



# 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 1

# Sommaire

---

1. Factorisation matricielle non négative
  1. Factorisations matricielles
  2. Factorisation non négative
  3. Applications NMF
2. Analyse de survie
  1. Concepts
  2. Modèle de Cox

# 1.1 Factorisations matricielles

---

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

# 1.1 Factorisations matricielles

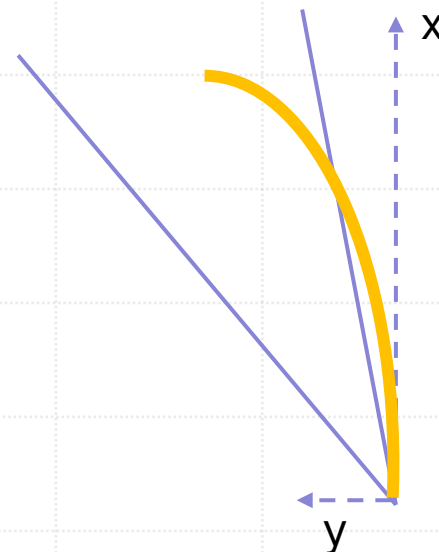
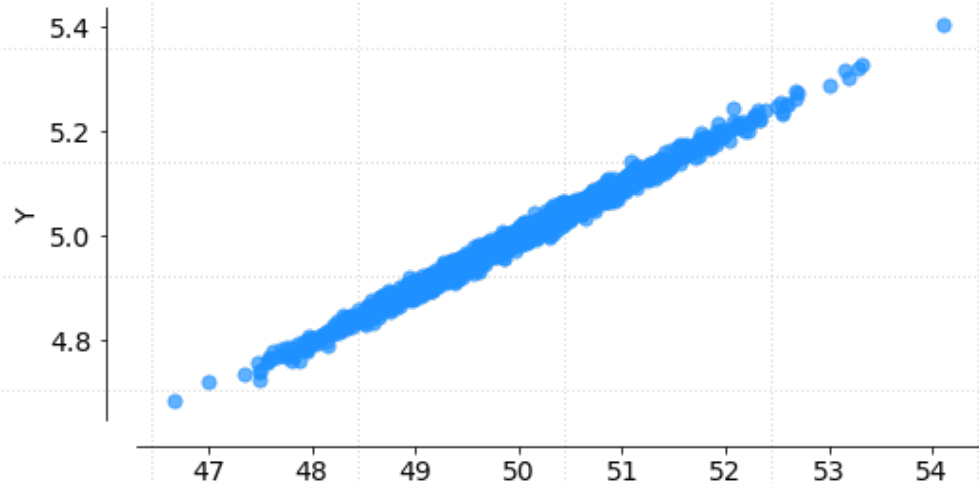
Notations  $N \in \mathbb{N}^*$  : nombre d'observations (=individus, échantillons)  
 $P \in \mathbb{N}^*$  : nombre de variables (=covariables, prédicteurs, features)  
 $V \in \mathbb{R}^{P \times N}$  : ensemble d'observations (=dataset)

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

# 1.1 Factorisations matricielles

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

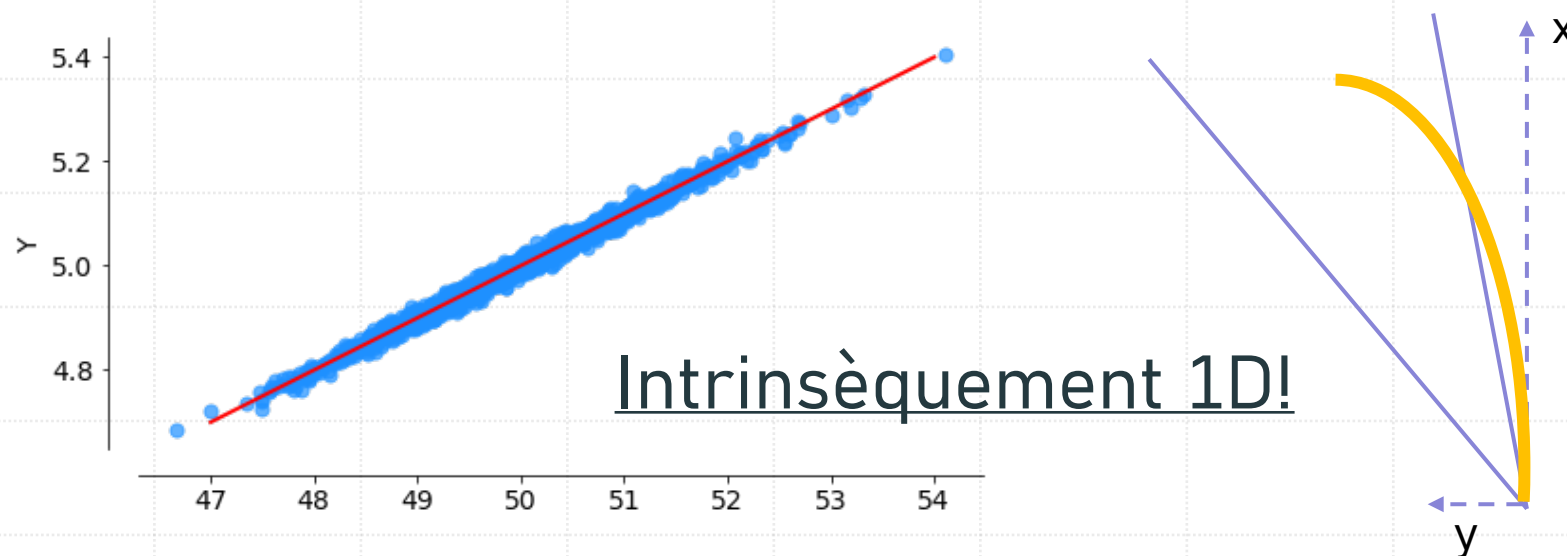
Idée Les observations sont générées par un petit nombre processus  $K \ll P$   
Les observations sont bien approchées par une représentation en basse dimension



# 1.1 Factorisations matricielles

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

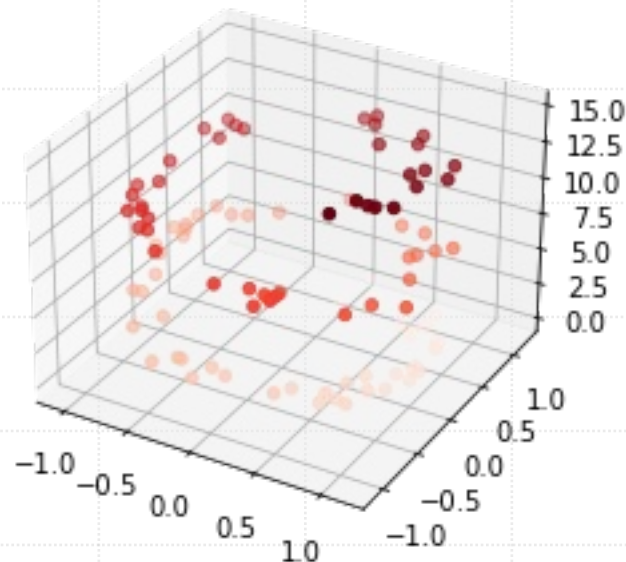
Idée Les observations sont générées par un petit nombre processus  $K \ll P$   
Les observations sont bien approchées par une représentation en basse dimension



# 1.1 Factorisations matricielles

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

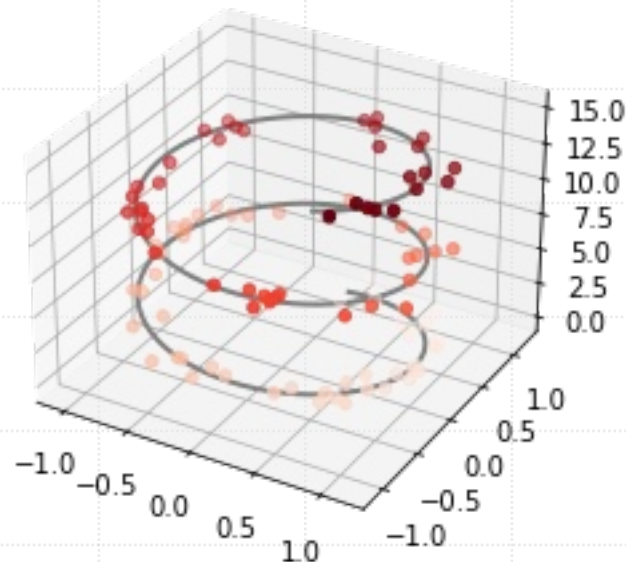
Idée Les observations sont générées par un petit nombre processus  $K \ll P$   
Les observations sont bien approchées par une représentation en basse dimension



# 1.1 Factorisations matricielles

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

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



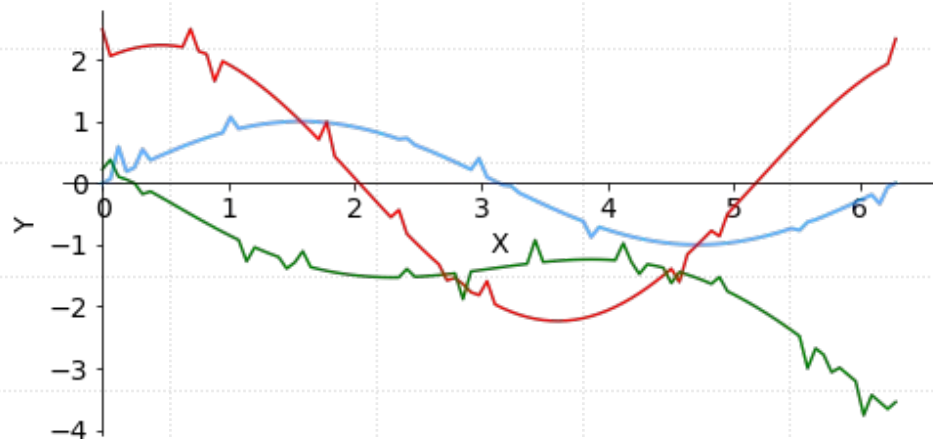
Intrinsèquement 1D!



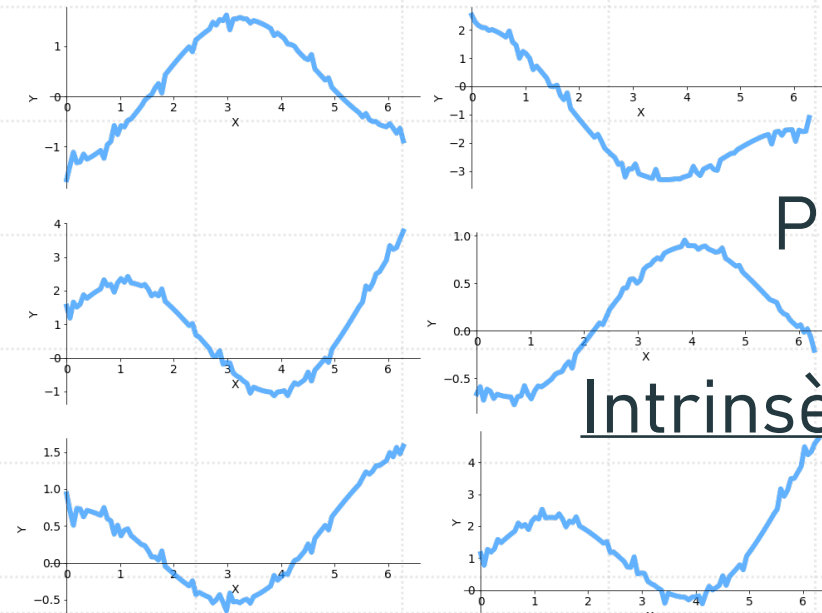
# 1.1 Factorisations matricielles

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

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



Sources



$P = 100$  points

Intrinsèquement 3D!

# 1.1 Factorisations matricielles

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

Idée Les observations sont générées par un petit nombre processus  $K \ll P$   
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

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

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

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

# 1.1 Factorisations matricielles

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

Idée Les observations sont générées par un petit nombre processus  $K \ll P$   
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  $N \in \mathbb{N}^*$  : nombre d'observations (=individus, échantillons)  
 $P \in \mathbb{N}^*$  : nombre de variables (=covariables, prédicteurs, features)  
 $V \in \mathbb{R}^{P \times N}$  : ensemble d'observations (=dataset)

Idée Les observations sont générées par un petit nombre processus  $K \ll P$   
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      Poids source 1 ou coord axe 1

# 1.1 Factorisations matricielles

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

Idée Les observations sont générées par un petit nombre processus  $K \ll P$   
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

Poids source 1 ou coord axe 1


Poids source K ou coord axe K

# 1.1 Factorisations matricielles

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

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

$$\mathbf{v} = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \cdots + v_{P-1} \mathbf{e}_{P-1} + v_P \mathbf{e}_P \quad \text{Data dimension} = P$$


$$\mathbf{v}_{\text{approx}} = h_1 \mathbf{w}_1 + \cdots + h_K \mathbf{w}_K$$


Approx dimension =  $K \ll P$

# 1.1 Factorisations matricielles

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

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

$$\mathbf{v} = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \cdots + v_{P-1} \mathbf{e}_{P-1} + v_P \mathbf{e}_P \quad \text{Data dimension} = P$$


$$\mathbf{v}_{\text{approx}} = h_1 \mathbf{w}_1 + \cdots + h_K \mathbf{w}_K$$

Approx dimension =  $K \ll P$

Clé: distance  $d(v, v_{\text{approx}})$



# 1.1 Factorisations matricielles

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

Formulation mathématique Factorisation matricielle

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

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

# 1.1 Factorisations matricielles

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

Formulation mathématique Factorisation matricielle

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

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

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

# 1.1 Factorisations matricielles

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

## Formulation mathématique Factorisation matricielle

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

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

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

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

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

# 1.1 Factorisations matricielles

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

## Formulation mathématique Factorisation matricielle

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

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

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

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

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

$$D(\mathbf{x} || \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$$

Divergence

# 1.1 Factorisations matricielles

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

Formulation mathématique Factorisation matricielle

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

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

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

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

Divergence

# 1.1 Factorisations matricielles

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

Formulation mathématique Factorisation matricielle

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

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

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

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

Divergence

# 1.1 Factorisations matricielles

---

## Différentes formes de factorisation

### 1. Analyse en composante principales (PCA)

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

# 1.1 Factorisations matricielles

## Différentes formes de factorisation

### 1. Analyse en composante principales (PCA)

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



# 1.1 Factorisations matricielles

## Différentes formes de factorisation

### 1. Analyse en composante principales (PCA)

- $D(\mathbf{x}||\mathbf{y}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2$  -> « facilement » résolu par SVD (Eckart-Young, 1936)

- $\mathcal{W} = \mathbb{R}^{P \times K}$ ,  $\mathcal{H} = \mathbb{R}^{K \times N}$ .  $\mathbf{V} = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}$  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

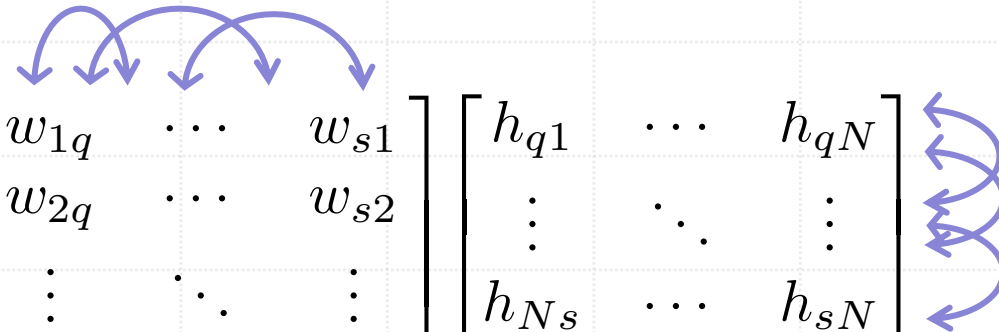
### 1. Analyse en composante principales (PCA)

- $D(\mathbf{x}||\mathbf{y}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2$  -> « facilement » résolu par SVD (Eckart-Young, 1936)

- $\mathcal{W} = \mathbb{R}^{P \times K}$ ,  $\mathcal{H} = \mathbb{R}^{K \times N}$ .  $\mathbf{V} = \mathbf{P} \mathbf{\Lambda} \mathbf{Q}$  alors  $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$   $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

NOTE 1: invariance d'échelle

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)

- $D(\mathbf{x}||\mathbf{y}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2$  -> « facilement » résolu par SVD (Eckart-Young, 1936)

- $\mathcal{W} = \mathbb{R}^{P \times K}$ ,  $\mathcal{H} = \mathbb{R}^{K \times N}$ .  $\mathbf{V} = \mathbf{P} \mathbf{\Lambda} \mathbf{Q}$  alors  $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$   $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

NOTE 1: invariance d'échelle

NOTE 2: invariance de rotation

Fixations des invariances:

- rotation, axes triés par valeur propre dec.  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_K \geq 0$
- échelle, axes "unitaires" i.e  $\mathbf{w}_k^\top \mathbf{w}_k = 1$ ,  $1 \leq k \leq K$

# 1.1 Factorisations matricielles

## Différentes formes de factorisation

### 1. Analyse en composante principales (PCA)

- $D(\mathbf{x}||\mathbf{y}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2$  -> « facilement » résolu par SVD (Eckart-Young, 1936)

- $\mathcal{W} = \mathbb{R}^{P \times K}$ ,  $\mathcal{H} = \mathbb{R}^{K \times N}$ .  $\mathbf{V} = \mathbf{P} \mathbf{\Lambda} \mathbf{Q}$  alors  $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$   $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

Propriétés PCA  $\mathbf{w}_q^\top \mathbf{w}_s = 0$ ,  $1 \leq s, q \leq K$ ,  $s \neq q$

# 1.1 Factorisations matricielles

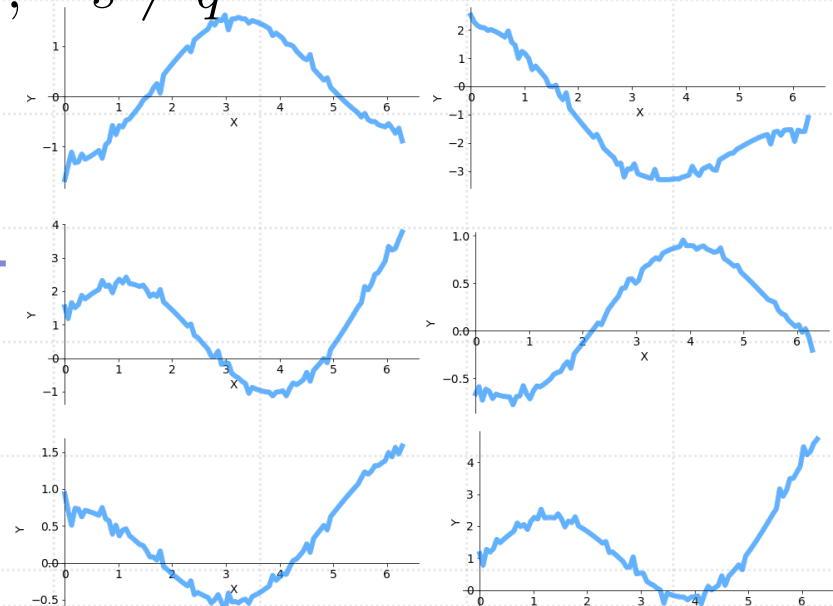
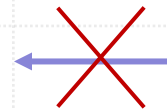
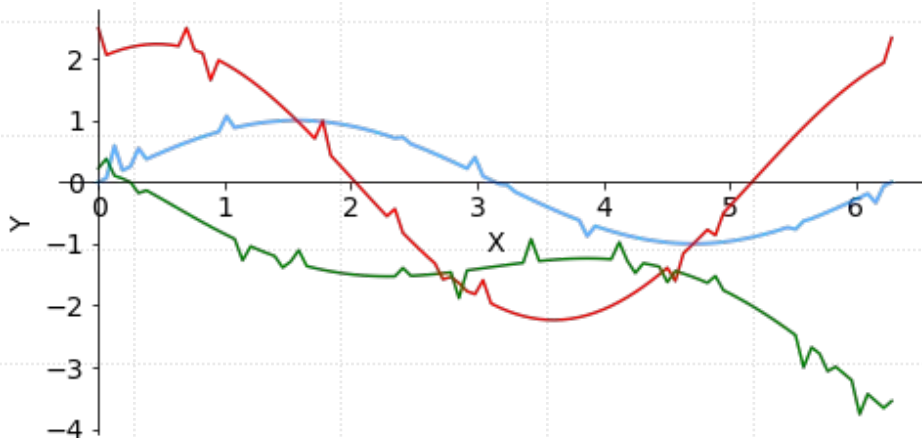
## Différentes formes de factorisation

### 1. Analyse en composante principales (PCA)

- $D(\mathbf{x}||\mathbf{y}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2$  -> « facilement » résolu par SVD (Eckart-Young, 1936)

- $\mathcal{W} = \mathbb{R}^{P \times K}$ ,  $\mathcal{H} = \mathbb{R}^{K \times N}$ .  $\mathbf{V} = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}$  alors  $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$   $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

Propriétés PCA  $\mathbf{w}_q^\top \mathbf{w}_s = 0$ ,  $1 \leq s, q \leq K$ ,  $s \neq q$



# 1.1 Factorisations matricielles

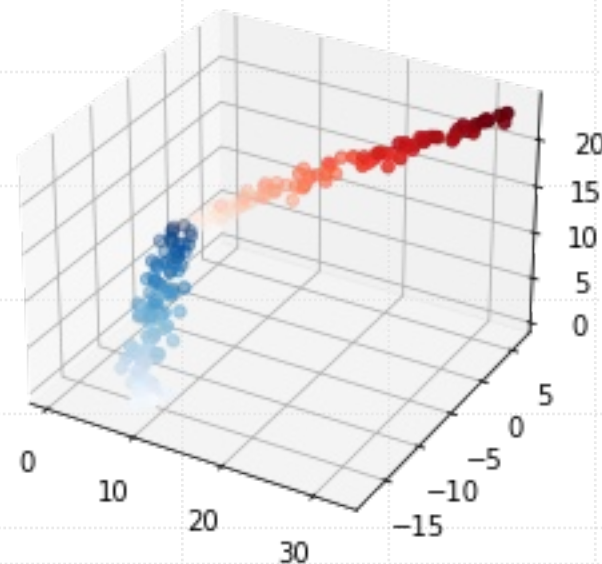
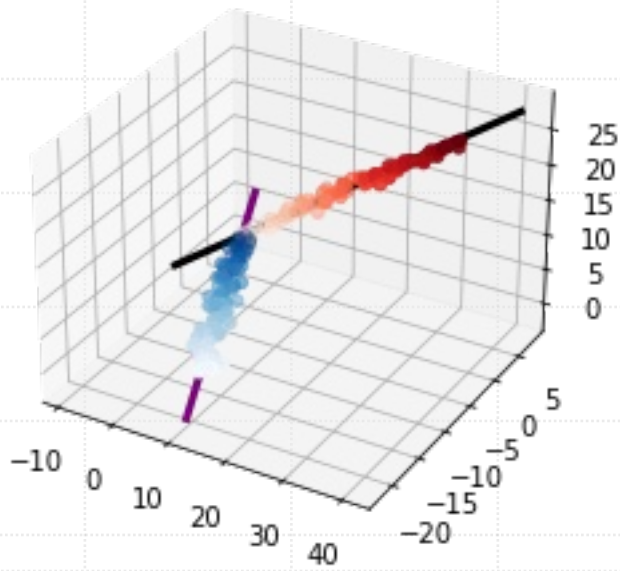
## Différentes formes de factorisation

### 1. Analyse en composante principales (PCA)

- $D(\mathbf{x}||\mathbf{y}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2$  -> « facilement » résolu par SVD (Eckart-Young, 1936)

- $\mathcal{W} = \mathbb{R}^{P \times K}$ ,  $\mathcal{H} = \mathbb{R}^{K \times N}$ .  $\mathbf{V} = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}$  alors  $\hat{\mathbf{W}} \propto \mathbf{P}_{:,1:K}$   $\hat{\mathbf{H}} \propto \mathbf{Q}_{1:K,:}$

Propriétés PCA  $\mathbf{w}_q^\top \mathbf{w}_s = 0$ ,  $1 \leq s, q \leq K$ ,  $s \neq q$



# 1.2 Factorisation non négative

## Différentes formes de factorisation

### 2. Factorisation matricielle non-négative (NMF)

Paatero et Tapper.  
Environmetrics. 1994.

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

# 1.2 Factorisation non négative

## Différentes formes de factorisation

### 2. Factorisation matricielle non-négative (NMF)

Paatero et Tapper.  
Environmetrics. 1994.

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

Lee et Seung. 2001.



# 1.2 Factorisation non négative

## Différentes formes de factorisation

### 2. Factorisation matricielle non-négative (NMF)

Paatero et Tapper.  
Environmetrics. 1994.

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

Lee et Seung. 2001.

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

Cichocki, Zdunek, et Amari.  
ICBSS. 2006

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

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

Févotte et Idier.  
arXiv. 2011

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

# 1.2 Factorisation non négative

## Différentes formes de factorisation

### 2. Factorisation matricielle non-négative (NMF)

Févotte et Idier.  
arXiv. 2011

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

$$D_{\beta=0}(\mathbf{x}|\mathbf{y}) = D_{\text{Itakura-Saito}},$$

$$D_{\beta=1}(\mathbf{x}|\mathbf{y}) = D_{\text{Kullback-Leibler}},$$

$$D_{\beta=2}(\mathbf{x}|\mathbf{y}) = D_{\text{Euclidean}}$$

# 1.2 Factorisation non négative

## Différentes formes de factorisation

### 2. Factorisation matricielle non-négative (NMF)

Févotte et Idier.  
arXiv. 2011

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

$$D_{\beta=0}(\mathbf{x}|\mathbf{y}) = D_{\text{Itakura-Saito}},$$

$$D_{\beta=1}(\mathbf{x}|\mathbf{y}) = D_{\text{Kullback-Leibler}},$$

$$D_{\beta=2}(\mathbf{x}|\mathbf{y}) = D_{\text{Euclidean}}$$

## Algorithme de résolution

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

# 1.2 Factorisation non négative

## Différentes formes de factorisation

### 2. Factorisation matricielle non-négative (NMF)

Févotte et Idier.  
arXiv. 2011

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

$$D_{\beta=0}(\mathbf{x}|\mathbf{y}) = D_{\text{Itakura-Saito}},$$

$$D_{\beta=1}(\mathbf{x}|\mathbf{y}) = D_{\text{Kullback-Leibler}},$$

$$D_{\beta=2}(\mathbf{x}|\mathbf{y}) = D_{\text{Euclidean}}$$

## Algorithme de résolution

$C(\mathbf{W}, \mathbf{H}) := D_{\beta}(\mathbf{V}||\mathbf{WH})$  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

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

$$D_{\beta=0}(\mathbf{x}|\mathbf{y}) = D_{\text{Itakura-Saito}},$$

$$D_{\beta=1}(\mathbf{x}|\mathbf{y}) = D_{\text{Kullback-Leibler}},$$

$$D_{\beta=2}(\mathbf{x}|\mathbf{y}) = D_{\text{Euclidean}}$$

## Algorithme de résolution

$C(\mathbf{W}, \mathbf{H}) := D_{\beta}(\mathbf{V}||\mathbf{WH})$  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**

$\mathbf{H}^{t+1} = \operatorname{argmin}_{\mathbf{H} \geq 0} D_{\beta}(\mathbf{V}||\mathbf{W}^t \mathbf{H});$

$\mathbf{W}^{t+1} = \operatorname{argmin}_{\mathbf{W} \geq 0} D_{\beta}(\mathbf{V}||\mathbf{W} \mathbf{H}^{t+1});$

**end**

---

# 1.2 Factorisation non négative

## Différentes formes de factorisation

### 2. Factorisation matricielle non-négative (NMF)

Févotte et Idier.  
arXiv. 2011

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

$$D_{\beta=0}(\mathbf{x}|\mathbf{y}) = D_{\text{Itakura-Saito}},$$

$$D_{\beta=1}(\mathbf{x}|\mathbf{y}) = D_{\text{Kullback-Leibler}},$$

$$D_{\beta=2}(\mathbf{x}|\mathbf{y}) = D_{\text{Euclidean}}$$

## Algorithme de résolution

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

## Propriétés:

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

# 1.3 Applications NMF

---

## Différentes applications

Comment déterminer le nombre optimal de sources  $K$  ?!

# 1.3 Applications NMF

## Différentes applications

Comment déterminer 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**

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



# 1.3 Applications NMF

## Différentes applications

Comment déterminer 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 déterminer le nombre optimal de sources K ?!

1. Classification de tumeurs [Brunet, Tamayo, Golub, et Mesirov. PNAS, 2004.](#)

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

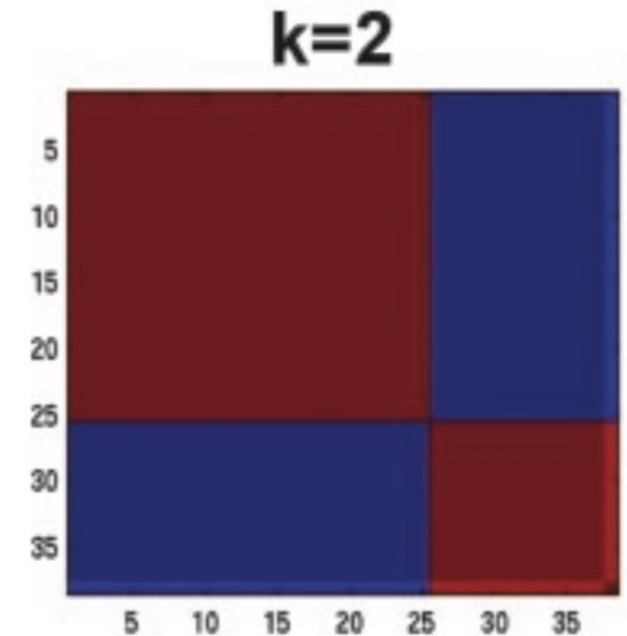
38 éch. moelle

-> 11 AML

-> 8 T-ALL

-> 19 B-ALL

999 genes (HU6800)



# 1.3 Applications NMF

## Différentes applications

Comment déterminer le nombre optimal de sources K ?!

1. Classification de tumeurs [Brunet, Tamayo, Golub, et Mesirov. PNAS, 2004.](#)

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

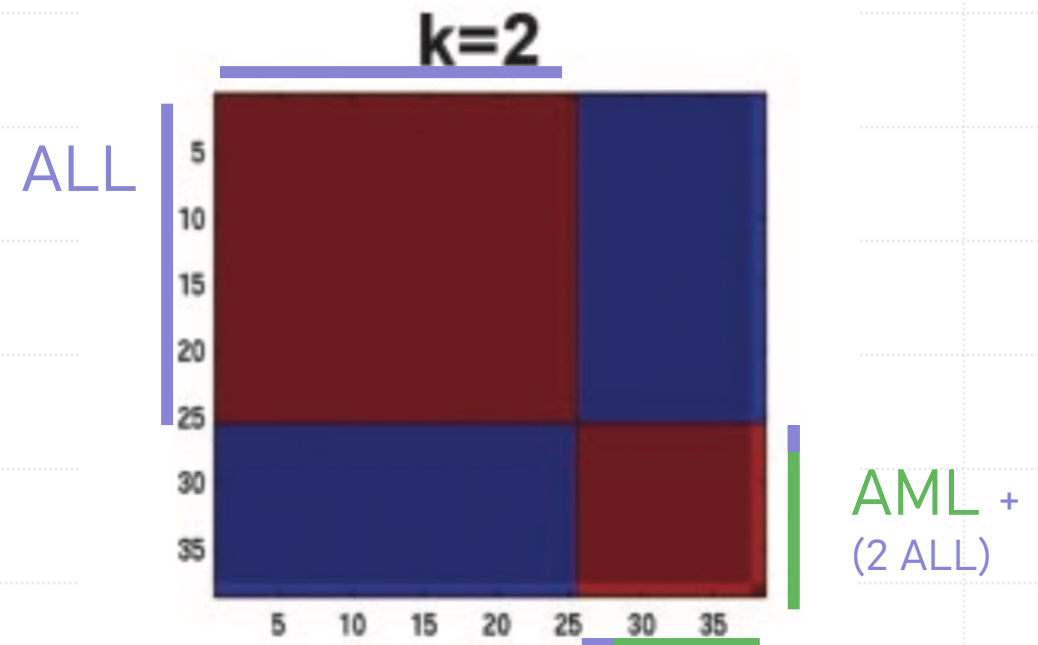
38 éch. moelle

-> 11 AML

-> 8 T-ALL

-> 19 B-ALL

999 genes (HU6800)

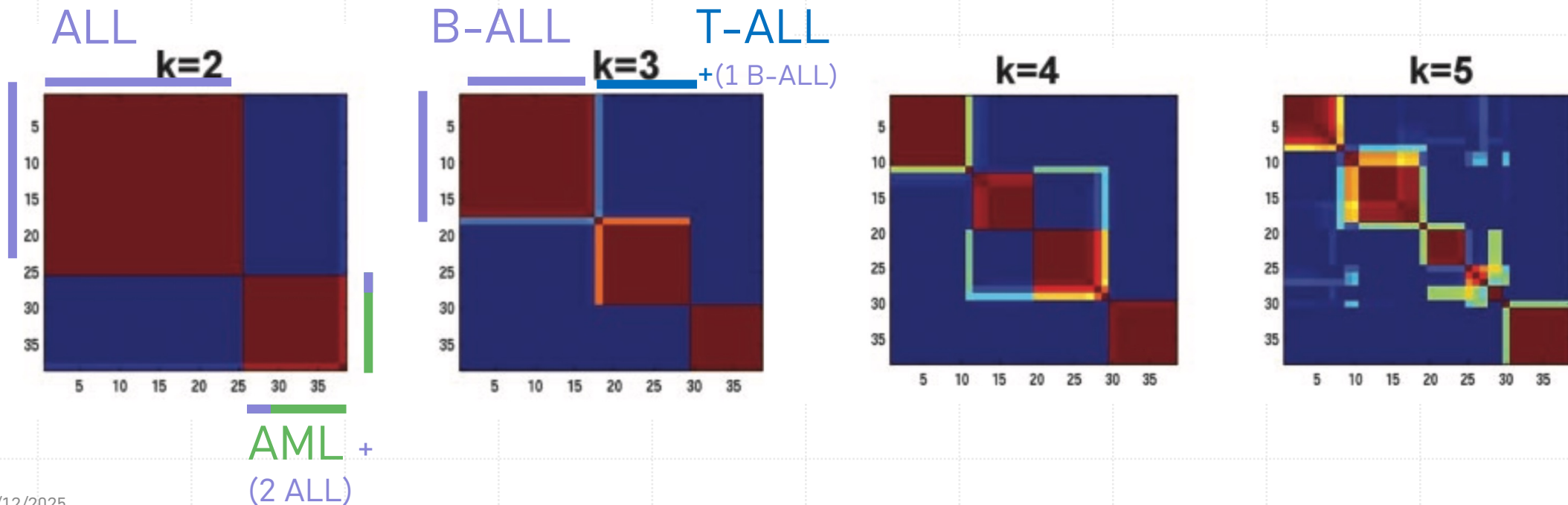


# 1.3 Applications NMF

## Différentes applications

Comment déterminer le nombre optimal de sources  $K$  ?!

1. Classification de tumeurs [Brunet, Tamayo, Golub, et Mesirov. PNAS, 2004.](#)



# 1.3 Applications NMF

Nouvelle sous-division ?

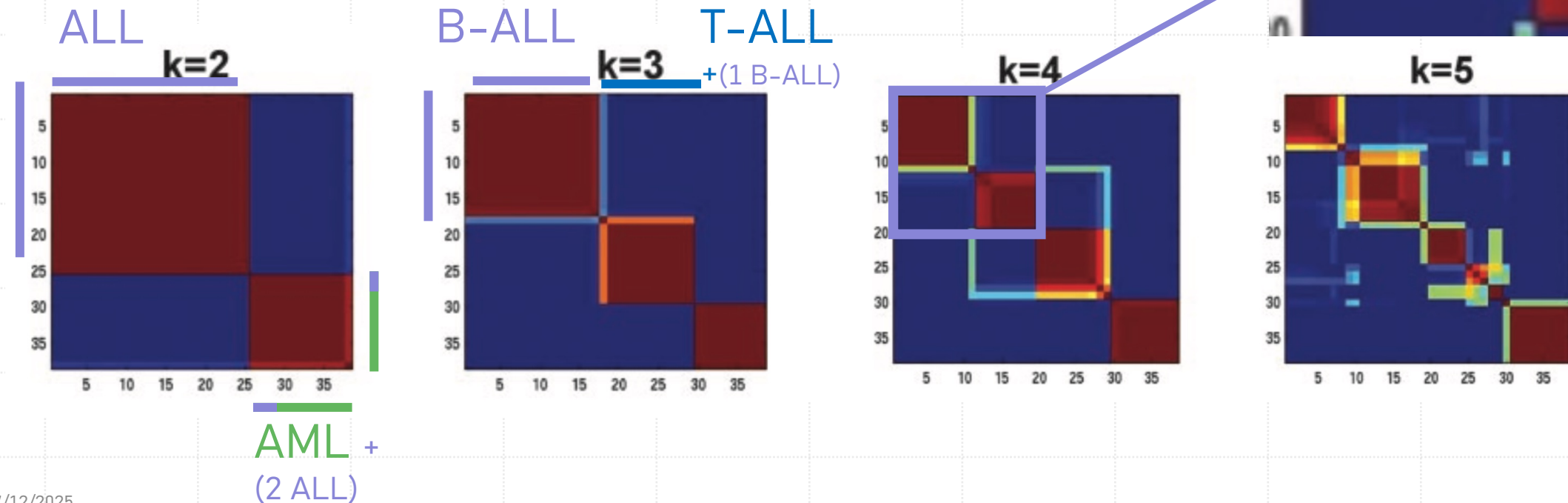
k=4

## Différentes applications

Comment déterminer le nombre optimal de sources K ?!

### 1. Classification de tumeurs

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



# 1.3 Applications NMF

---

## Différentes applications

Comment déterminer 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 déterminer 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  $4 \times 3 = 12$  mutations monobase

# 1.3 Applications NMF

## Différentes applications

Comment déterminer 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  $4 \times 3 = 12$  mutations monobase  
+ 1 nucl. à g. et 1 nucl. à d. donc  $4 \times 4 = 16$  contextes }  $12 \times 16 = 192$  mutations monobase



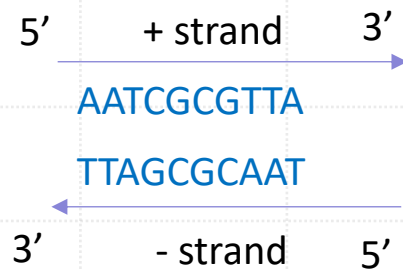
# 1.3 Applications NMF

## Différentes applications

Comment déterminer 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  $4 \times 3 = 12$  mutations monobase  
+ 1 nucl. à g. et 1 nucl. à d. donc  $4 \times 4 = 16$  contextes }  $12 \times 16 = 192$  mutations monobase



# 1.3 Applications NMF

## Différentes applications

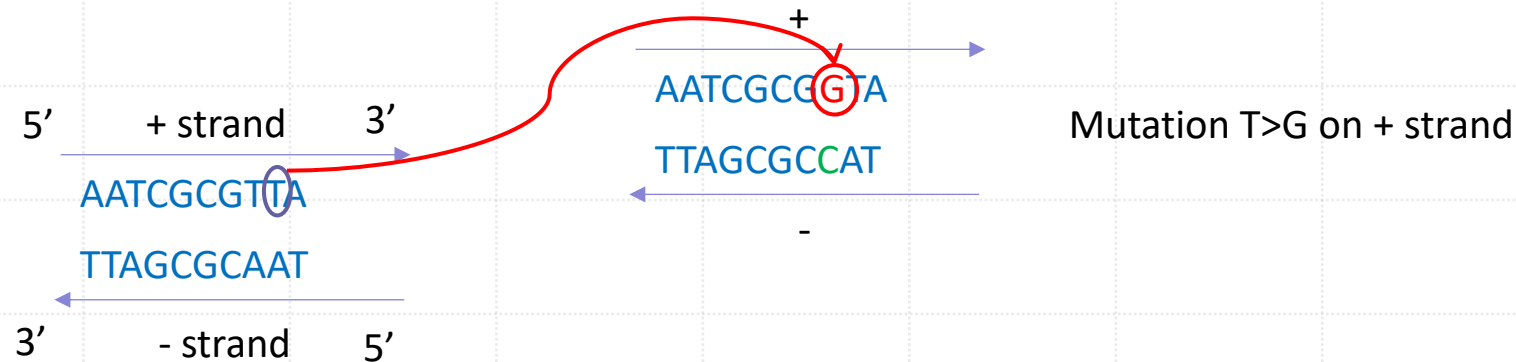
Comment déterminer le nombre optimal de sources K ?!

## 2. Signatures mutationnelles

[Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012](#)

[Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013](#)

**Contexte ?** 4 nucléotides A, C, G, T donc  $4 \times 3 = 12$  mutations monobase  
+ 1 nucl. à g. et 1 nucl. à d. donc  $4 \times 4 = 16$  contextes }  $12 \times 16 = 192$  mutations monobase



# 1.3 Applications NMF

## Différentes applications

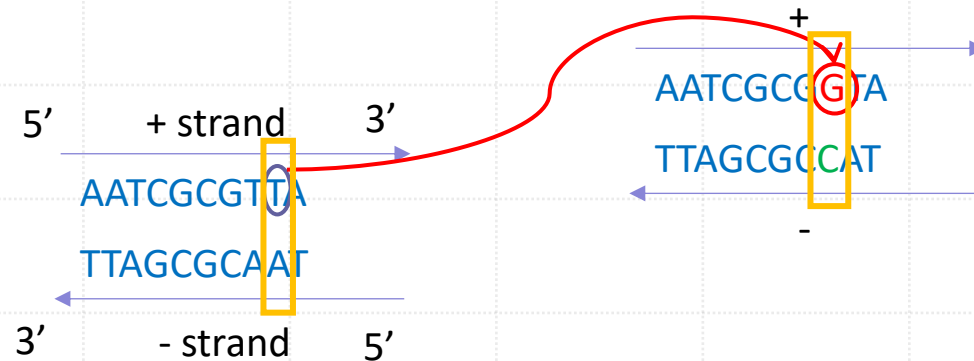
Comment déterminer le nombre optimal de sources K ?!

## 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  $4 \times 3 = 12$  mutations monobase  
+ 1 nucl. à g. et 1 nucl. à d. donc  $4 \times 4 = 16$  contextes }  $12 \times 16 = 192$  mutations monobase



# 1.3 Applications NMF

## Différentes applications

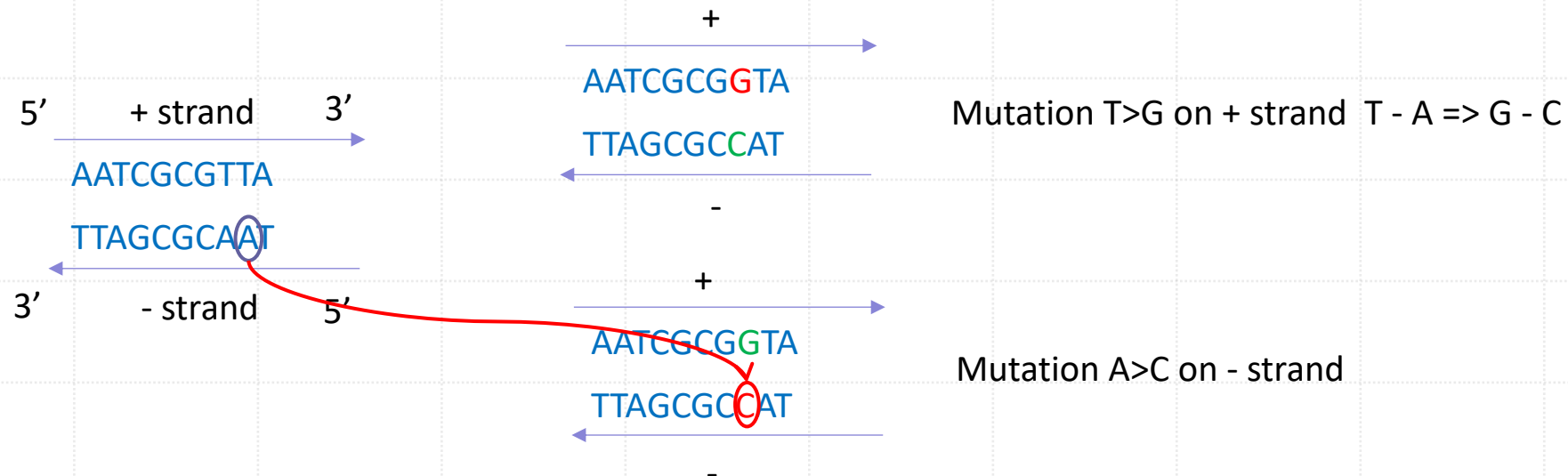
Comment déterminer le nombre optimal de sources K ?!

## 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  $4 \times 3 = 12$  mutations monobase  
+ 1 nucl. à g. et 1 nucl. à d. donc  $4 \times 4 = 16$  contextes }  $12 \times 16 = 192$  mutations monobase



# 1.3 Applications NMF

## Différentes applications

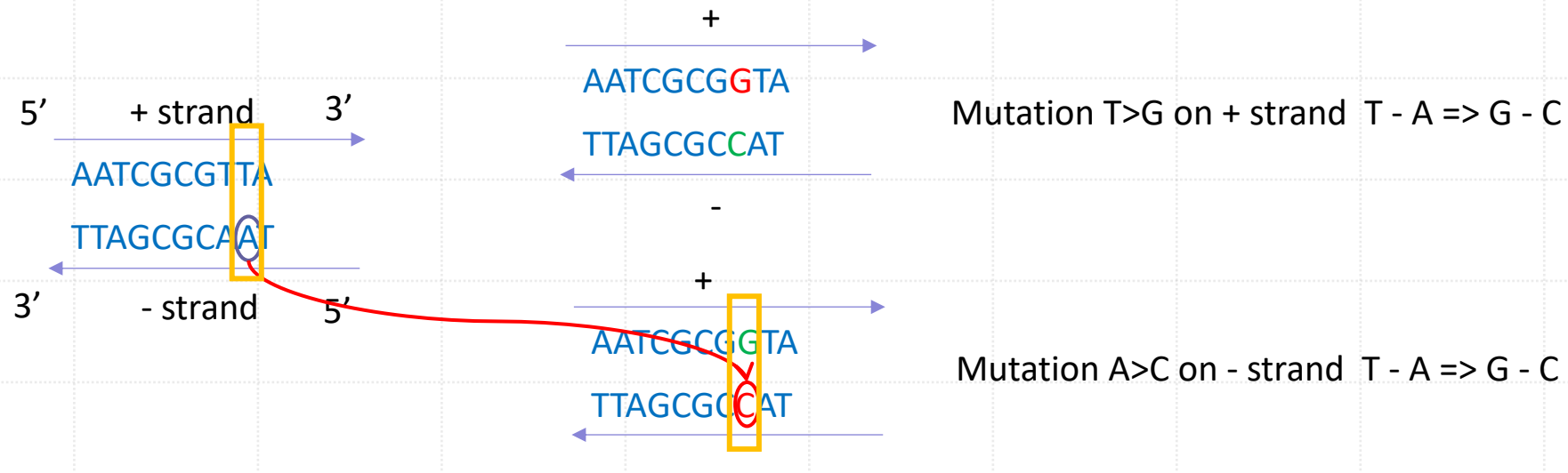
Comment déterminer le nombre optimal de sources K ?!

## 2. Signatures mutationnelles

[Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012](#)

[Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013](#)

**Contexte ?** 4 nucléotides A, C, G, T donc  $4 \times 3 = 12$  mutations monobase  
+ 1 nucl. à g. et 1 nucl. à d. donc  $4 \times 4 = 16$  contextes }  $12 \times 16 = 192$  mutations monobase



# 1.3 Applications NMF

## Différentes applications

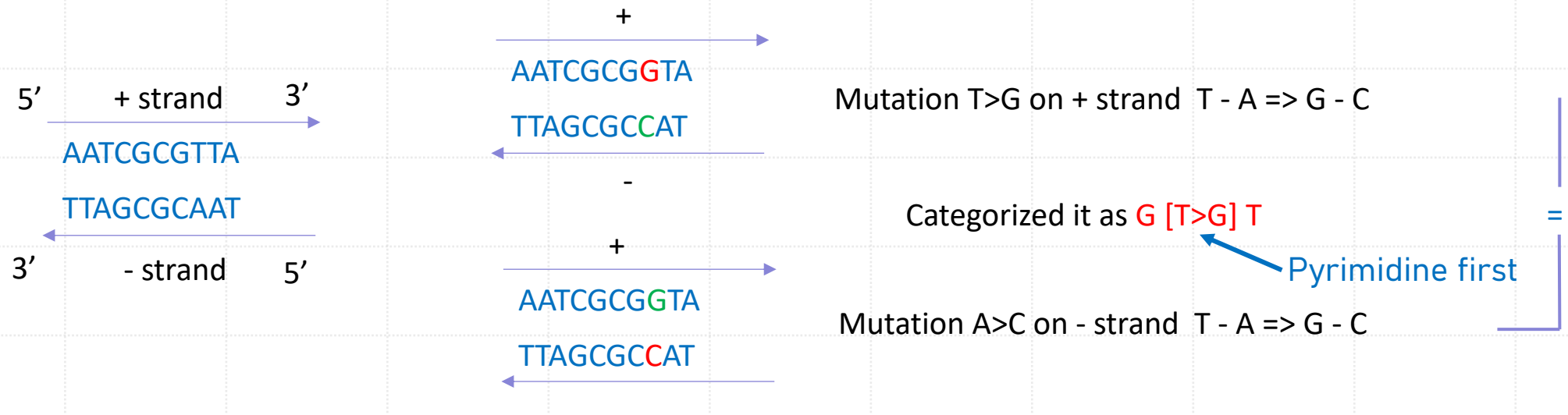
Comment déterminer le nombre optimal de sources K ?!

## 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  $4 \times 3 = 12$  mutations monobase  
+ 1 nucl. à g. et 1 nucl. à d. donc  $4 \times 4 = 16$  contextes }  $12 \times 16 = 192$  mutations monobase



# 1.3 Applications NMF

## Différentes applications

Comment déterminer 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  $4 \times 3 = 12$  mutations monobase  
+ 1 nucl. à g. et 1 nucl. à d. donc  $4 \times 4 = 16$  contextes }  $12 \times 16 = 192$  mutations monobase

$4 \times 3 / 2 = 6$  mutations monobase  
+ 1 nucl. à g. et 1 nucl. à d. donc  $4 \times 4 = 16$  contextes }  $12 \times 8 = \mathbf{96}$  mutations monobase

# 1.3 Applications NMF

## Différentes applications

Comment déterminer le nombre optimal de sources  $K$  ?!

## 2. Signatures mutationnelles

[Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012](#)

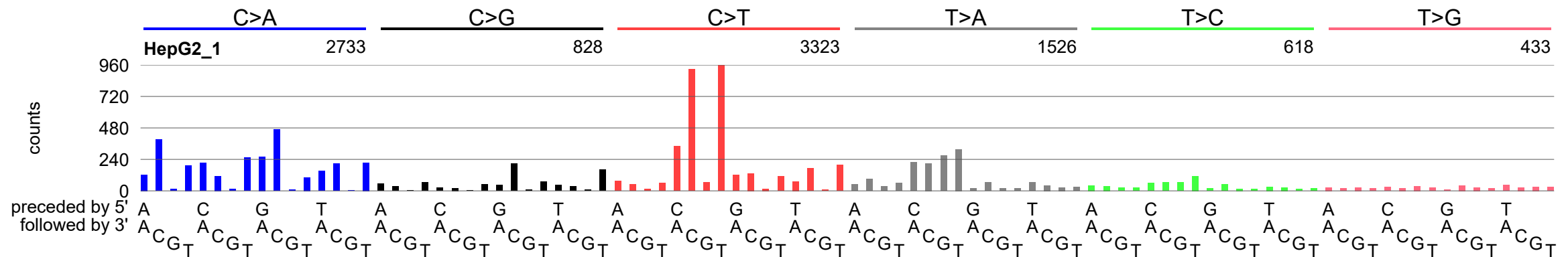
[Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013](#)

### Contexte ?

+ 1 nucl. à g. et 1 nucl. à d. donc  $4 \times 4 = 16$  contextes

$4 \times 3 / 2 = 6$  mutations monobase

12 x 8 = **96 mutations monobase**





# 1.3 Applications NMF

## Différentes applications

Comment déterminer le nombre optimal de sources K ?!

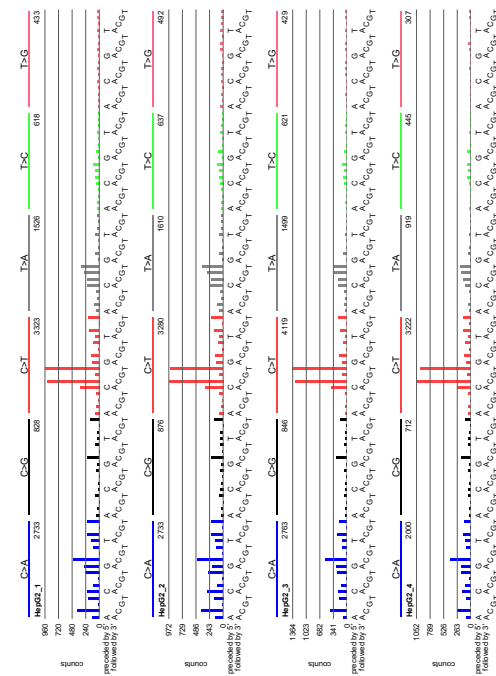
## 2. Signatures mutationnelles

[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 déterminer le nombre optimal de sources K ?!

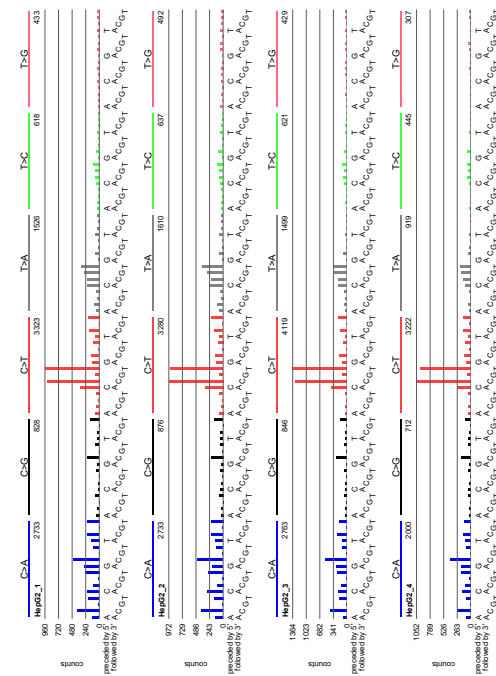
## 2. Signatures mutationnelles

[Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012](#)

[Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013](#)

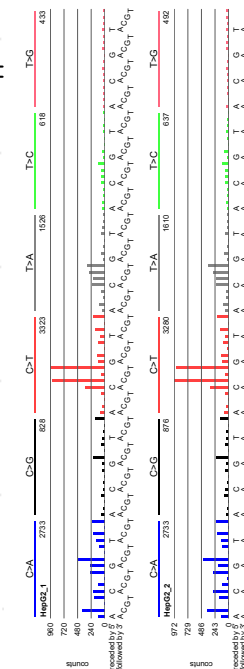
$$\mathbf{V} = \begin{bmatrix} v_{1,1} & \cdots & v_{1,N} \\ \vdots & \ddots & \vdots \\ v_{96,1} & \cdots & v_{96,N} \end{bmatrix}$$

Exemple: N=4 génomes



≈

$$\mathbf{W} =$$



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

# 1.3 Applications NMF

## Différentes applications

Comment déterminer le nombre optimal de sources K ?!

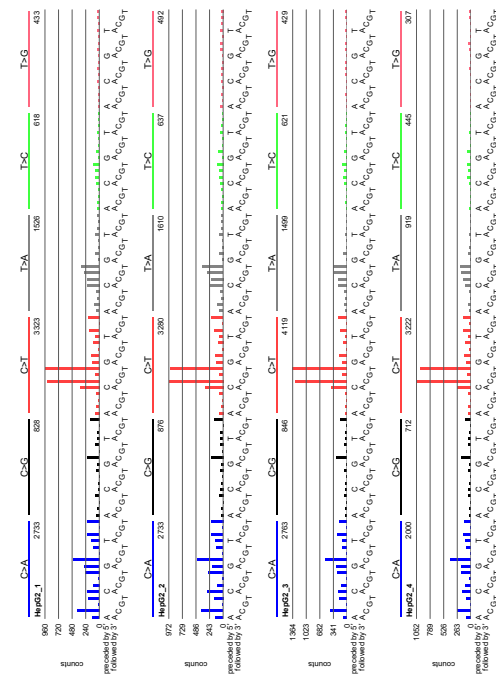
## 2. Signatures mutationnelles

[Nik-Zainal, Alexandrov, Wedge, et al. Cell 2012](#)

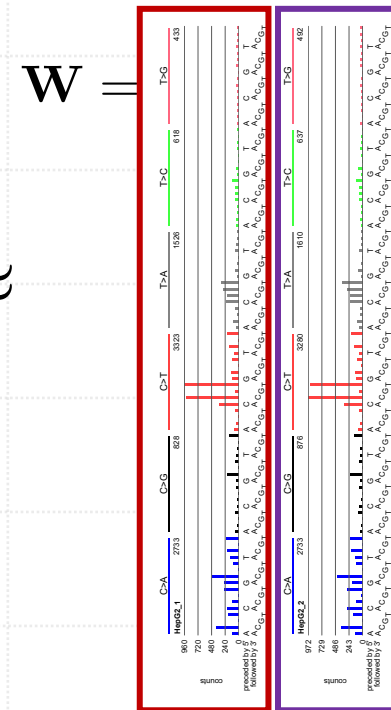
[Alexandrov, Nik-Zainal, Wedge, Campbell, Stratton. Cell Reports 2013](#)

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

Exemple: N=4 génomes



≈



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

Signature 1

Signature 2

# 1.3 Applications NMF

---

## Différentes applications

Comment déterminer 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

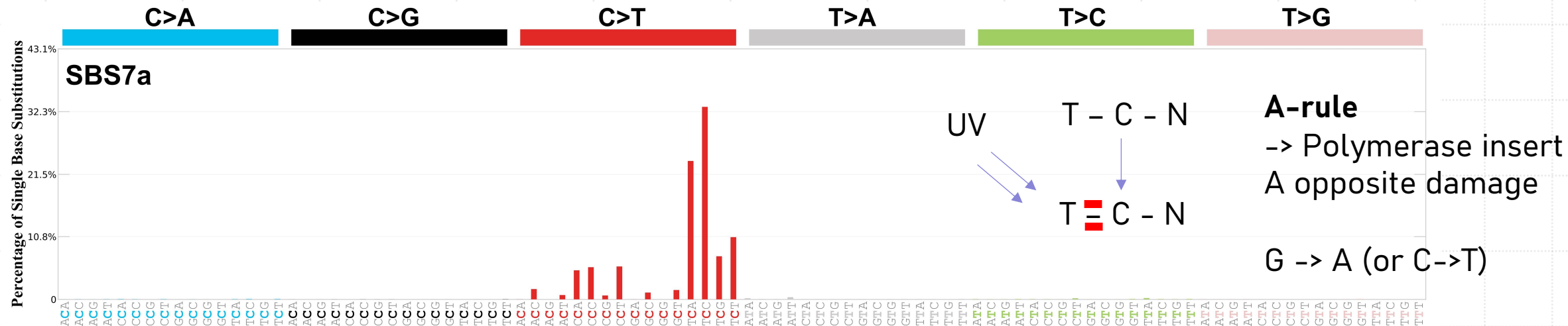
<https://cancer.sanger.ac.uk/signatures/sbs/>

# 1.3 Applications NMF

## Différentes applications

Comment déterminer le nombre optimal de sources K ?!

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



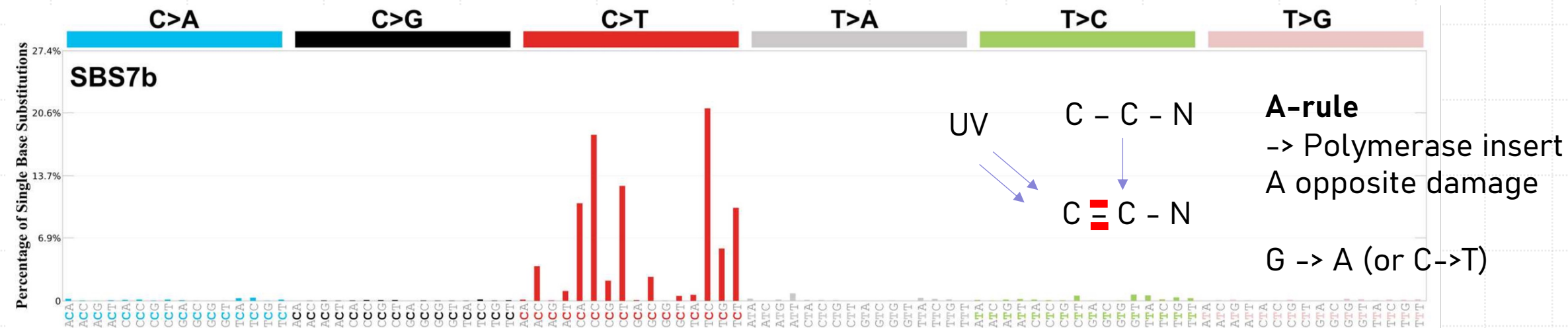
Exposition aux UV

# 1.3 Applications NMF

## Différentes applications

Comment déterminer le nombre optimal de sources K ?!

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



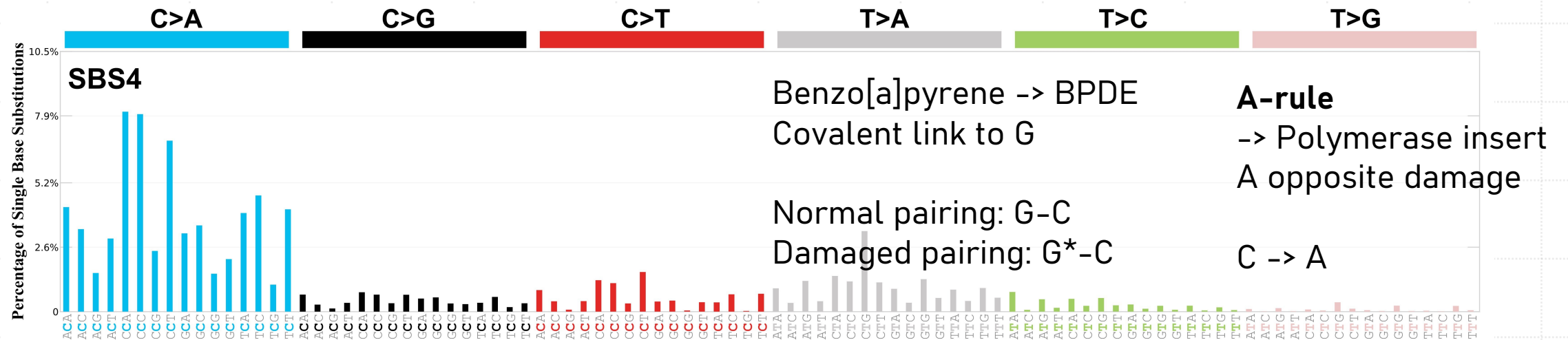
Exposition aux UV

# 1.3 Applications NMF

## Différentes applications

Comment déterminer le nombre optimal de sources K ?!

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



Benzo[a]pyrene -> BPDE  
Covalent link to G

Normal pairing: G-C  
Damaged pairing: G\*-C

**A-rule**  
-> Polymerase insert  
A opposite damage

C -> A

Exposition au tabac

# Sommaire

---

## 1. Factorisation matricielle non négative

1. Factorisations matricielles
2. Factorisation non négative
3. Applications NMF

## 2. Analyse de survie

1. Concepts
2. Modèle de Cox



# 2.1 Concepts

---

## Notations

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

# 2.1 Concepts

## Notations

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

## Définition fonction survie

La **fonction de survie** est donnée par

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

# 2.1 Concepts

## Définition taux de risque instantané

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

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

Soit:  $f_D$  (resp  $F_D$ ) densité (resp f.r.) de  $T_D$ . 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

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

Soit:  $f_D$  (resp  $F_D$ ) densité (resp f.r.) de  $T_D$ . 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

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

Soit:  $f_D$  (resp  $F_D$ ) densité (resp f.r.) de  $T_D$ . Alors

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

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

# 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  $T_{D,1:n}$ , 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  $T_{D,1:n}$ , 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,

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

# 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  $T_{D,1:n}$ , 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,

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

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



## 2.2 Modèle de Cox

### Entraînement par maximum de vraisemblance

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

## 2.2 Modèle de Cox

### Entraînement par maximum de vraisemblance

- $\mathbb{P}(T_i = t_i | \Delta_i = 0) = \mathbb{P}(T_{D,i} \geq t_i) = S_i(t_i)$
  - $\mathbb{P}(T_i = t_i | \Delta_i = 1) = \mathbb{P}(T_{D,i} = t_i) = \lambda_i(t_i) S_i(t_i)$
- }  $\mathbb{P}(T_i = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$

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

## 2.2 Modèle de Cox

### Entraînement par maximum de vraisemblance

- $\mathbb{P}(T_i = t_i | \Delta_i = 0) = \mathbb{P}(T_{D,i} \geq t_i) = S_i(t_i)$
  - $\mathbb{P}(T_i = t_i | \Delta_i = 1) = \mathbb{P}(T_{D,i} = t_i) = \lambda_i(t_i) S_i(t_i)$
- }  $\mathbb{P}(T_i = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$

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

## 2.2 Modèle de Cox

### Entraînement par maximum de vraisemblance

- $\mathbb{P}(T_i = t_i | \Delta_i = 0) = \mathbb{P}(T_{D,i} \geq t_i) = S_i(t_i)$
  - $\mathbb{P}(T_i = t_i | \Delta_i = 1) = \mathbb{P}(T_{D,i} = t_i) = \lambda_i(t_i) S_i(t_i)$
- }  $\mathbb{P}(T_i = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$

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

**Patients encore à risque au temps i**

## 2.2 Modèle de Cox

### Entraînement par maximum de vraisemblance

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

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

## 2.2 Modèle de Cox

### Entraînement par maximum de vraisemblance

- $\mathbb{P}(T_i = t_i | \Delta_i = 0) = \mathbb{P}(T_{D,i} \geq t_i) = S_i(t_i)$
  - $\mathbb{P}(T_i = t_i | \Delta_i = 1) = \mathbb{P}(T_{D,i} = t_i) = \lambda_i(t_i) S_i(t_i)$
- }  $\mathbb{P}(T_i = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$

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

**Vraisemblance partielle**

## 2.2 Modèle de Cox

### Entraînement par maximum de vraisemblance

- $\mathbb{P}(T_i = t_i | \Delta_i = 0) = \mathbb{P}(T_{D,i} \geq t_i) = S_i(t_i)$
  - $\mathbb{P}(T_i = t_i | \Delta_i = 1) = \mathbb{P}(T_{D,i} = t_i) = \lambda_i(t_i) S_i(t_i)$
- }  $\mathbb{P}(T_i = t_i | \Delta_i = \delta_i) = \lambda(t_i)^{\delta_i} S_i(t_i)$

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

**Vraisemblance partielle**

## 2.2 Modèle de Cox

### Exemple

individual	$T_i$	$\delta_i$	$Z_i$
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3



# 2.2 Modèle de Cox

## Exemple

individual	$T_i$	$\delta_i$	$Z_i$
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

ordered failure					Likelihood contribution
$j$	time	$T_i$	$R(T_i)$	$i_j$	$\left[ e^{\beta Z_i} / \sum_{j \in \mathcal{R}(T_i)} e^{\beta Z_j} \right]^{\delta_i}$
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	$T_i$	$\delta_i$	$Z_i$	ordered failure		$R(T_i)$	$i_j$	Likelihood contribution $\left[ e^{\beta Z_i} / \sum_{j \in \mathcal{R}(T_i)} e^{\beta Z_j} \right]^{\delta_i}$
				$j$	time $T_i$			
1	9	1	4					
2	8	0	5	1	6	{1,2,3,4}	3	$e^{7\beta} / [e^{4\beta} + e^{5\beta} + e^{7\beta} + e^{3\beta}]$
3	6	1	7					
4	10	1	3	2	8	{1,2,4}	2	1

## 2.2 Modèle de Cox

### Exemple

individual	$T_i$	$\delta_i$	$Z_i$	ordered failure		$R(T_i)$	$i_j$	Likelihood contribution $\left[ e^{\beta Z_i} / \sum_{j \in \mathcal{R}(T_i)} e^{\beta Z_j} \right]^{\delta_i}$
				$j$	time $T_i$			
1	9	1	4					
2	8	0	5	1	6	{1,2,3,4}	3	$e^{7\beta} / [e^{4\beta} + e^{5\beta} + e^{7\beta} + e^{3\beta}]$
3	6	1	7					
4	10	1	3	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	$T_i$	$\delta_i$	$Z_i$	ordered failure		$i_j$	Likelihood contribution $\left[ e^{\beta Z_i} / \sum_{j \in \mathcal{R}(T_i)} e^{\beta Z_j} \right]^{\delta_i}$
				$j$	time $T_i$	$R(T_i)$	
1	9	1	4				
2	8	0	5	1	6	$\{1,2,3,4\}$	$e^{7\beta} / [e^{4\beta} + e^{5\beta} + e^{7\beta} + e^{3\beta}]$
3	6	1	7				
4	10	1	3	2	8	$\{1,2,4\}$	1
				3	9	$\{1,4\}$	$e^{4\beta} / [e^{4\beta} + e^{3\beta}]$
				4	10	$\{4\}$	$e^{3\beta} / e^{3\beta} = 1$