.

1.1 Factorisations matricielles

Différentes formes de factorisation

1. Analyse en composante principales (PCA)

- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2$ -> « facilement » résolu par SVD (Eckart-Young, 1936)
- $\mathcal{W} = \mathbb{R}^{\mathrm{P}\times\mathrm{K}}$, $\mathcal{H} = \mathbb{R}^{\mathrm{K}\times\mathrm{N}}$. $\mathbf{V} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{Q}$ alors $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$ $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

1.1 Factorisations matricielles

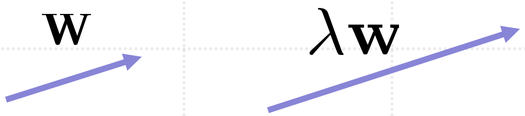
Différentes formes de factorisation

1. Analyse en composante principales (PCA)

- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2$ -> « facilement » résolu par SVD (Eckart-Young, 1936)
- $\mathcal{W} = \mathbb{R}^{\mathrm{P}\times\mathrm{K}}, \mathcal{H} = \mathbb{R}^{\mathrm{K}\times\mathrm{N}}.$ $\mathbf{V} = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}$ alors $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$ $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

NOTE 1: invariance d'échelle

$$\begin{bmatrix} w_{11} & \cdots & w_{1K} \\ w_{21} & \cdots & w_{2K} \\ \vdots & \ddots & \vdots \\ w_{P1} & \cdots & w_{PK} \end{bmatrix} \begin{bmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{N1} & \cdots & h_{KN} \end{bmatrix} = \begin{bmatrix} \lambda_1 w_{11} & \cdots & \lambda_K w_{1K} \\ \lambda_1 w_{21} & \cdots & \lambda_K w_{2K} \\ \vdots & \ddots & \vdots \\ \lambda_1 w_{P1} & \cdots & \lambda_K w_{PK} \end{bmatrix} \begin{bmatrix} \lambda_1^{-1} h_{11} & \cdots & \lambda_1^{-1} h_{1N} \\ \vdots & \ddots & \vdots \\ \lambda_K^{-1} h_{N1} & \cdots & \lambda_K^{-1} h_{KN} \end{bmatrix}$$



1.1 Factorisations matricielles

Différentes formes de factorisation

1. Analyse en composante principales (PCA)

- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2$ -> « facilement » résolu par SVD (Eckart-Young, 1936)
- $\mathcal{W} = \mathbb{R}^{\mathrm{P}\times\mathrm{K}}, \mathcal{H} = \mathbb{R}^{\mathrm{K}\times\mathrm{N}}.$ $\mathbf{V} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{Q}$ alors $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$ $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

NOTE 1: invariance d'échelle

NOTE 2: invariance de rotation

$$\begin{bmatrix} w_{11} & \cdots & w_{1K} \\ w_{21} & \cdots & w_{2K} \\ \vdots & \ddots & \vdots \\ w_{P1} & \cdots & w_{PK} \end{bmatrix} \begin{bmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{N1} & \cdots & h_{KN} \end{bmatrix} = \begin{bmatrix} w_{1q} & \cdots & w_{s1} \\ w_{2q} & \cdots & w_{s2} \\ \vdots & \ddots & \vdots \\ w_{Pq} & \cdots & w_{sK} \end{bmatrix} \begin{bmatrix} h_{q1} & \cdots & h_{qN} \\ \vdots & \ddots & \vdots \\ h_{Ns} & \cdots & h_{sN} \end{bmatrix}$$

$$1 \leq s, q \leq K, \quad s \neq q$$

1.1 Factorisations matricielles

Différentes formes de factorisation

1. Analyse en composante principales (PCA)

- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2$ -> « facilement » résolu par SVD (Eckart-Young, 1936)
- $\mathcal{W} = \mathbb{R}^{\mathrm{P}\times\mathrm{K}}, \mathcal{H} = \mathbb{R}^{\mathrm{K}\times\mathrm{N}}.$ $\mathbf{V} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{Q}$ alors $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$ $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

NOTE 1: invariance d'échelle

NOTE 2: invariance de rotation

Fixations des invariances:

- rotation, axes triés par valeur propre dec. $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_K \geq 0$
- échelle, axes “unitaires” i.e $\mathbf{w_k}^\top \mathbf{w_k} = 1, \quad 1 \leq k \leq K$

1.1 Factorisations matricielles

Différentes formes de factorisation

1. Analyse en composante principales (PCA)

- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2$ -> « facilement » résolu par SVD (Eckart-Young, 1936)
- $\mathcal{W} = \mathbb{R}^{\mathrm{P}\times\mathrm{K}}, \mathcal{H} = \mathbb{R}^{\mathrm{K}\times\mathrm{N}}.$ $\mathbf{V} = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}$ alors $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$ $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

Propriétés PCA $\mathbf{w_q}^\top \mathbf{w_s} = 0, \quad 1 \leq s,q \leq K, \quad s \neq q$

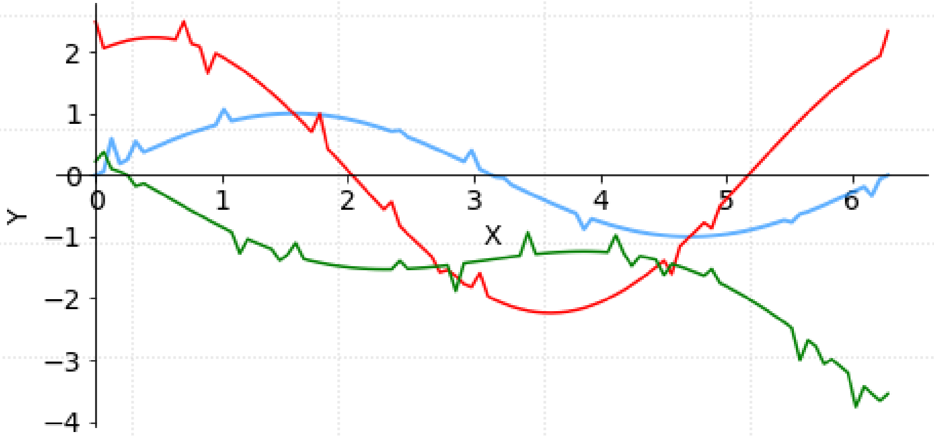
1.1 Factorisations matricielles

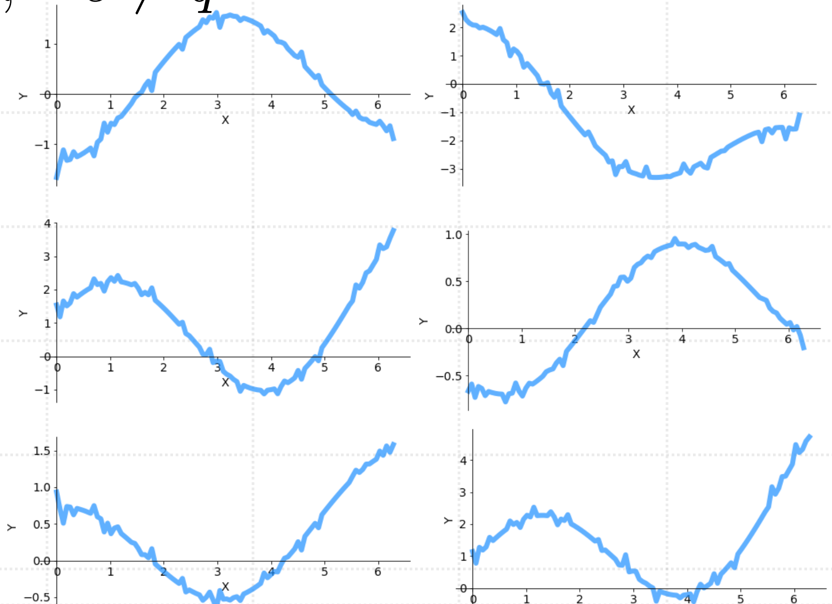
Différentes formes de factorisation

1. Analyse en composante principales (PCA)

- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \frac{1}{2}||\mathbf{x}-\mathbf{y}||_2$ -> « facilement » résolu par SVD (Eckart-Young, 1936)
- $\mathcal{W} = \mathbb{R}^{\mathrm{P}\times\mathrm{K}}, \mathcal{H} = \mathbb{R}^{\mathrm{K}\times\mathrm{N}}.$ $\mathbf{V} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{Q}$ alors $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$ $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

Propriétés PCA $\mathbf{w_q}^\top \mathbf{w_s} = 0, \quad 1 \leq s,q \leq K, \quad s \neq q$





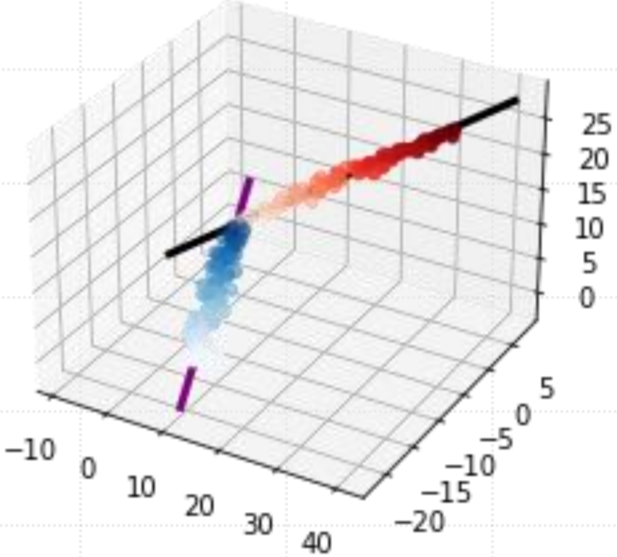
1.1 Factorisations matricielles

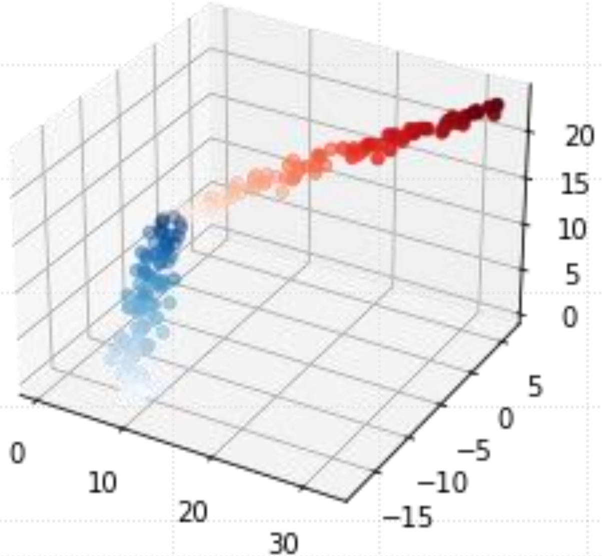
Différentes formes de factorisation

1. Analyse en composante principales (PCA)

- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2$ -> « facilement » résolu par SVD (Eckart-Young, 1936)
- $\mathcal{W} = \mathbb{R}^{\mathrm{P}\times\mathrm{K}}, \mathcal{H} = \mathbb{R}^{\mathrm{K}\times\mathrm{N}}.$ $\mathbf{V} = \mathbf{P\Lambda Q}$ alors $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$ $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

Propriétés PCA $\mathbf{w_q}^\top \mathbf{w_s} = 0, \quad 1 \leq s, q \leq K, \quad s \neq q$





1.2 Factorisation non négative

Différentes formes de factorisation

2. Factorisation matricielle non-négative (NMF)

Paatero et Tapper. *Environmentrics.* 1994.

- $\mathcal{W} = \mathbb{R}_{+}^{\mathrm{P}\times\mathrm{K}}$, $\mathcal{H} = \mathbb{R}_{+}^{\mathrm{K}\times\mathrm{N}}$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_F = \sqrt{\sum_{p=1}^{P}(x_p - y_p)^2}$

1.2 Factorisation non négative

Différentes formes de factorisation

2. Factorisation matricielle non-négative (NMF)

Paatero et Tapper. *Environmentrics.* 1994.

Lee et Seung. 2001.

- $\mathcal{W} = \mathbb{R}_{+}^{\mathrm{P\times K}}, \mathcal{H} = \mathbb{R}_{+}^{\mathrm{K\times N}}$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_F = \sqrt{\sum_{p=1}^{P}(x_p - y_p)^2}$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_F = \sqrt{\sum_{p=1}^{P}(x_p - y_p)^2}_P$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \mathrm{D_{KL}}(\mathbf{x}|\mathbf{y}) = \sum_{p=1} x_p \log(\frac{x_p}{y_p}) - x_p + y_p$

1.2 Factorisation non négative

Différentes formes de factorisation

2. Factorisation matricielle non-négative (NMF)

Paatero et Tapper. *Environmentrics*. 1994.

- $\mathcal{W} = \mathbb{R}_{+}^{\mathrm{P}\times\mathrm{K}}, \ \mathcal{H} = \mathbb{R}_{+}^{\mathrm{K}\times\mathrm{N}}$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_F = \sqrt{\sum_{p=1}^{P}(x_p - y_p)^2}$

Lee et Seung. 2001.

- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_F = \sqrt{\sum_{p=1}^{P}(x_p - y_p)^2}$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \mathrm{D}_{\mathrm{KL}}(\mathbf{x}|\mathbf{y}) = \sum_{p=1}^{P} x_p \log(\frac{x_p}{y_p}) - x_p + y_p$

Cichocki, Zdunek, et Amari. ICBSS. 2006

- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \mathrm{D}_{\alpha}(\mathbf{x}|\mathbf{y}) = \frac{1}{\alpha(1-\alpha)}\sum_{p=1}^{P} \alpha x_p + (1-\alpha)y_p - x_p^{\alpha}y_p^{1-\alpha}$

Cichocki, Zdunek Phan, et Amari. John Wiley & Sons. 2008

Févotte et Idier. arXiv. 2011

- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \mathrm{D}_{\beta}(\mathbf{x}|\mathbf{y}) = \frac{1}{\beta(\beta-1)}\sum_{p=1}^{P} (x_p^{\beta} + (\beta-1)y_p^{\beta} - \beta x_p y_p^{\beta-1})$

1.2 Factorisation non négative

Différentes formes de factorisation

2. Factorisation matricielle non-négative (NMF)

Févotte et Idier. arXiv. 2011

- $\mathcal{W} = \mathbb{R}_+^{\mathrm{P}\times\mathrm{K}}$, $\mathcal{H} = \mathbb{R}_+^{\mathrm{K}\times\mathrm{N}}$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \mathrm{D}_\beta(\mathbf{x}|\mathbf{y})$

$$\mathrm{D}_{\beta=0}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Itakura-Saito}},$$
$$\mathrm{D}_{\beta=1}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Kullback-Leibler}},$$
$$\mathrm{D}_{\beta=2}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Euclidean}}$$

1.2 Factorisation non négative

Différentes formes de factorisation

2. Factorisation matricielle non-négative (NMF)

Févotte et Idier. arXiv. 2011

- $\mathcal{W} = \mathbb{R}_{+}^{\mathrm{P}\times\mathrm{K}}$, $\mathcal{H} = \mathbb{R}_{+}^{\mathrm{K}\times\mathrm{N}}$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \mathrm{D}_{\beta}(\mathbf{x}|\mathbf{y})$

$$\mathrm{D}_{\beta=0}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Itakura-Saito}},$$
$$\mathrm{D}_{\beta=1}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Kullback-Leibler}},$$
$$\mathrm{D}_{\beta=2}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Euclidean}}$$

Algorithme de résolution

$C(\mathbf{W},\mathbf{H}) := \mathrm{D}_{\beta}(\mathbf{V}||\mathbf{WH})$ non convexe mais ...

1.2 Factorisation non négative

Différentes formes de factorisation

2. Factorisation matricielle non-négative (NMF)

Févotte et Idier. arXiv. 2011

- $\mathcal{W} = \mathbb{R}_+^{\mathrm{P}\times\mathrm{K}}$, $\mathcal{H} = \mathbb{R}_+^{\mathrm{K}\times\mathrm{N}}$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \mathrm{D}_\beta(\mathbf{x}|\mathbf{y})$

$$\mathrm{D}_{\beta=0}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Itakura-Saito}},$$
$$\mathrm{D}_{\beta=1}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Kullback-Leibler}},$$
$$\mathrm{D}_{\beta=2}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Euclidean}}$$

Algorithme de résolution

$C(\mathbf{W},\mathbf{H}) := \mathrm{D}_\beta(\mathbf{V}||\mathbf{WH})$ non convexe mais ... $C(\mathbf{W},\cdot)$ et $C(\cdot,\mathbf{H})$ convexes

1.2 Factorisation non négative

Différentes formes de factorisation

2. Factorisation matricielle non-négative (NMF)

Févotte et Idier. arXiv. 2011

- $\mathcal{W} = \mathbb{R}_{+}^{P\times K}$, $\mathcal{H} = \mathbb{R}_{+}^{K\times N}$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \mathrm{D}_{\beta}(\mathbf{x}|\mathbf{y})$

$$\mathrm{D}_{\beta=0}(\mathbf{x}|\mathbf{y}) = D_{\text{Itakura-Saito}},$$
$$\mathrm{D}_{\beta=1}(\mathbf{x}|\mathbf{y}) = D_{\text{Kullback-Leibler}},$$
$$\mathrm{D}_{\beta=2}(\mathbf{x}|\mathbf{y}) = D_{\text{Euclidean}}$$

Algorithme de résolution

$C(\mathbf{W},\mathbf{H}) := \mathrm{D}_{\beta}(\mathbf{V}||\mathbf{WH})$ non convexe mais ... $C(\mathbf{W},\cdot)$ et $C(\cdot,\mathbf{H})$ convexes

Algorithm 1: algorithme NMF générique

Result: $\mathbf{W},\mathbf{H}$ solving NMF

initialisation $\mathbf{W^1},\mathbf{H^1}$;

for t = *1:T* **do**

$\mathbf{H}^{t+1} = \mathrm{argmin}_{\mathbf{H}\leq 0}\, \mathrm{D}_{\beta}(\mathbf{V}||\mathbf{W^t H})$;

$\mathbf{W}^{t+1} = \mathrm{argmin}_{\mathbf{W}\leq 0}\, \mathrm{D}_{\beta}(\mathbf{V}||\mathbf{W H^{t+1}})$;

end

1.2 Factorisation non négative

Différentes formes de factorisation

2. Factorisation matricielle non-négative (NMF)

Févotte et Idier. arXiv. 2011

- $\mathcal{W} = \mathbb{R}_{+}^{\mathrm{P}\times\mathrm{K}}$, $\mathcal{H} = \mathbb{R}_{+}^{\mathrm{K}\times\mathrm{N}}$
- $\mathrm{D}(\mathbf{x}||\mathbf{y}) = \mathrm{D}_\beta(\mathbf{x}|\mathbf{y})$

$$\mathrm{D}_{\beta=0}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Itakura-Saito}},$$
$$\mathrm{D}_{\beta=1}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Kullback-Leibler}},$$
$$\mathrm{D}_{\beta=2}(\mathbf{x}|\mathbf{y}) = D_{\mathrm{Euclidean}}$$

Algorithme de résolution

$C(\mathbf{W},\mathbf{H}) := \mathrm{D}_\beta(\mathbf{V}||\mathbf{WH})$ non convexe mais ... $C(\mathbf{W},\cdot)$ et $C(\cdot,\mathbf{H})$ convexes

Propriétés:

- invariances d'échelle et de rotation
- pas de minimum global assure
- coefficients positifs
- axes non orthogonaux

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

1. Classification de tumeurs — Brunet, Tamayo, Golub, et Mesirov. PNAS, 2004.

Algorithm 1: Selection de K dans la NMF

Result: Matrice de consensus pour K donné

for $m = 1 : M$ **do**

  initialisation $\mathbf{W}^{(\mathbf{m},\mathbf{K})}, \mathbf{H}^{(\mathbf{m},\mathbf{K})}$;

  $\mathbf{H}^{*,(m,K)}, \mathbf{W}^{*,(m,K)} \leftarrow \mathrm{NMF}(\mathbf{V}, K, \mathbf{W}^{(\mathbf{m},\mathbf{K})}, \mathbf{H}^{(\mathbf{m},\mathbf{K})})$;

  assignation éch. i à un cluster en prenant max de la colonne $\mathbf{H}_{:,i}$;

  calcul matrice $\mathbf{C}^{(m,K)}$ avec $\mathbf{C}^{(m,K)}_{i,j} = 1$ si i et j même cluster;

end

$$\mathbf{C}^{(m,K)} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & \cdots & 1 \end{bmatrix}$$

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

1. Classification de tumeurs Brunet, Tamayo, Golub, et Mesirov. PNAS, 2004.

Algorithm 1: Selection de K dans la NMF

Result: Matrice de consensus pour K donné
for $m = 1 : M$ **do**
  initialisation $\mathbf{W^{(m,K)}}, \mathbf{H^{(m,K)}}$;
  $\mathbf{H}^{*,(m,K)}, \mathbf{W}^{*,(m,K)} \leftarrow \text{NMF}(\mathbf{V}, K, \mathbf{W^{(m,K)}}, \mathbf{H^{(m,K)}})$;
  assignation éch. i à un cluster en prenant max de la colonne $\mathbf{H}_{:,i}$;
  calcul matrice $\mathbf{C}^{(m,K)}$ avec $\mathbf{C}_{i,j}^{(m,K)} = 1$ si i et j même cluster;
end
$\bar{\mathbf{C}}^{(K)} = \frac{1}{m} \sum_{m=1}^{M} \mathbf{C}^{(m,K)}$;
clustering hiérarchique pour réordonner la matrice $\bar{\mathbf{C}}^{(K)}$;

1.3 Applications NMF

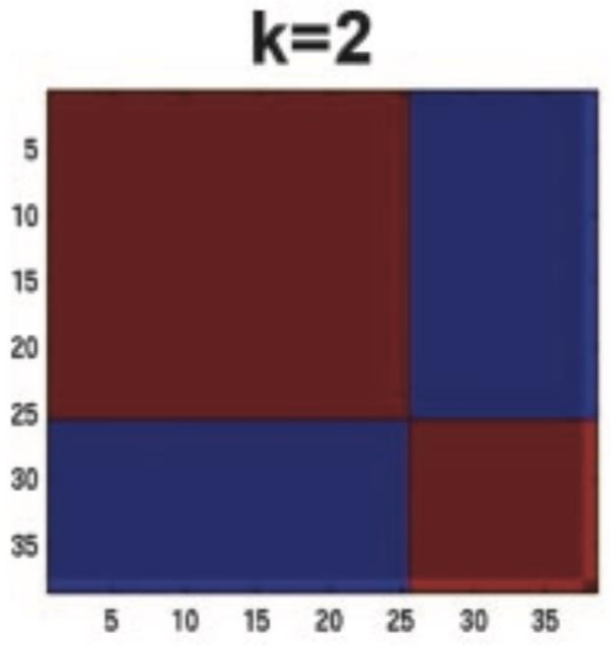
Différentes applications

Comment determiner le nombre optimal de sources K ?!

1. Classification de tumeurs

Brunet, Tamayo, Golub, et Mesirov. PNAS, 2004.

$$\mathbf{V} = \begin{bmatrix} v_{1,1} & \cdots & v_{1,38} \\ \vdots & \ddots & \vdots \\ v_{999,1} & \cdots & v_{999,38} \end{bmatrix}$$



38 éch. moelle
-> 11 AML
-> 8 T-ALL
-> 19 B-ALL
999 genes (HU6800)

1.3 Applications NMF

Différentes applications

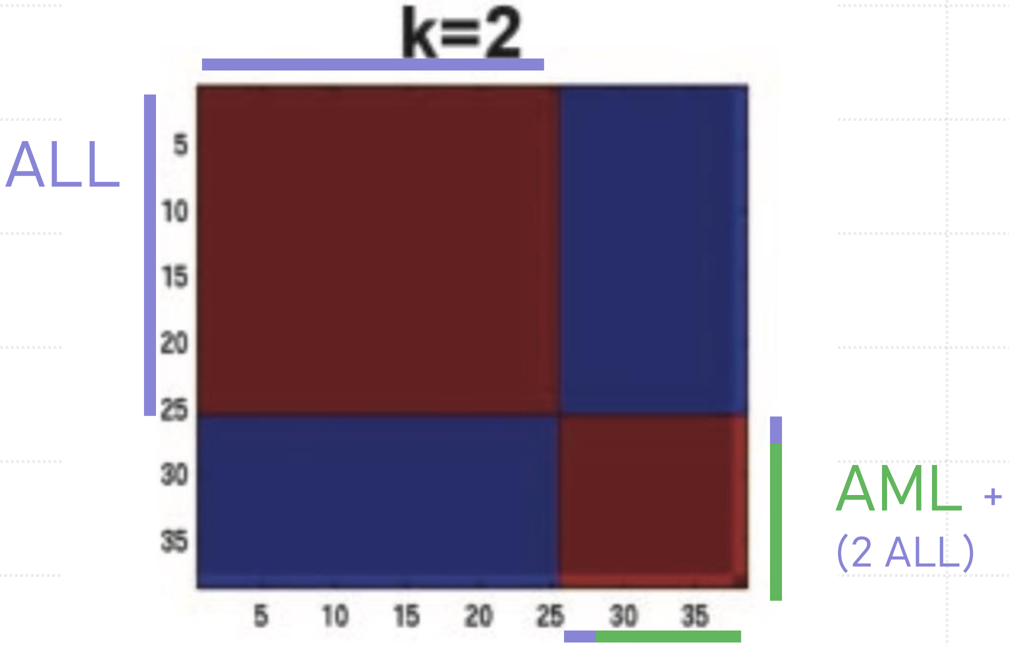
Comment determiner le nombre optimal de sources K ?!

1. Classification de tumeurs

Brunet, Tamayo, Golub, et Mesirov. PNAS, 2004.

$$\mathbf{V} = \begin{bmatrix} v_{1,1} & \cdots & v_{1,38} \\ \vdots & \ddots & \vdots \\ v_{999,1} & \cdots & v_{999,38} \end{bmatrix}$$

38 éch. moelle
-> 11 AML
-> 8 T-ALL
-> 19 B-ALL
999 genes (HU6800)

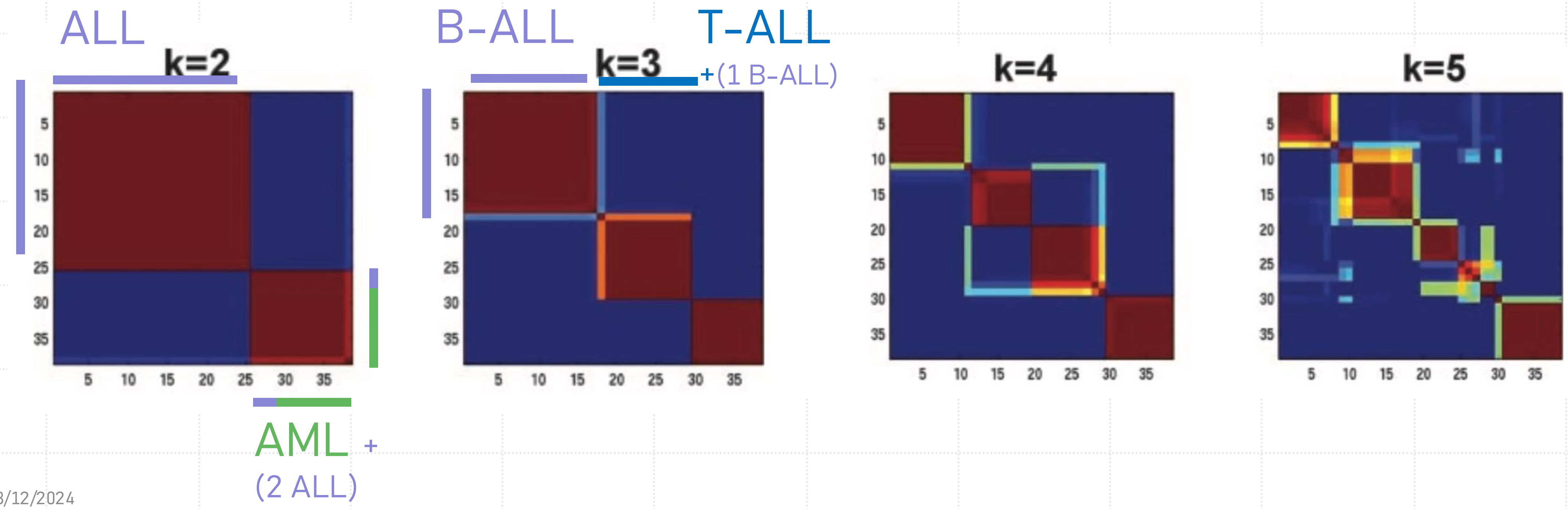


1.3 Applications NMF

Différentes applications

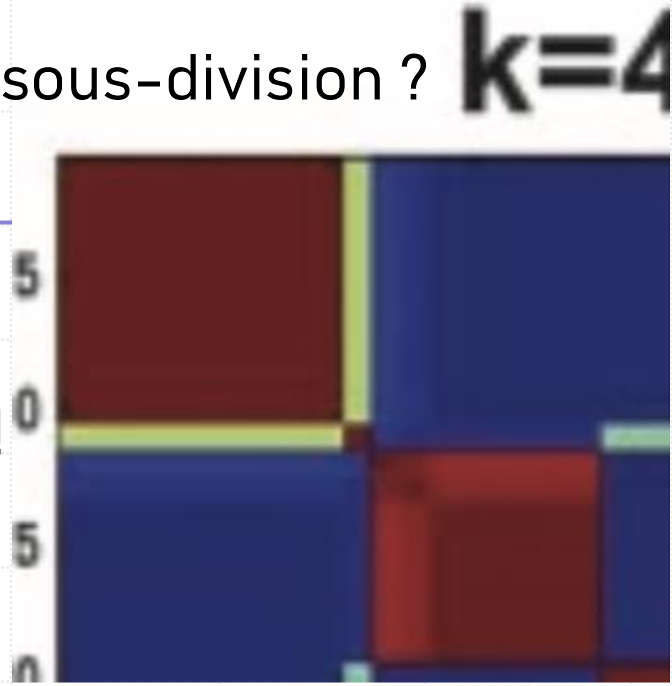
Comment determiner le nombre optimal de sources K ?!

1. Classification de tumeurs   Brunet, Tamayo, Golub, et Mesirov. PNAS, 2004.



1.3 Applications NMF

Nouvelle sous-division ?

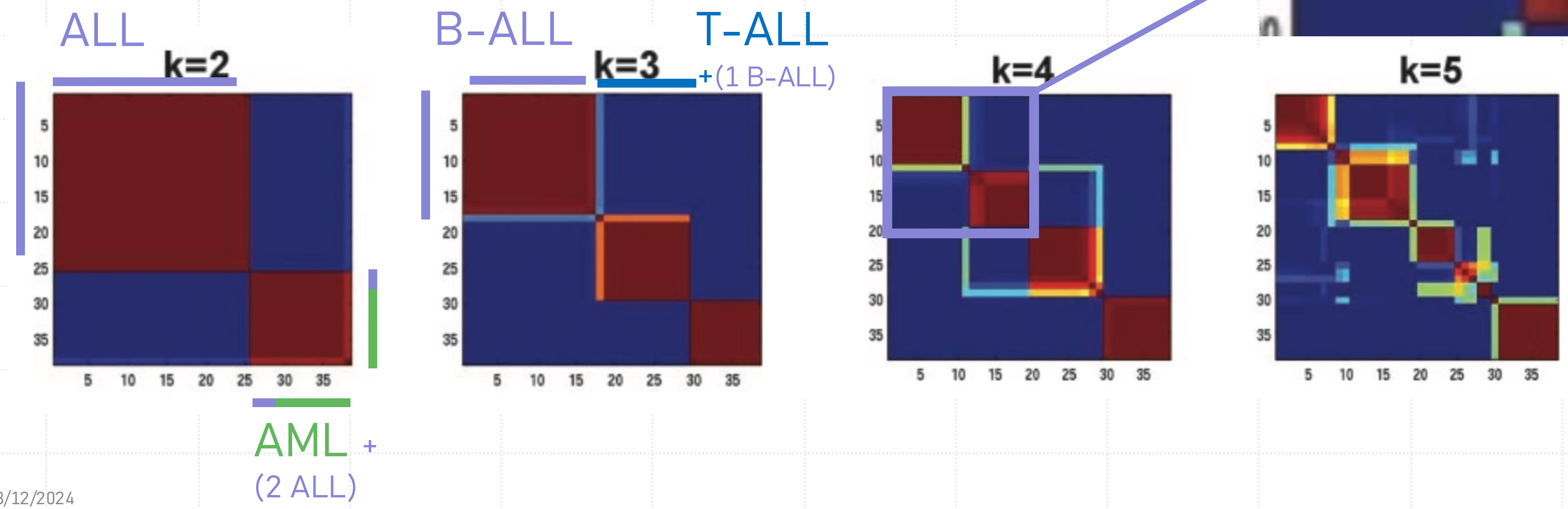


Différentes applications

Comment determiner le nombre optimal de sources K ?!

1. Classification de tumeurs

Brunet, Tamayo, Golub, et Mesirov. PNAS, 2004.



1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012
Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Idée: Il existe un petit nombre de processus mutagéniques (endogène ou exogène) ayant chacun une empreinte particulière.

Chaque processus affecterait des **contextes** particuliers

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012
Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Contexte ? 4 nucleotides A, C, G, T donc 4x3 = 12 mutations monobase

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles [Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012] [Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013]

Contexte ? 4 nucleotides A, C, G, T donc 4x3 = 12 mutations monobase
+ 1 nucl. à g. et 1 nucl. à d. donc 4x4 = 16 contextes

12 x 16 = 192 mutations monobase

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012
Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Contexte ? 4 nucleotides A, C, G, T donc 4x3 = 12 mutations monobase
+ 1 nucl. à g. et 1 nucl. à d. donc 4x4 = 16 contextes

12 x 16 = 192 mutations monobase

5′ + strand 3′ →

AATCGCGTTA

TTAGCGCAAT

← 3′ - strand 5′

1.3 Applications NMF

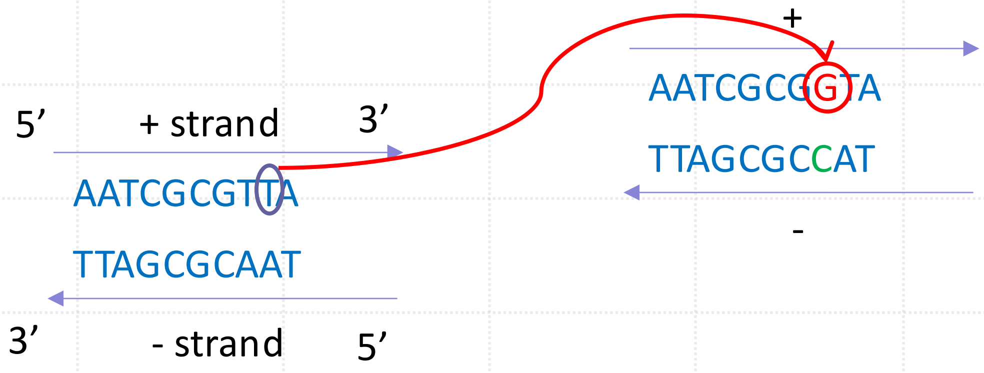
Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012
Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Contexte ? 4 nucleotides A, C, G, T donc 4x3 = 12 mutations monobase
+ 1 nucl. à g. et 1 nucl. à d. donc 4x4 = 16 contextes

12 x 16 = 192 mutations monobase



Mutation T>G on + strand

1.3 Applications NMF

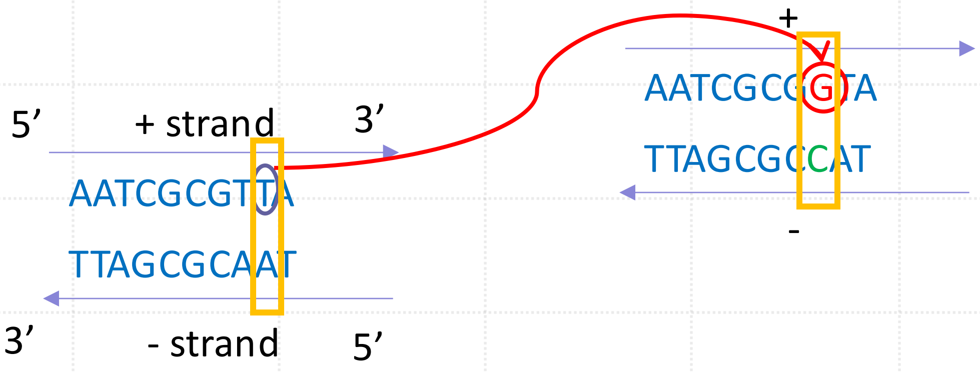
Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles
Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012
Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Contexte ? 4 nucleotides A, C, G, T donc 4x3 = 12 mutations monobase
+ 1 nucl. à g. et 1 nucl. à d. donc 4x4 = 16 contextes

12 x 16 = 192 mutations monobase



Mutation T>G on + strand T - A => G - C

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012
Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Contexte ? 4 nucleotides A, C, G, T donc 4x3 = 12 mutations monobase
+ 1 nucl. à g. et 1 nucl. à d. donc 4x4 = 16 contextes

12 x 16 = 192 mutations monobase

5' + strand 3'
AATCGCGTTA
TTAGCGCAAT
3' - strand 5'

+
AATCGCGGTA
TTAGCGCCAT
-

Mutation T>G on + strand T - A => G - C

+
AATCGCGGTA
TTAGCGCCAT
-

Mutation A>C on - strand

1.3 Applications NMF

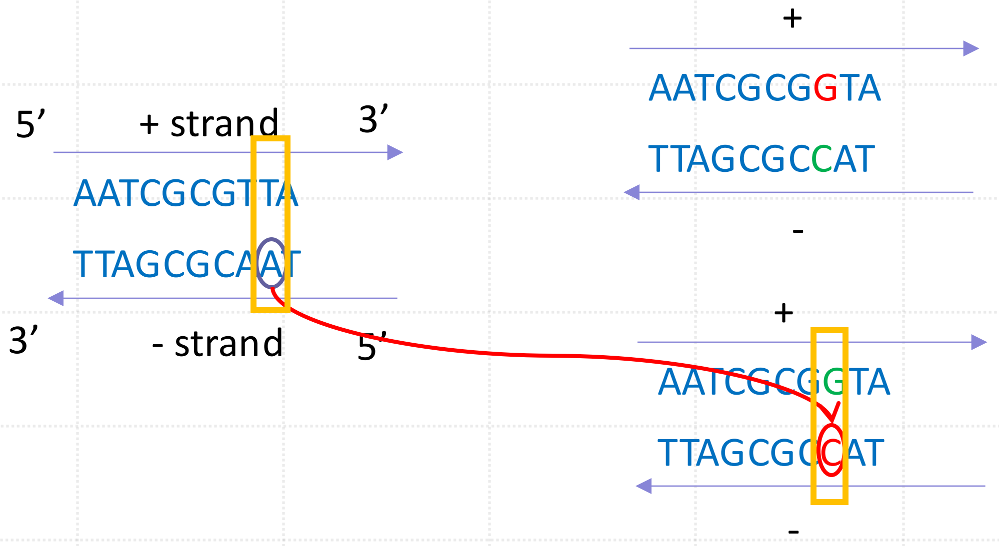
Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012
Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Contexte ? 4 nucleotides A, C, G, T donc 4x3 = 12 mutations monobase
+ 1 nucl. à g. et 1 nucl. à d. donc 4x4 = 16 contextes

12 x 16 = 192 mutations monobase



Mutation T>G on + strand T - A => G - C

Mutation A>C on - strand T - A => G - C

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles — Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012; Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Contexte ? 4 nucleotides A, C, G, T donc 4x3 = 12 mutations monobase
+ 1 nucl. à g. et 1 nucl. à d. donc 4x4 = 16 contextes

12 x 16 = 192 mutations monobase

5' + strand 3'
AATCGCGTTA
TTAGCGCAAT
3' - strand 5'

\+
AATCGCGGTA
TTAGCGCCAT
\-

Mutation T>G on + strand T - A => G - C

Categorized it as G [T>G] T

Pyrimidine first

=

\+
AATCGCGGTA
TTAGCGCCAT
\-

Mutation A>C on - strand T - A => G - C

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012
Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Contexte ? 4 nucleotides A, C, G, T donc 4x3 = 12 mutations monobase
+ 1 nucl. à g. et 1 nucl. à d. donc 4x4 = 16 contextes

12 x 16 = 192 mutations monobase

4x3/2 = 6 mutations monobase
+ 1 nucl. à g. et 1 nucl. à d. donc 4x4 = 16 contextes

12 x 8 = **96 mutations monobase**

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

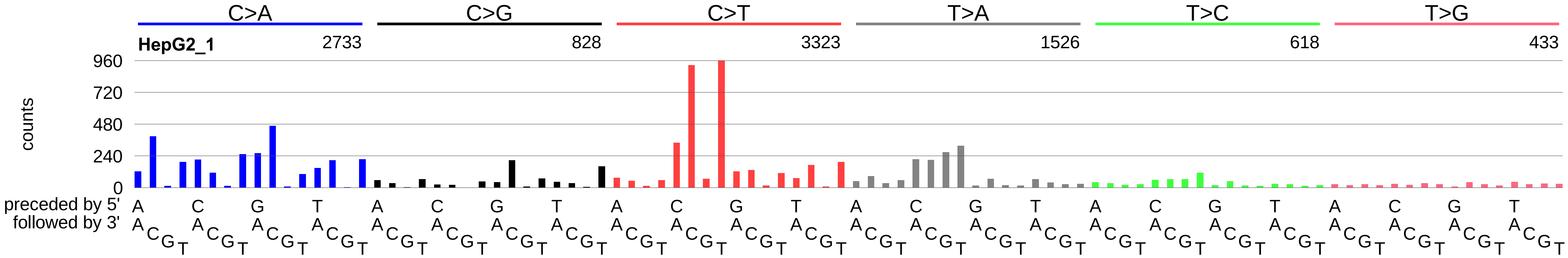
2. Signatures mutationnelles
Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012
Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Contexte ?

4x3/2 = 6 mutations monobase

+ 1 nucl. à g. et 1 nucl. à d. donc 4x4 = 16 contextes

12 x 8 = **96 mutations monobase**



1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

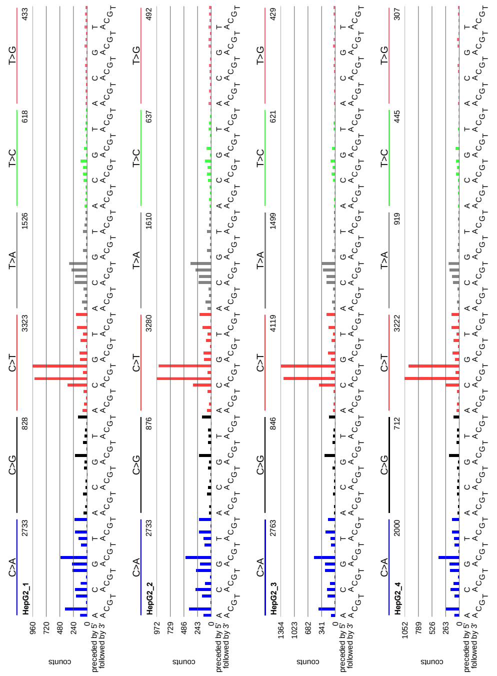
2. Signatures mutationnelles

Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012

Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

$$\mathbf{V} = \begin{bmatrix} v_{1,1} & \cdots & v_{1,N} \\ \vdots & \ddots & \vdots \\ v_{96,1} & \cdots & v_{96,N} \end{bmatrix}$$

Exemple: N=4 génomes



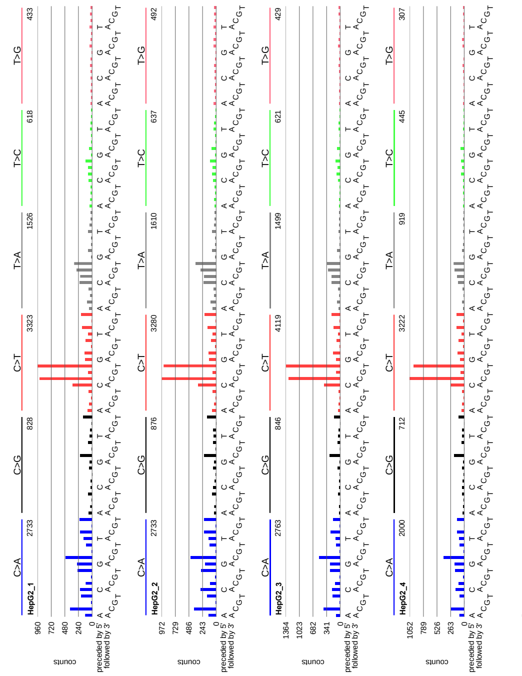
1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

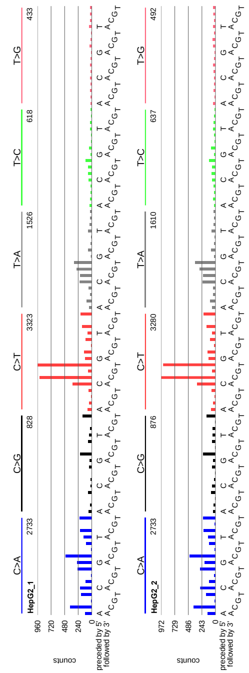
2. Signatures mutationnelles
Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012
Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

$$\mathbf{V} = \begin{bmatrix} v_{1,1} & \cdots & v_{1,N} \\ \vdots & \ddots & \vdots \\ v_{96,1} & \cdots & v_{96,N} \end{bmatrix}$$



$\approx$

$\mathbf{W} =$



$$\begin{bmatrix} h_{1,1} & \cdots & h_{1,N} \\ h_{K,1} & \cdots & h_{K,N} \end{bmatrix} = \mathbf{H}$$

Exemple: N=4 génomes

1.3 Applications NMF

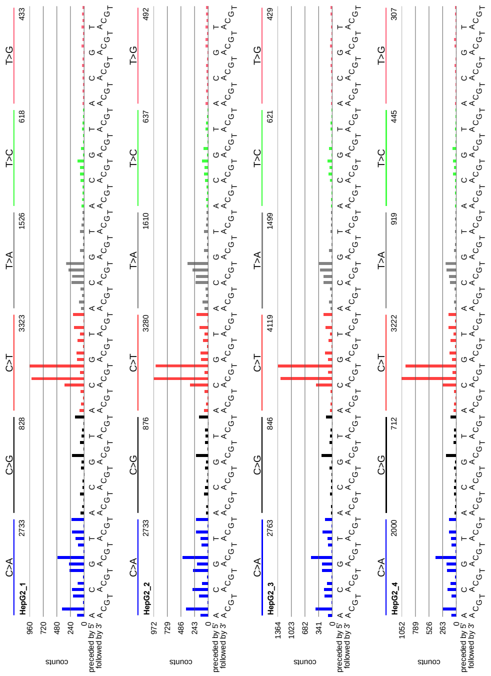
Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles [Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012](#) [Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013](#)

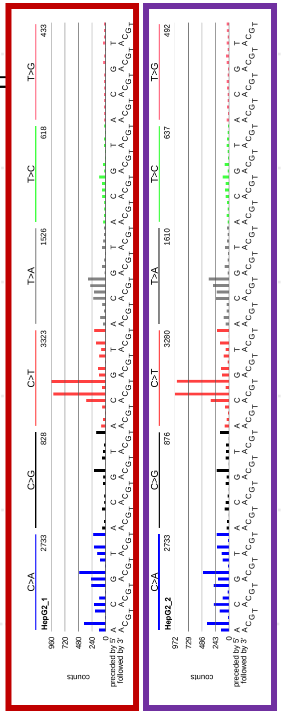
$$\mathbf{V} = \begin{bmatrix} v_{1,1} & \cdots & v_{1,N} \\ \vdots & \ddots & \vdots \\ v_{96,1} & \cdots & v_{96,N} \end{bmatrix}$$

<u>Exemple</u>: N=4 génomes



$\approx$

$\mathbf{W} =$



Signature 1 Signature 2

$$\begin{bmatrix} h_{1,1} & \cdots & h_{1,N} \\ h_{K,1} & \cdots & h_{K,N} \end{bmatrix} = \mathbf{H}$$

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012

Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013

Alexandrov, Nik-Zainal, Wedge, et al.. Nature 2013

-> 4,9M de mutations, 7k exomes de cancers, 30 signatures

Alexandrov, Kim, Haradvala, et al. Nature 2020

-> 79,8M de mutations, 19k exomes de cancers, 4k genomes de cancers, 65 signatures SBS

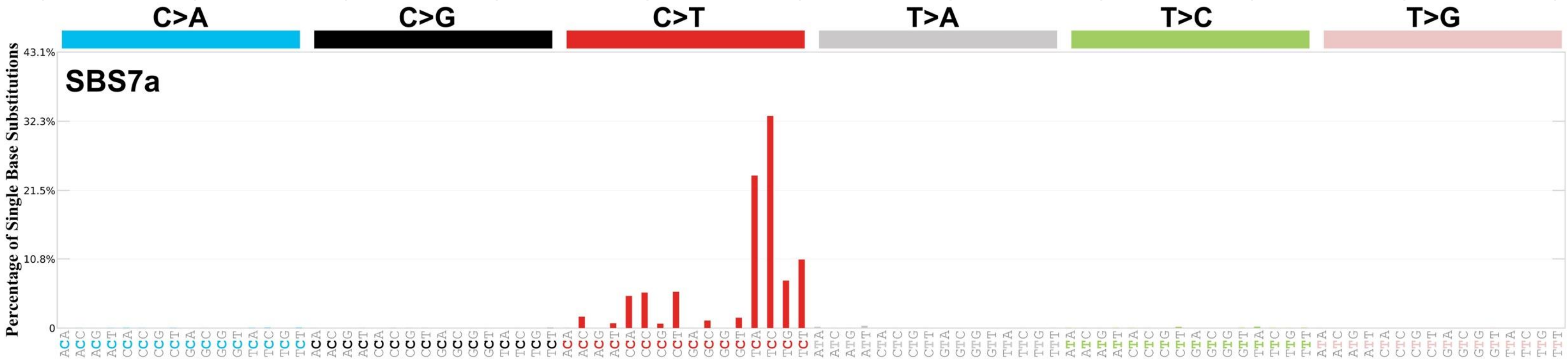
https://cancer.sanger.ac.uk/signatures/sbs/

1.3 Applications NMF

Différentes applications

Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles https://cancer.sanger.ac.uk/signatures/sbs/



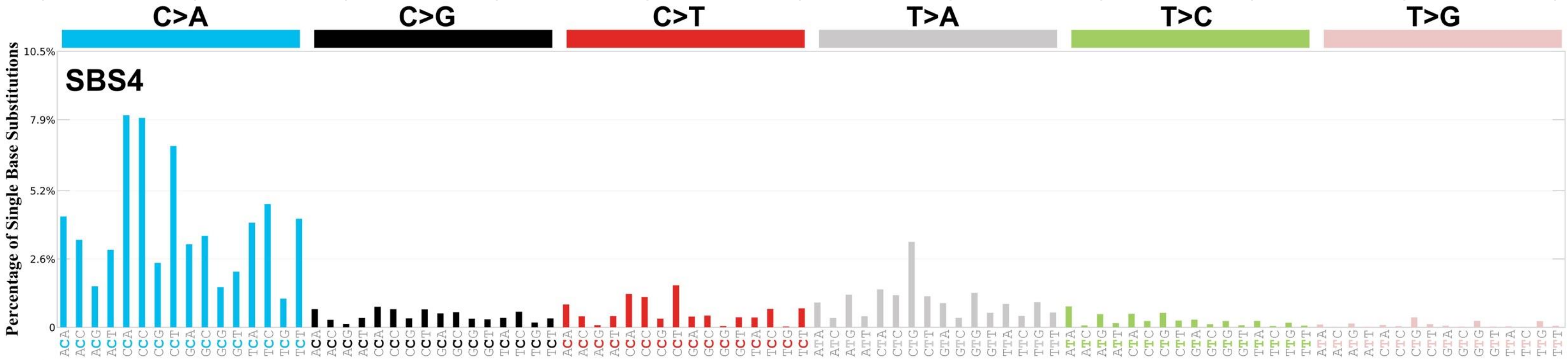
Exposition aux UV

1.3 Applications NMF

Différentes applications

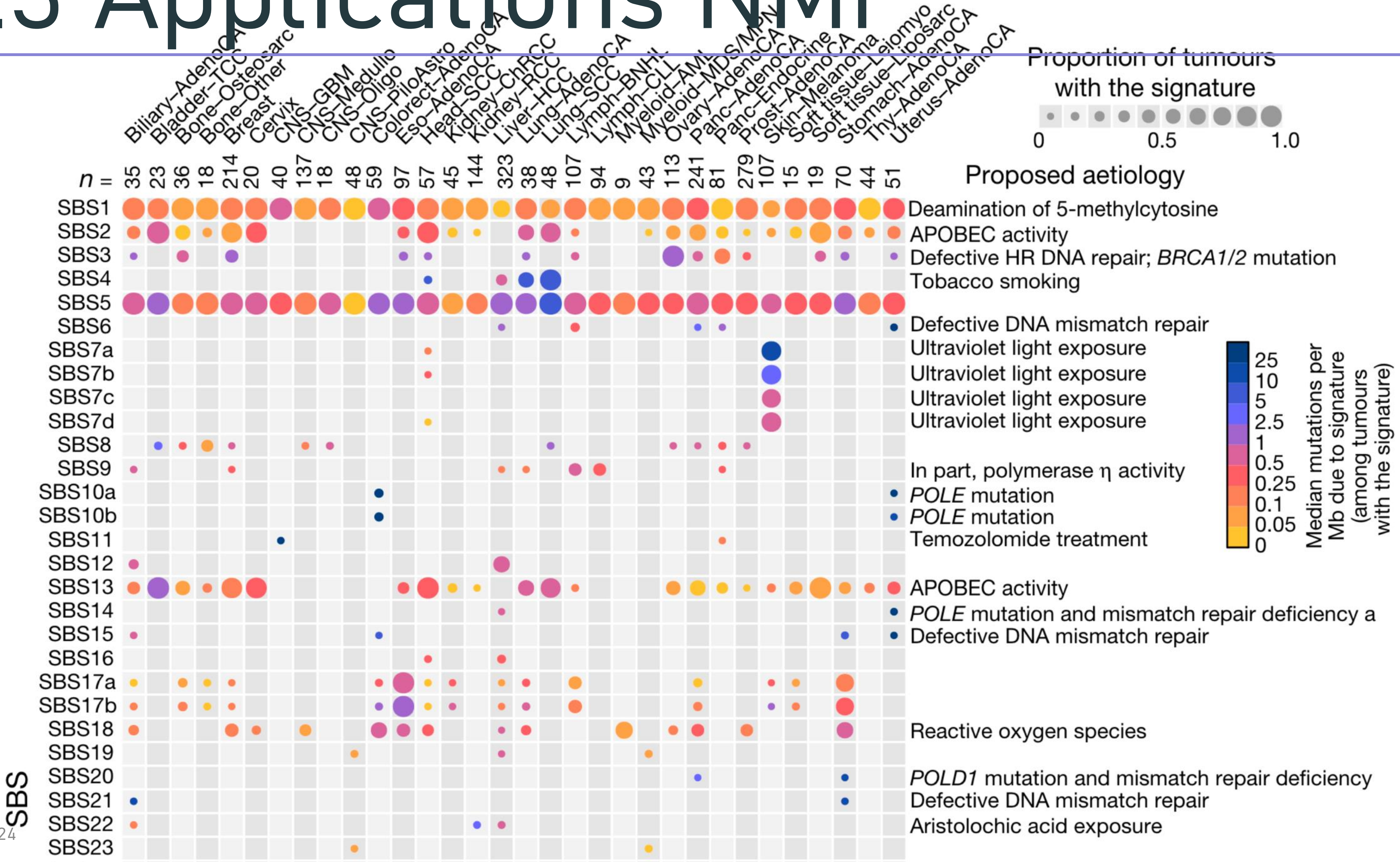
Comment determiner le nombre optimal de sources K ?!

2. Signatures mutationnelles https://cancer.sanger.ac.uk/signatures/sbs/



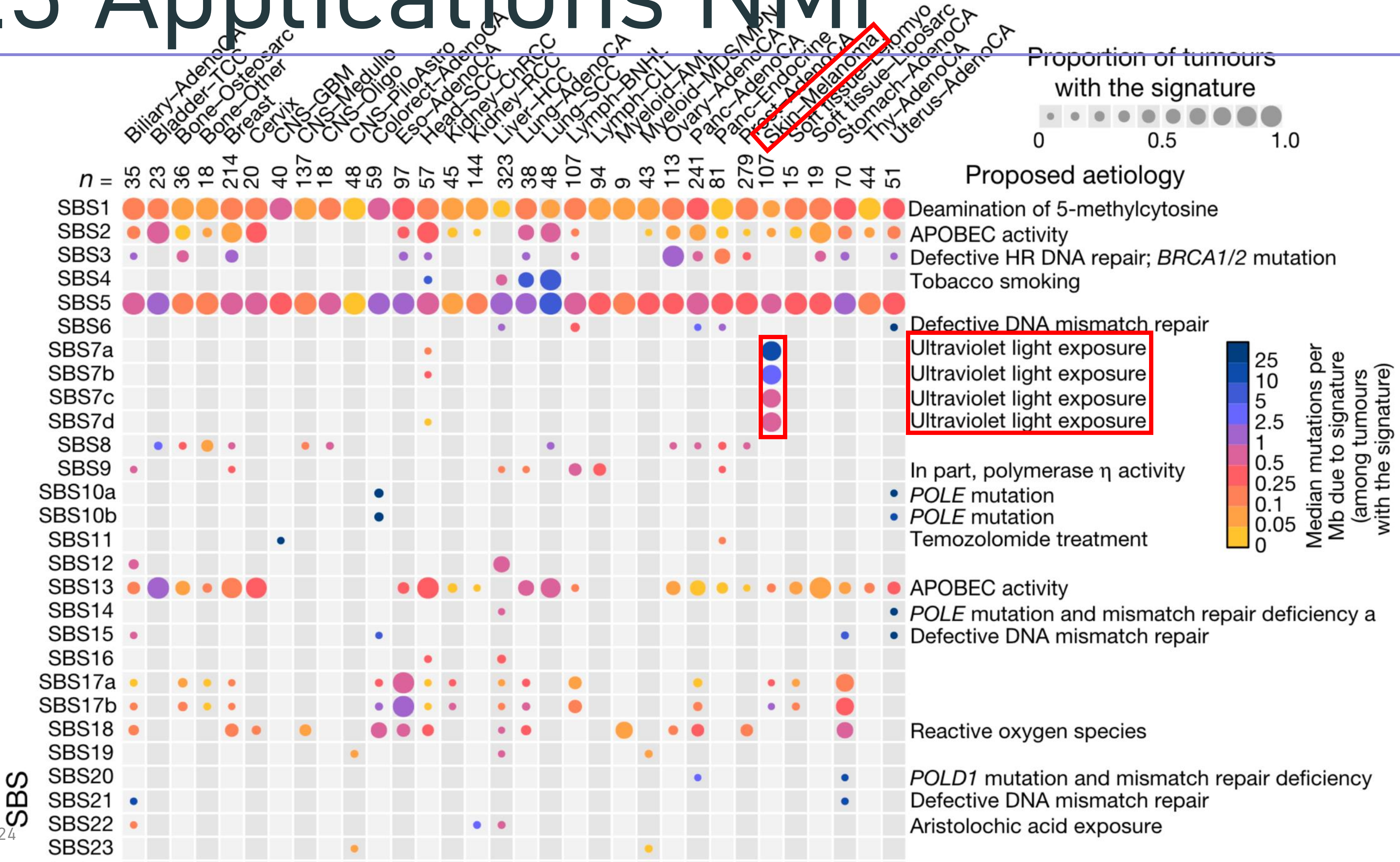
Exposition au tabac

1.3 Applications NMF



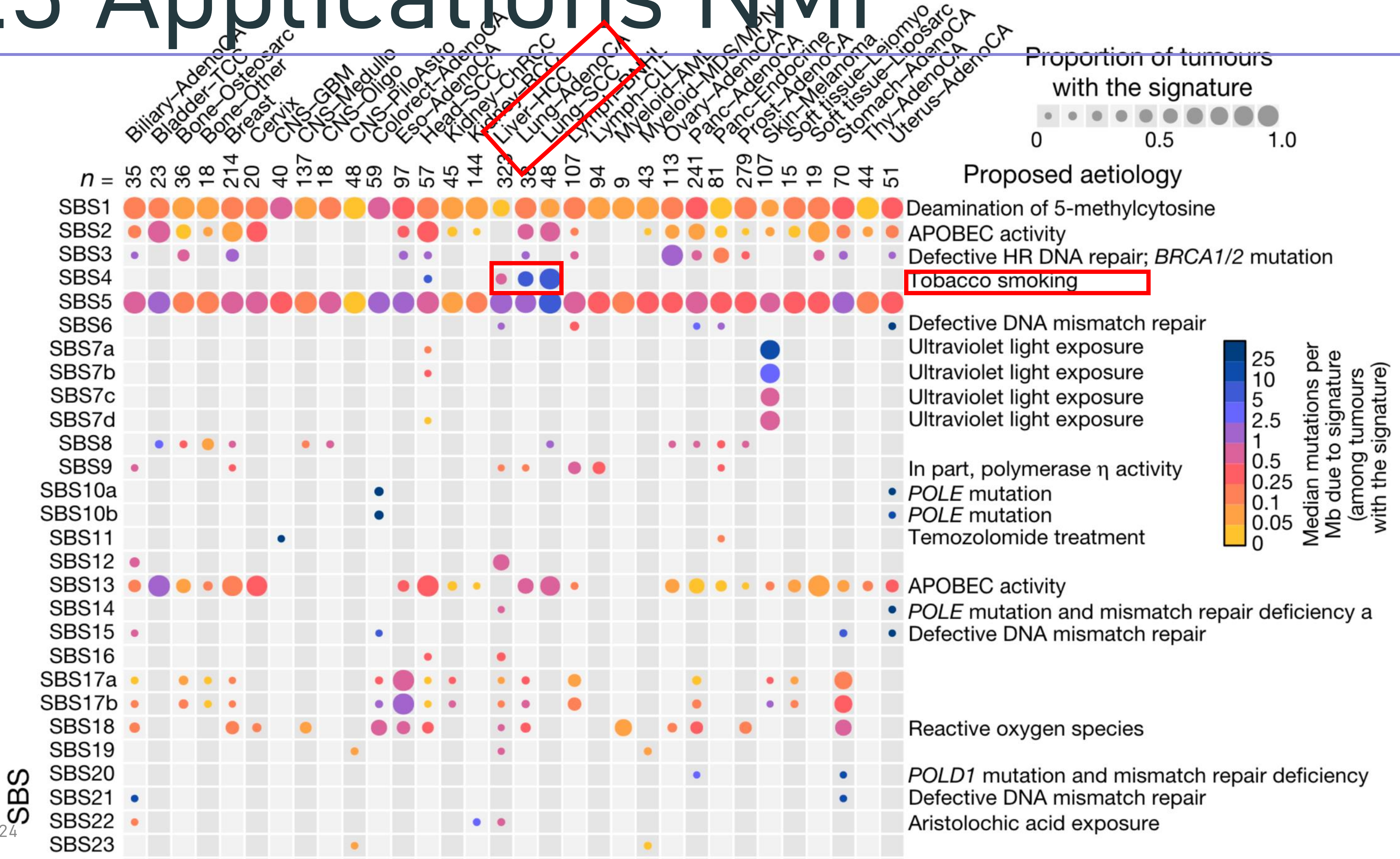
Alexandrov, Kim, Haradvala, et al. Nature 2020

1.3 Applications NMF



Alexandrov, Kim, Haradvala, et al. Nature 2020

1.3 Applications NMF



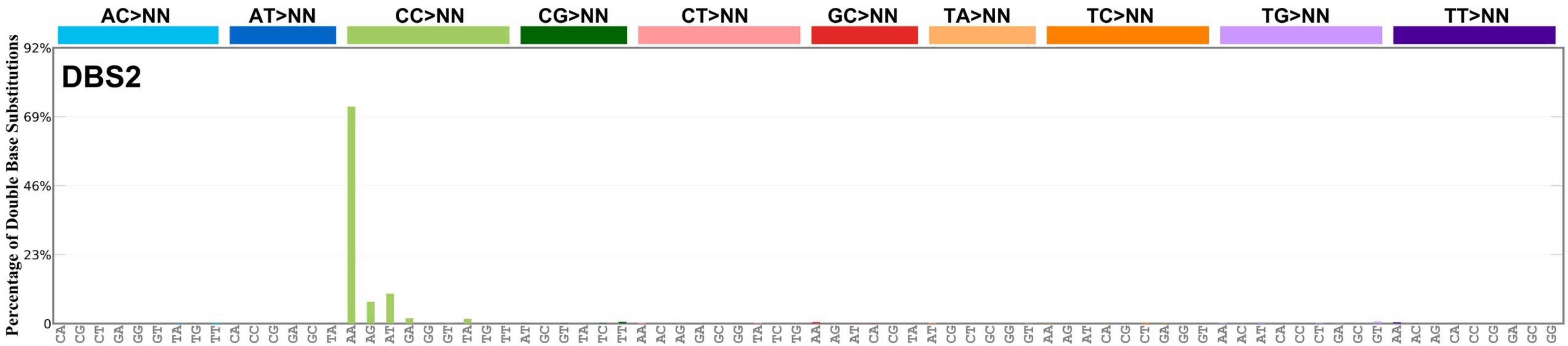
Alexandrov, Kim, Haradvala, et al. Nature 2020

1.3 Applications NMF

2. Signatures mutationnelles

Extensions à d’autres “contextes” ou altérations

-> 5 nucleotide contexts
-> doublet-base substitutions



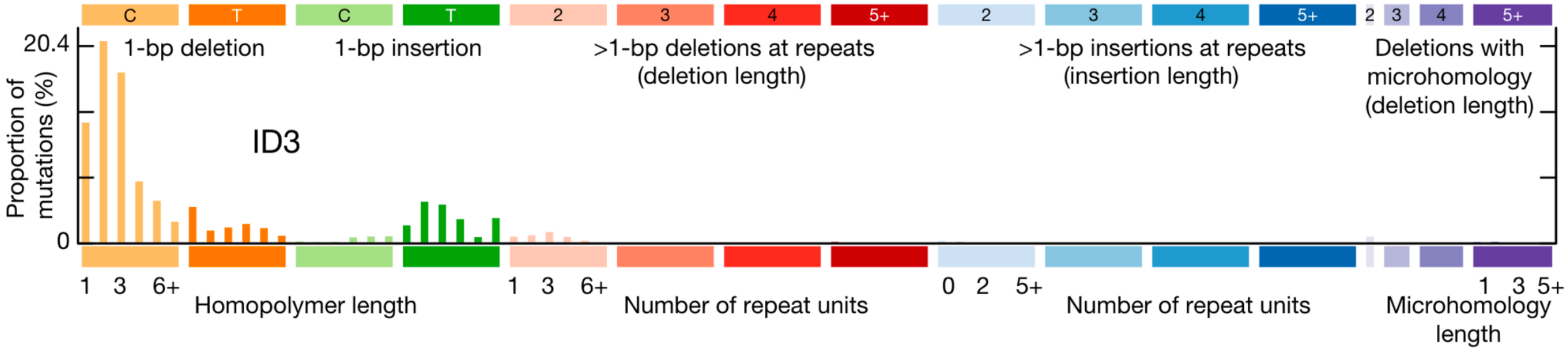
Alexandrov, Kim, Haradvala, et al. Nature 2020

1.3 Applications NMF

2. Signatures mutationnelles

Extensions à d'autres "contextes" ou altérations

-> 5 nucleotide contexts
-> doublet-base substitutions
-> indels



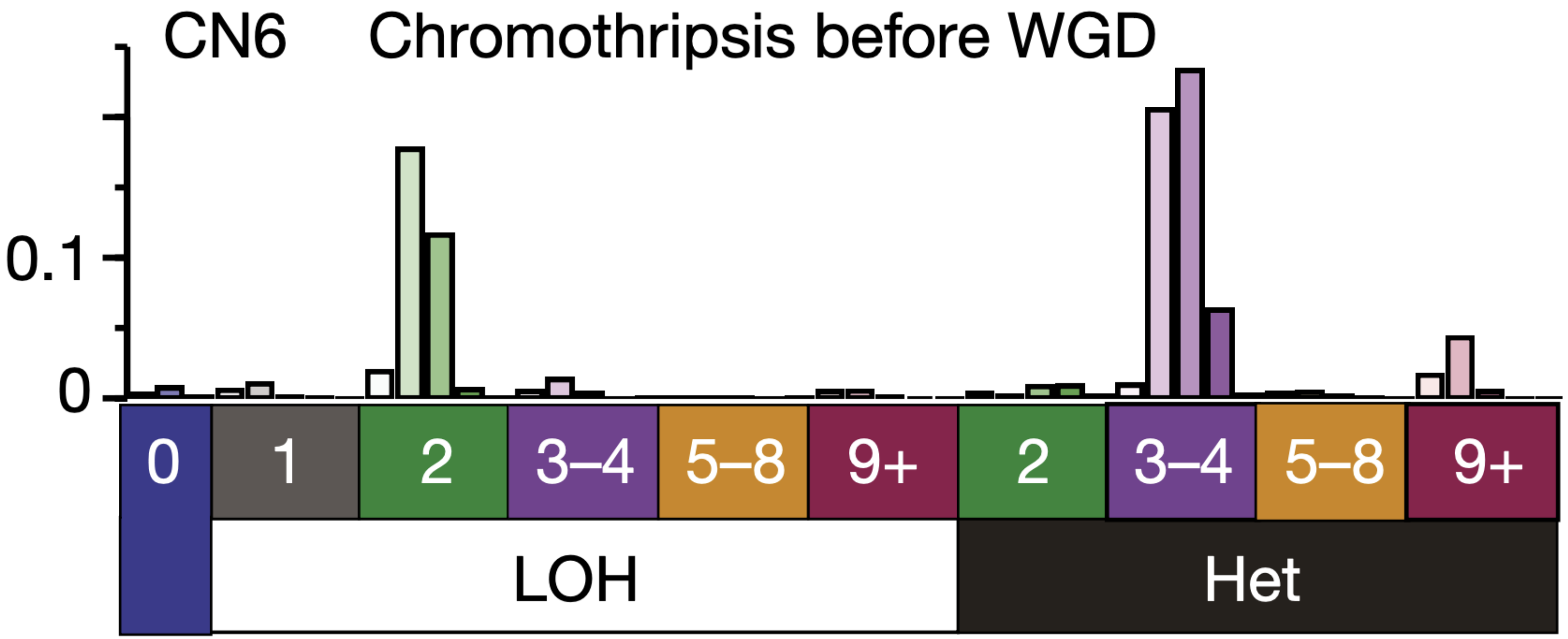
Alexandrov, Kim, Haradvala, et al. Nature 2020

1.3 Applications NMF

2. Signatures mutationnelles

Extensions à d'autres "contextes" ou altérations

-> 5 nucleotide contexts
-> doublet-base substitut
-> indels
-> CNAs



Steele, et al. Nature 2022

Sommaire

2.1 Concepts

Notations

1. $\mathrm{T_D}, \mathrm{T_C}$ variable aléatoires positives, temps à l'évènement et temps à la censure respectivement;
2. Δ variable aléatoire binaire indiquant l'occurrence de l'évènement;
3. $\mathrm{T} = \min(\mathrm{T_D}, \mathrm{T_C})$ variable aléatoire observée;
4. $\mathbf{Z}$ vecteur aléatoire des covariables;

2.1 Concepts

Notations

1. $\mathrm{T_D}, \mathrm{T_C}$ variable aléatoires positives, temps à l'évènement et temps à la censure respectivement;
2. Δ variable aléatoire binaire indiquant l'occurrence de l'évènement;
3. $\mathrm{T} = \min(\mathrm{T_D}, \mathrm{T_C})$ variable aléatoire observée;
4. $\mathbf{Z}$ vecteur aléatoire des covariables;

Définition fonction survie

La **fonction de survie** est donnée par

$$S : \begin{array}{ccc} \mathbb{R}_+ & \to & [0,1] \\ t & \mapsto & \mathbb{P}(\mathrm{T_D} \geq t), \end{array}$$

2.1 Concepts

Définition taux de risque instantané

La **taux de risque (instantané)** est donné par

$$\lambda : \begin{array}{ccc} \mathbb{R}_+ & \to & \mathbb{R}_+ \\ t & \mapsto & \lim_{h\to 0} \mathbb{P}(\mathrm{T_D} \leq t + h | \mathrm{T_D} \geq t) \end{array}$$

Soit: f_D (resp F_D) densité (resp f.r.) de $\mathrm{T_D}$. Alors

$$\lambda(t) = \frac{f_D(t)}{1 - F_D(t)}$$

2.1 Concepts

Définition taux de risque instantané

La **taux de risque (instantané)** est donné par

$$\lambda : \begin{array}{ccc} \mathbb{R}_+ & \to & \mathbb{R}_+ \\ t & \mapsto & \lim_{h \to 0} \mathbb{P}(\mathrm{T_D} \leq t + h | \mathrm{T_D} \geq t) \end{array}$$

Soit: f_D (resp F_D) densité (resp f.r.) de $\mathrm{T_D}$. Alors

$$\begin{aligned} \lambda(t) &= \frac{f_D(t)}{1 - F_D(t)} \\ &= \frac{-S'(t)}{S(t)} \end{aligned}$$

2.1 Concepts

Définition taux de risque instantané

La **taux de risque (instantané)** est donné par

$$\lambda : \begin{array}{ccc} \mathbb{R}_+ & \to & \mathbb{R}_+ \\ t & \mapsto & \lim_{h\to 0} \mathbb{P}(\mathrm{T_D} \leq t + h | \mathrm{T_D} \geq t) \end{array}$$

Soit: f_D (resp F_D) densité (resp f.r.) de $\mathrm{T_D}$. Alors

$$\lambda(t) = \frac{f_D(t)}{1 - F_D(t)} = \frac{-S'(t)}{S(t)}$$

d'où $$\int_0^t \lambda(s)ds = -\log(S(t)) \quad (\text{car } S(0)=1)$$

2.2 Modèle de Cox

Le modèle

Soient $\delta_{1:n}, \mathbf{z_{1:n}}$ n-échantillon de $\Delta_{1:n}, \mathbf{Z_{1:n}}$. Le modèle de Cox modélise $\mathrm{T_{D,1:n}}$, à $\mathbf{z_{1:n}}$ fixés, via le taux de risque selon la relation

$$\lambda_i(t) = \lambda_0(t) e^{\mathbf{z_i}^\top \beta}$$

2.2 Modèle de Cox

Le modèle

Soient $\delta_{1:n}, \mathbf{z_{1:n}}$ n-échantillon de $\Delta_{1:n}, \mathbf{Z_{1:n}}$. Le modèle de Cox modélise $\mathrm{T_{D,1:n}}$, à $\mathbf{z_{1:n}}$ fixés, via le taux de risque selon la relation

$$\lambda_i(t) = \lambda_0(t) e^{\mathbf{z_i}^\top \beta}$$

Explications

-> **<u>Proportionnalité</u>** ou "risques proportionnels" ?

Supposons $\mathbf{z} = 1$ pour le groupe traité et $\mathbf{z} = 0$ pour le groupe de contrôle. Alors,

$$\forall t \geq 0, \qquad \frac{\lambda(t,1)}{\lambda(t,0)} = \frac{\lambda_0(t) e^{\beta}}{\lambda_0(t)} = e^{\beta}$$

2.2 Modèle de Cox

Le modèle

Soient $\delta_{1:n}, \mathbf{z_{1:n}}$ n-échantillon de $\Delta_{1:n}, \mathbf{Z_{1:n}}$. Le modèle de Cox modélise $\mathrm{T_{D,1:n}}$, à $\mathbf{z_{1:n}}$ fixés, via le taux de risque selon la relation

$$\lambda_i(t) = \lambda_0(t) e^{\mathbf{z_i}^\top \beta}$$

Explications

-> **Proportionnalité** ou "risques proportionnels" ?
Supposons $\mathbf{z} = 1$ pour le groupe traité et $\mathbf{z} = 0$ pour le groupe de contrôle. Alors,

$$\forall t \geq 0, \qquad \frac{\lambda(t,1)}{\lambda(t,0)} = \frac{\lambda_0(t) e^{\beta}}{\lambda_0(t)} = e^{\beta}$$

-> **Linéarité** $\log \lambda$ est une combinaison linéaire des covariables $\mathbf{z} = (z^1, \cdots, z^p)$

2.2 Modèle de Cox

Entrainement par maximum de vraisemblance

- $\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 0) = \mathbb{P}(\mathrm{T_{D,i}} \geq t_i) = \mathcal{S}_i(t_i)$
- $\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 1) = \mathbb{P}(\mathrm{T_{D,i}} = t_i) = \lambda_i(t_i)\mathcal{S}_i(t_i)$

$$\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$$

2.2 Modèle de Cox

Entrainement par maximum de vraisemblance

- $\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 0) = \mathbb{P}(\mathrm{T_{D,i}} \geq t_i) = \mathcal{S}_i(t_i)$
- $\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 1) = \mathbb{P}(\mathrm{T_{D,i}} = t_i) = \lambda_i(t_i)\mathcal{S}_i(t_i)$

$$\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$$

$$\mathcal{L}(\beta; \mathbf{z_{1:n}}, \mathbf{t_{1:n}}, \delta_{\mathbf{1:n}}) = \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S_i(t_i)$$

2.2 Modèle de Cox

Entrainement par maximum de vraisemblance

- $\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 0) = \mathbb{P}(\mathrm{T_{D,i}} \geq t_i) = \mathcal{S}_i(t_i)$
- $\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 1) = \mathbb{P}(\mathrm{T_{D,i}} = t_i) = \lambda_i(t_i)\mathcal{S}_i(t_i)$

$\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$

$$\mathcal{L}(\beta; \mathbf{z_{1:n}}, \mathbf{t_{1:n}}, \delta_{\mathbf{1:n}}) = \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S_i(t_i)$$

$$= \prod_{i=1}^{n} \left[\frac{\lambda_i(t_i)}{\sum_{j \in R(t_i)} \lambda_j(t_j)} \right]^{\delta_i} [\sum_{j \in R(t_i)} \lambda_j(t_i)]^{\delta_i} S_i(t_i)$$

2.2 Modèle de Cox

Entrainement par maximum de vraisemblance

- $\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 0) = \mathbb{P}(\mathrm{T_{D,i}} \geq t_i) = \mathcal{S}_i(t_i)$
- $\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 1) = \mathbb{P}(\mathrm{T_{D,i}} = t_i) = \lambda_i(t_i)\mathcal{S}_i(t_i)$

$$\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$$

$$\mathcal{L}(\beta; \mathbf{z_{1:n}}, \mathbf{t_{1:n}}, \delta_{\mathbf{1:n}}) = \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S_i(t_i)$$

$$= \prod_{i=1}^{n} \left[\frac{\lambda_i(t_i)}{\sum_{j \in \boxed{R(t_i)}} \lambda_j(t_j)} \right]^{\delta_i} [\sum_{j \in \boxed{R(t_i)}} \lambda_j(t_i)]^{\delta_i} S_i(t_i)$$

Patients encore à risque au temps i

2.2 Modèle de Cox

Entrainement par maximum de vraisemblance

$$\left.\begin{array}{l} \bullet\ \mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 0) = \mathbb{P}(\mathrm{T_{D,i}} \geq t_i) = \mathcal{S}_i(t_i) \\ \bullet\ \mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 1) = \mathbb{P}(\mathrm{T_{D,i}} = t_i) = \lambda_i(t_i)\mathcal{S}_i(t_i) \end{array}\right\} \mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$$

$$\begin{aligned}
\mathcal{L}(\beta; \mathbf{z_{1:n}}, \mathbf{t_{1:n}}, \delta_{\mathbf{1:n}}) &= \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S_i(t_i) \\
&= \prod_{i=1}^{n} \left[\frac{\lambda_i(t_i)}{\sum_{j \in R(t_i)} \lambda_j(t_j)} \right]^{\delta_i} [\sum_{j \in R(t_i)} \lambda_j(t_i)]^{\delta_i} S_i(t_i) \\
&= \prod_{i=1}^{n} \left[\frac{\lambda_0(t_i) e^{z_i^\top \beta}}{\sum_{j \in R(t_i)} \lambda_0(t_i) e^{z_j^\top \beta}} \right]^{\delta_i} [\sum_{j \in R(t_i)} \lambda_j(t_j)]^{\delta_i} S_i(t_i)
\end{aligned}$$

2.2 Modèle de Cox

Entrainement par maximum de vraisemblance

- $\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 0) = \mathbb{P}(\mathrm{T_{D,i}} \geq t_i) = \mathcal{S}_i(t_i)$
- $\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 1) = \mathbb{P}(\mathrm{T_{D,i}} = t_i) = \lambda_i(t_i)\mathcal{S}_i(t_i)$

$$\mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$$

$$\begin{aligned}
\mathcal{L}(\beta; \mathbf{z_{1:n}}, \mathbf{t_{1:n}}, \delta_{\mathbf{1:n}}) &= \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S_i(t_i) \\
&= \prod_{i=1}^{n} \left[\frac{\lambda_i(t_i)}{\sum_{j \in R(t_i)} \lambda_j(t_j)} \right]^{\delta_i} [\sum_{j \in R(t_i)} \lambda_j(t_i)]^{\delta_i} S_i(t_i) \\
&= \boxed{\prod_{i=1}^{n} \left[\frac{\lambda_0(t_i) e^{z_i^\top \beta}}{\sum_{j \in R(t_i)} \lambda_0(t_i) e^{z_j^\top \beta}} \right]^{\delta_i}} [\sum_{j \in R(t_i)} \lambda_j(t_j)]^{\delta_i} S_i(t_i)
\end{aligned}$$

Vraisemblance partielle

2.2 Modèle de Cox

Entrainement par maximum de vraisemblance

$$\left.\begin{array}{l} \bullet\ \mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 0) = \mathbb{P}(\mathrm{T_{D,i}} \geq t_i) = \mathcal{S}_i(t_i) \\ \bullet\ \mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = 1) = \mathbb{P}(\mathrm{T_{D,i}} = t_i) = \lambda_i(t_i)\mathcal{S}_i(t_i) \end{array}\right\} \mathbb{P}(\mathrm{T_i} = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$$

$$\begin{aligned} \mathcal{L}(\beta; \mathbf{z_{1:n}}, \mathbf{t_{1:n}}, \delta_{\mathbf{1:n}}) &= \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S_i(t_i) \\ &= \prod_{i=1}^{n} \left[\frac{\lambda_i(t_i)}{\sum_{j \in R(t_i)} \lambda_j(t_j)} \right]^{\delta_i} [\sum_{j \in R(t_i)} \lambda_j(t_i)]^{\delta_i} S_i(t_i) \\ &= \boxed{\prod_{i=1}^{n} \left[\frac{\cancel{\lambda_0(t_i)} e^{z_i^\top \beta}}{\sum_{j \in R(t_i)} \cancel{\lambda_0(t_i)} e^{z_j^\top \beta}} \right]^{\delta_i}} [\sum_{j \in R(t_i)} \lambda_j(t_j)]^{\delta_i} S_i(t_i) \end{aligned}$$

Vraisemblance partielle

2.2 Modèle de Cox

Exemple

individual	T_i	δ_i	Z_i
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

2.2 Modèle de Cox

Exemple

individual	T_i	δ_i	Z_i
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

j	ordered failure time T_i	$R(T_i)$	i_j	Likelihood contribution $\left[e^{\beta Z_i} / \sum_{j \in \mathcal{R}(T_i)} e^{\beta Z_j}\right]^{\delta_i}$
1	6	{1,2,3,4}	3	$e^{7\beta}/[e^{4\beta}+e^{5\beta}+e^{7\beta}+e^{3\beta}]$

2.2 Modèle de Cox

Exemple

individual	T_i	δ_i	Z_i
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

j	ordered failure time T_i	$R(T_i)$	i_j	Likelihood contribution $\left[e^{\beta Z_i} / \sum_{j \in \mathcal{R}(T_i)} e^{\beta Z_j}\right]^{\delta_i}$
1	6	{1,2,3,4}	3	$e^{7\beta}/[e^{4\beta}+e^{5\beta}+e^{7\beta}+e^{3\beta}]$
2	8	{1,2,4}	2	1

2.2 Modèle de Cox

Exemple

individual	T_i	δ_i	Z_i
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

j	ordered failure time T_i	$R(T_i)$	i_j	Likelihood contribution $\left[e^{\beta Z_i} / \sum_{j \in \mathcal{R}(T_i)} e^{\beta Z_j}\right]^{\delta_i}$
1	6	{1,2,3,4}	3	$e^{7\beta}/[e^{4\beta}+e^{5\beta}+e^{7\beta}+e^{3\beta}]$
2	8	{1,2,4}	2	1
3	9	{1,4}	1	$e^{4\beta}/[e^{4\beta}+e^{3\beta}]$

2.2 Modèle de Cox

Exemple

individual	T_i	δ_i	Z_i
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

j	ordered failure time T_i	$R(T_i)$	i_j	Likelihood contribution $\left[e^{\beta Z_i} / \sum_{j \in \mathcal{R}(T_i)} e^{\beta Z_j}\right]^{\delta_i}$
1	6	{1,2,3,4}	3	$e^{7\beta}/[e^{4\beta} + e^{5\beta} + e^{7\beta} + e^{3\beta}]$
2	8	{1,2,4}	2	1
3	9	{1,4}	1	$e^{4\beta}/[e^{4\beta} + e^{3\beta}]$
4	10	{4}	4	$e^{3\beta}/e^{3\beta} = 1$