

MathExp 2018: Chaînes de Markov

# Simulation des chaînes de Markov

Ana Busic

Inria Paris - DI ENS

<http://www.di.ens.fr/~busic/>

Saint Flour, Mai 2018

# Plan

Système à événements discrets

Echantillonnage de la distribution stationnaire

Temps de couplage

Annexe : Distance de variation et couplage de chaînes de Markov

Annexe : Critère de Foster

Annexe : Rappels et table d'alias

# Système à événements discrets

SED :  $(\mathcal{X}, \pi^0, A, p, \cdot)$

- ▶  $\mathcal{X}$  est un **espace d'états fini**.
- ▶  $\pi^0$  est une **distribution initiale** :  
 $\pi_x^0 \geq 0$  est la probabilité que le système soit dans l'état  $x \in \mathcal{X}$  à l'instant 0.
- ▶  $A$  est un **ensemble d'événements fini**.
- ▶  $p$  est une mesure de probabilité sur  $A$  :  
 $p_a > 0$  est la **probabilité de l'événement**  $a \in A$ .
- ▶  $\cdot$  est un opérateur qui décrit l'action des événements dans  $A$  sur les états dans  $\mathcal{X}$  - **fonction de transition** :

$$\cdot : \mathcal{X} \times A \rightarrow \mathcal{X}, \quad (x, a) \mapsto x \cdot a.$$

# Fonction de transition

Extension aux séquences (mots) d'événements fini :

$$\cdot : \mathcal{X} \times A^n \rightarrow \mathcal{X}, \quad n \in \mathbb{N}.$$

Soit  $a_{1 \rightarrow n} \stackrel{\Delta}{=} a_1 \dots a_n$  dans  $A^n$  :

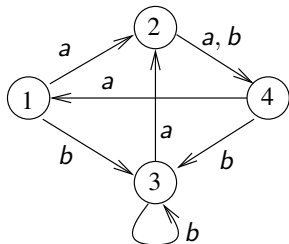
$$(x, a_{1 \rightarrow n}) \mapsto x \cdot a_{1 \rightarrow n} \stackrel{\Delta}{=} x \cdot a_1 \cdot a_2 \cdot \dots \cdot a_n.$$

# Évolution du système

Sur  $n$  étapes :

1. On choisit l'état initial  $X_0$  selon la distribution  $\pi^0$ .
2. Pour  $i = 1$  à  $n$  faire :
  - ▶ Choisir un événement  $a_i \in A$  selon  $p$
  - ▶  $X_i := X_{i-1} \cdot a_i = X_0 \cdot a_{1 \rightarrow i}$

Exemple :



Soit  $p_a = 1/3$ ,  $p_b = 2/3$ , et  
 $\pi^0 = (1/4, 1/4, 1/4, 1/4)$ .

Une **trajectoire** possible :  
 $1 - 3 - 3 - 2 - 4 - 1 - 3 - 3 - \dots$   
en partant de l'état 1 et pour la  
séquence d'événements  
 $bbababb \dots$

# SED vs. chaîne de Markov

Un SED  $\mathcal{A} = (\mathcal{X}, \pi^0, A, p, \cdot)$  induit une chaîne de Markov (en temps discret homogène) :

▶ chaîne de Markov

- ▶ Soient  $\{a_n\}_{n \in \mathbb{N}}$  une suite *i.i.d.* d'événements de  $A$  distribués selon  $p$  et  $X_0$  distribué selon  $\pi^0$ .
- ▶ Alors  $\{X_n \stackrel{\Delta}{=} X_0 \cdot a_{1 \rightarrow n}\}_{n \in \mathbb{N}}$  est une chaîne de Markov avec la matrice de transition  $P$  :

$$\text{pour tout } x, y \text{ dans } \mathcal{X}, \quad P_{x,y} = \sum_{a \in A : x \cdot a = y} p_a. \quad (1)$$

# SED vs. chaîne de Markov

Un SED  $\mathcal{A} = (\mathcal{X}, \pi^0, A, p, \cdot)$  induit une chaîne de Markov (en temps discret homogène) :

▶ chaîne de Markov

- ▶ Soient  $\{a_n\}_{n \in \mathbb{N}}$  une suite *i.i.d.* d'événements de  $A$  distribués selon  $p$  et  $X_0$  distribué selon  $\pi^0$ .
- ▶ Alors  $\{X_n \stackrel{\Delta}{=} X_0 \cdot a_{1 \rightarrow n}\}_{n \in \mathbb{N}}$  est une chaîne de Markov avec la matrice de transition  $P$  :

$$\text{pour tout } x, y \text{ dans } \mathcal{X}, \quad P_{x,y} = \sum_{a \in A : x \cdot a = y} p_a. \quad (1)$$

De plus, pour toute chaîne de Markov finie avec la distribution initiale  $\pi^0$  et matrice de transition  $P$  il existe un SED  $\mathcal{A} = (\mathcal{X}, \pi^0, A, p, \cdot)$  vérifiant (1). Pas unique.

# SED vs. chaîne de Markov

Un SED  $\mathcal{A} = (\mathcal{X}, \pi^0, A, p, \cdot)$  induit une chaîne de Markov (en temps discret homogène) :

▶ chaîne de Markov

- ▶ Soient  $\{a_n\}_{n \in \mathbb{N}}$  une suite *i.i.d.* d'événements de  $A$  distribués selon  $p$  et  $X_0$  distribué selon  $\pi^0$ .
- ▶ Alors  $\{X_n \stackrel{\Delta}{=} X_0 \cdot a_{1 \rightarrow n}\}_{n \in \mathbb{N}}$  est une chaîne de Markov avec la matrice de transition  $P$  :

$$\text{pour tout } x, y \text{ dans } \mathcal{X}, \quad P_{x,y} = \sum_{a \in A : x \cdot a = y} p_a. \quad (1)$$

De plus, pour toute chaîne de Markov finie avec la distribution initiale  $\pi^0$  et matrice de transition  $P$  il existe un SED  $\mathcal{A} = (\mathcal{X}, \pi^0, A, p, \cdot)$  vérifiant (1). Pas unique.

Si  $|\mathcal{X}| = N$ , il existe un SED avec au plus  $N^2$  événements.



# Echantillonnage de la distribution stationnaire

$\{X_n\}_{n \in \mathbb{N}}$  une chaîne irréductible et apériodique.

## Question

*Comment échantillonner sa distribution stationnaire  $\pi$  ?*

# Echantillonnage de la distribution stationnaire

$\{X_n\}_{n \in \mathbb{N}}$  une chaîne irréductible et apériodique.

## Question

*Comment échantillonner sa distribution stationnaire  $\pi$  ?*

Solution A : résoudre le système linéaire  $\pi = \pi P$  pour calculer  $\pi$ , puis utiliser l'échantillonnage d'une distribution discrète.

# Echantillonnage de la distribution stationnaire

$\{X_n\}_{n \in \mathbb{N}}$  une chaîne irréductible et apériodique.

## Question

*Comment échantillonner sa distribution stationnaire  $\pi$  ?*

Solution A : résoudre le système linéaire  $\pi = \pi P$  pour calculer  $\pi$ , puis utiliser l'échantillonnage d'une distribution discrète.

## Question

*Comment éviter le calcul de  $\pi$  (et même  $P$ ) ?*


# Simulation Monte-Carlo (MCMC)

Algorithme :

- ▶ On choisit l'état initial  $X_0$  selon  $\pi^0$ .
- ▶ Pour  $i = 1$  jusqu'à  $n$  :
  - ▶ On choisit  $a_i$  selon  $p$ .
  - ▶  $X_i = X_{i-1} \cdot a_i$ .

Sortie : un échantillon de  $\pi^n = \pi^0 P^n$ .

Complexité :  $O(\mathcal{C}n)$ , où  $\mathcal{C}$  est la complexité du calcul de la fonction de transition .

Remarque : échantillonnage d'une distribution discrète en  $\Theta(1)$  en utilisant la méthode de alias [Walker, 77]. 

Inconvénient : **approximation**.

Estimation d'erreur est difficile : dépend de la deuxième valeur propre de  $P$  - difficile à calculer [Brémaud, Asmussen and Glynn].

# Simulation parfaite (exacte)

Objectif :

- ▶ des échantillons distribués selon la distribution  $\pi$  exactement sans jamais la calculer (ni  $P$ ).
- ▶ temps d'arrêt fini.

Propp et Wilson (1996) ont proposé d'utiliser le **couplage depuis le passé** comme une technique effective de simulation parfaite.

Couplage depuis le passé [Borovkov 75, Glynn 96] - travaux d'un intérêt essentiellement théorique et existentiel (reposant sur la théorie du renouvellement).

## Couplage vers l'avant

On considère une suite  $a_1, \dots, a_n, \dots$  d'événements i.i.d. selon la distribution  $p$  et l'ensemble des trajectoires qu'elle induit en partant de tous les états du SED.

Notation :  $E \cdot a = \{x \cdot a : x \in E\}$ .

On pose  $E_0 = \mathcal{X}$ ,  $E_1 = \mathcal{X} \cdot a_1$ ,  $\dots$ ,  $E_n = E_{n-1} \cdot a_n = \mathcal{X} \cdot a_{1 \rightarrow n}$ .

# Couplage vers l'avant

On considère une suite  $a_1, \dots, a_n, \dots$  d'événements i.i.d. selon la distribution  $p$  et l'ensemble des trajectoires qu'elle induit en partant de tous les états du SED.

Notation :  $E \cdot a = \{x \cdot a : x \in E\}$ .

On pose  $E_0 = \mathcal{X}$ ,  $E_1 = \mathcal{X} \cdot a_1$ ,  $\dots$ ,  $E_n = E_{n-1} \cdot a_n = \mathcal{X} \cdot a_{1 \rightarrow n}$ .

Par construction les tailles des ensembles  $E_n$  sont décroissantes, donc il existe  $\ell = \lim_{n \rightarrow \infty} |E_n|$ .

## Définition

Le SED a la *propriété de couplage (couple)* si  $\ell = 1$  pour presque toutes les suites d'événements  $a_1, \dots, a_n, \dots$

Le *temps de couplage* est alors  $\tau^f \stackrel{\Delta}{=} \min\{n \in \mathbb{N} \text{ t.q. } |E_n| = 1\}$ .

# Couplage vers l'avant

## Théorème

*Un SED couple si et seulement si il admet un mot couplant (synchronisant).*

La variable aléatoire  $\tau^f$  a une queue de distribution géométrique et son espérance est finie :

soit  $a_1 \cdots a_k$  un mot couplant, alors

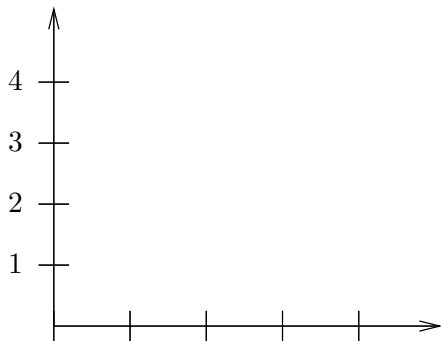
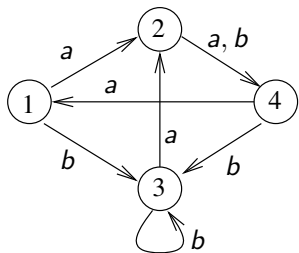
$$P(\tau^f > n.k) \leq (1 - p_{a_1} \cdots p_{a_k})^n$$

et

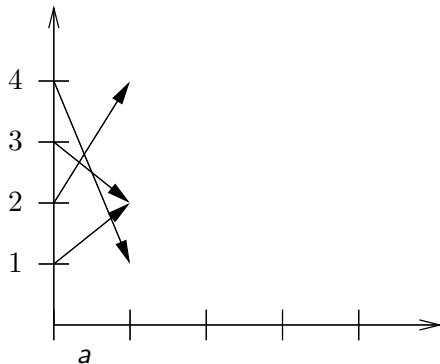
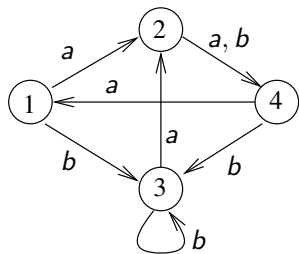
$$\mathbf{E}\tau^f \leq k/(p_{a_1} \cdots p_{a_k}).$$



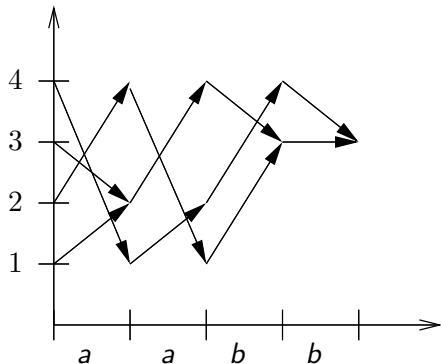
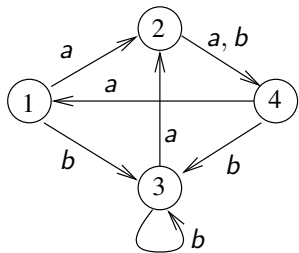
# Couplage vers l'avant



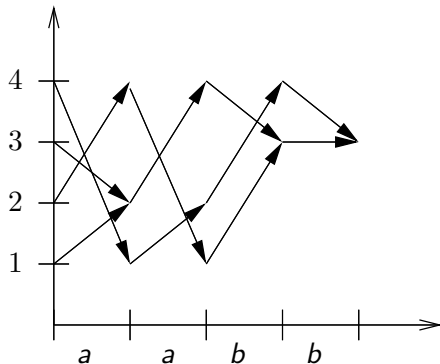
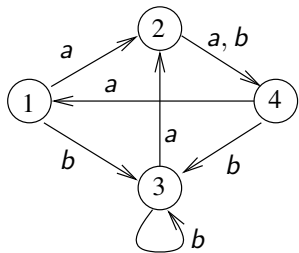
# Couplage vers l'avant



# Couplage vers l'avant



# Couplage vers l'avant

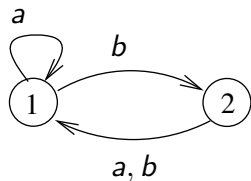


Couplage en état 3.

# Questions

## Question

*Couplage vers l'avant donne-t-il un échantillon stationnaire ?*

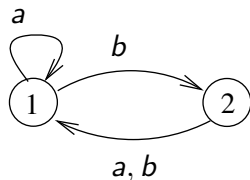


$$p_a = p_b = 0.5$$

# Questions

## Question

*Couplage vers l'avant donne-t-il un échantillon stationnaire ?*

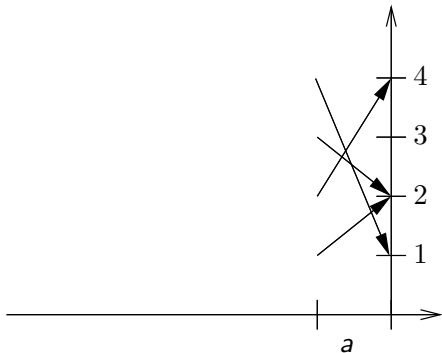
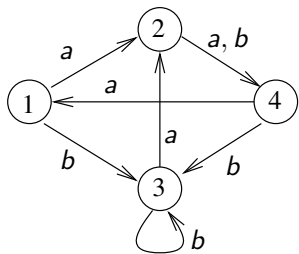


$$p_a = p_b = 0.5$$

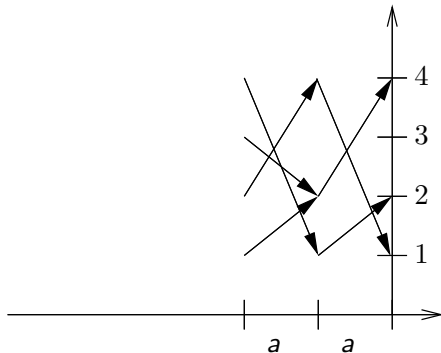
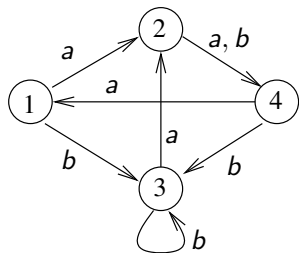
## Question

*Un SED dont la chaîne de Markov est irréductible et apériodique couple-t-il toujours ?*

# Couplage depuis le passé : l'idée

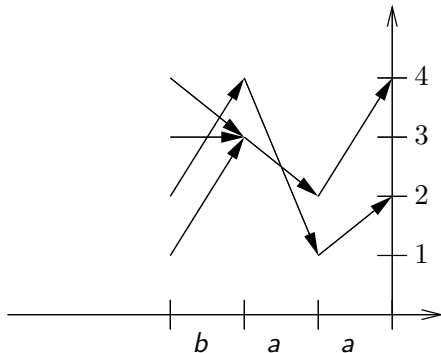
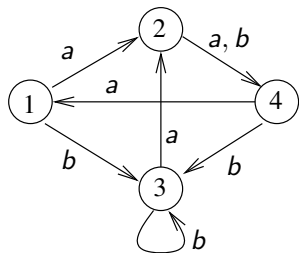


# Couplage depuis le passé : l'idée

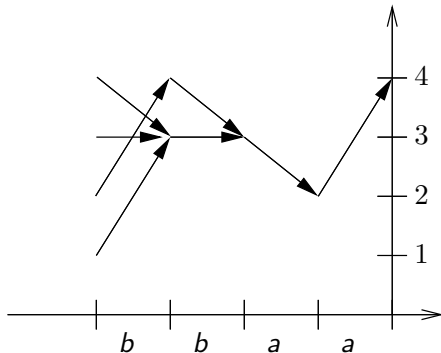
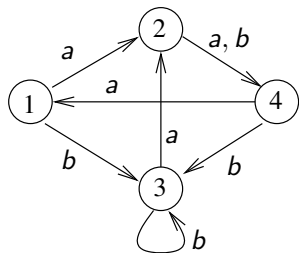




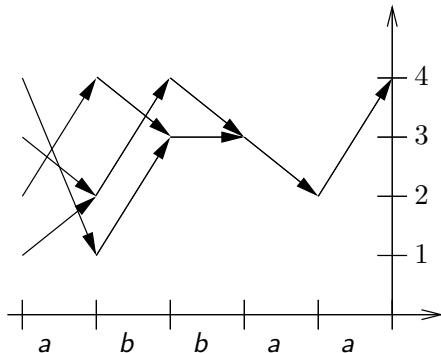
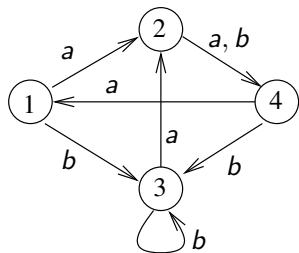
# Couplage depuis le passé : l'idée



## Couplage depuis le passé : l'idée



# Couplage depuis le passé : l'idée



# Couplage depuis le passé

$a_1, a_2, \dots, a_n, \dots$  une suite *i.i.d.* d'événements distribués selon  $p$ .

On regarde l'ensemble des trajectoires depuis le passé :

$$S_0 = \mathcal{X}, S_1 = \mathcal{X} \cdot a_1, S_2 = \mathcal{X} \cdot a_2 a_1,$$

$$S_n \stackrel{\Delta}{=} \mathcal{X} \cdot a_{n \rightarrow 1} \stackrel{\Delta}{=} \mathcal{X} \cdot a_n \cdot a_{n-1} \dots a_1.$$

## Couplage depuis le passé

$a_1, a_2, \dots, a_n, \dots$  une suite *i.i.d.* d'événements distribués selon  $p$ .  
On regarde l'ensemble des trajectoires depuis le passé :

$$S_0 = \mathcal{X}, S_1 = \mathcal{X} \cdot a_1, S_2 = \mathcal{X} \cdot a_2 a_1, \\ S_n \stackrel{\Delta}{=} \mathcal{X} \cdot a_{n \rightarrow 1} \stackrel{\Delta}{=} \mathcal{X} \cdot a_n \cdot a_{n-1} \dots a_1.$$

Par construction, on a  $S_n \subseteq S_{n-1} \subseteq \dots \subseteq S_1 \subseteq \mathcal{X}$ .

Il existe donc un ensemble limite  $S$ .

Si le SED est couplant, alors  $S$  est réduit à un seul élément p.s.. On définit alors le temps de couplage depuis le passé :

$$\tau^b \stackrel{\Delta}{=} \min\{n \in \mathbb{N} \text{ t.q. } |S_n| = 1\}.$$

## Couplage depuis le passé

$a_1, a_2, \dots, a_n, \dots$  une suite *i.i.d.* d'événements distribués selon  $p$ .  
On regarde l'ensemble des trajectoires depuis le passé :

$$S_0 = \mathcal{X}, S_1 = \mathcal{X} \cdot a_1, S_2 = \mathcal{X} \cdot a_2 a_1, \\ S_n \stackrel{\Delta}{=} \mathcal{X} \cdot a_{n \rightarrow 1} \stackrel{\Delta}{=} \mathcal{X} \cdot a_n \cdot a_{n-1} \dots a_1.$$

Par construction, on a  $S_n \subseteq S_{n-1} \subseteq \dots \subseteq S_1 \subseteq \mathcal{X}$ .

Il existe donc un ensemble limite  $S$ .

Si le SED est couplant, alors  $S$  est réduit à un seul élément p.s.. On définit alors le temps de couplage depuis le passé :

$$\tau^b \stackrel{\Delta}{=} \min\{n \in \mathbb{N} \text{ t.q. } |S_n| = 1\}.$$

Comme la probabilité d'occurrence de  $a_n \dots a_1$  et celle de  $a_1 \dots a_n$  sont les mêmes :  $\mathbb{P}(a_n \dots a_1) = p_{a_n} \dots p_{a_1} = \mathbb{P}(a_1 \dots a_n)$ , alors  $\tau^f \stackrel{D}{=} \tau^b$ .

# Couplage depuis le passé

## Théorème

- ▶ Si SED est couplant,  $\mathbf{E}(\tau^b) < \infty$ .
- ▶ Soit  $Y$  (l'unique) état dans  $S_{\tau^b}$ . Alors  $Y$  est distribué selon  $\pi$ .

# Couplage depuis le passé

## Théorème

- ▶ Si SED est couplant,  $\mathbf{E}(\tau^b) < \infty$ .
- ▶ Soit  $Y$  (l'unique) état dans  $S_{\tau^b}$ . Alors  $Y$  est distribué selon  $\pi$ .

## Démonstration.

- ▶ Puisque  $\tau^f \stackrel{D}{=} \tau^b$ ,  $\mathbf{E}(\tau^b) = \mathbf{E}(\tau^f) < \infty$ .



# Couplage depuis le passé

## Théorème

- ▶ Si SED est couplant,  $\mathbf{E}(\tau^b) < \infty$ .
- ▶ Soit  $Y$  (l'unique) état dans  $S_{\tau^b}$ . Alors  $Y$  est distribué selon  $\pi$ .

## Démonstration.

- ▶ Puisque  $\tau^f \stackrel{D}{=} \tau^b$ ,  $\mathbf{E}(\tau^b) = \mathbf{E}(\tau^f) < \infty$ .
- ▶ On doit montrer :  $\mathbb{P}(Y = x) = \pi_x, \forall x \in \mathcal{X}$ .

Soit  $x \in \mathcal{X}$  un état fixé. Il suffit de montrer que pour tout  $\epsilon > 0$ ,

$$|\mathbb{P}(Y = x) - \pi_x| \leq \epsilon.$$

## Démonstration (suite)

Soit  $\epsilon > 0$  fixé. Puisque  $\tau^b < \infty$  p.s., il existe un  $m$  suffisamment grand t.q.

$$\mathbb{P}(|S_m| = 1) \geq 1 - \epsilon.$$

## Démonstration (suite)

Soit  $\epsilon > 0$  fixé. Puisque  $\tau^b < \infty$  p.s., il existe un  $m$  suffisamment grand t.q.

$$\mathbb{P}(|S_m| = 1) \geq 1 - \epsilon.$$

Fixons  $m$ . Soit l'état  $X$  choisi selon  $\pi$  et observons la trajectoire en partant de  $X$  à l'instant  $-m$ , en utilisant la même suite d'événements  $a_m, \dots, a_1$ . Notons par  $\tilde{Y} = X \cdot a_{m \rightarrow 1}$ .

## Démonstration (suite)

Soit  $\epsilon > 0$  fixé. Puisque  $\tau^b < \infty$  p.s., il existe un  $m$  suffisamment grand t.q.

$$\mathbb{P}(|S_m| = 1) \geq 1 - \epsilon.$$

Fixons  $m$ . Soit l'état  $X$  choisi selon  $\pi$  et observons la trajectoire en partant de  $X$  à l'instant  $-m$ , en utilisant la même suite d'événements  $a_m, \dots, a_1$ . Notons par  $\tilde{Y} = X \cdot a_{m \rightarrow 1}$ .

Puisque  $X$  est stationnaire,  $\tilde{Y}$  l'est aussi. De plus,

$$\mathbb{P}(Y \neq \tilde{Y}) \leq \epsilon.$$

## Démonstration (suite)

Nous avons :

$$\begin{aligned}\mathbb{P}(Y = x) - \pi_x &= \mathbb{P}(Y = x) - \mathbb{P}(\tilde{Y} = x) \\ &\leq \mathbb{P}(Y = x, \tilde{Y} \neq x) \\ &\leq \mathbb{P}(Y \neq \tilde{Y}) \leq \epsilon.\end{aligned}$$

## Démonstration (suite)

Nous avons :

$$\begin{aligned}\mathbb{P}(Y = x) - \pi_x &= \mathbb{P}(Y = x) - \mathbb{P}(\tilde{Y} = x) \\ &\leq \mathbb{P}(Y = x, \tilde{Y} \neq x) \\ &\leq \mathbb{P}(Y \neq \tilde{Y}) \leq \epsilon.\end{aligned}$$

Aussi :

$$\begin{aligned}\pi_x - \mathbb{P}(Y = x) &= \mathbb{P}(\tilde{Y} = x) - \mathbb{P}(Y = x) \\ &\leq \mathbb{P}(\tilde{Y} = x, Y \neq x) \\ &\leq \mathbb{P}(Y \neq \tilde{Y}) \leq \epsilon.\end{aligned}$$

## Démonstration (suite)

Nous avons :

$$\begin{aligned}\mathbb{P}(Y = x) - \pi_x &= \mathbb{P}(Y = x) - \mathbb{P}(\tilde{Y} = x) \\ &\leq \mathbb{P}(Y = x, \tilde{Y} \neq x) \\ &\leq \mathbb{P}(Y \neq \tilde{Y}) \leq \epsilon.\end{aligned}$$

Aussi :

$$\begin{aligned}\pi_x - \mathbb{P}(Y = x) &= \mathbb{P}(\tilde{Y} = x) - \mathbb{P}(Y = x) \\ &\leq \mathbb{P}(\tilde{Y} = x, Y \neq x) \\ &\leq \mathbb{P}(Y \neq \tilde{Y}) \leq \epsilon.\end{aligned}$$

Donc,

$$|\mathbb{P}(Y = x) - \pi_x| \leq \epsilon.$$

# Couplage depuis le passé : l'algorithme

Algorithme :

**Data :**  $(\mathcal{X}, A, \cdot, p)$ .

**Result :** Un état de  $\mathcal{X}$  distribué selon  $\pi$ .

**begin**

$n := 1$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$T(x) := x$ ;

**end**

**repeat**

    Tirer  $a$  selon  $p$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$S(x) := T(x \cdot a)$ ;

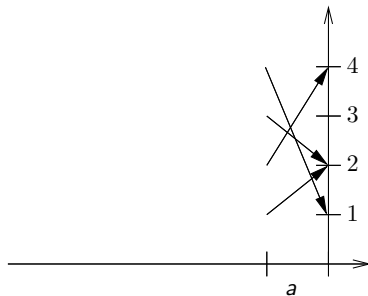
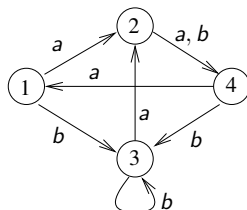
**end**

$T := S$ ;

**until**  $|\{T(x), x \in \mathcal{X}\}| = 1$ ;

**return**  $T(x_1)$  ( $x_1$  un élément dans  $\mathcal{X}$ )

**end**



	1	2	3	4
$T$	1	2	3	4
$S$	2	4	2	1



# Couplage depuis le passé : l'algorithme

Algorithme :

**Data :**  $(\mathcal{X}, A, \cdot, p)$ .

**Result :** Un état de  $\mathcal{X}$  distribué selon  $\pi$ .

**begin**

$n := 1$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$T(x) := x$ ;

**end**

**repeat**

    Tirer  $a$  selon  $p$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$S(x) := T(x \cdot a)$ ;

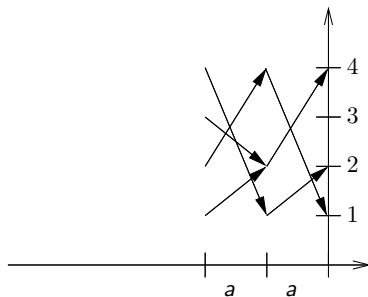
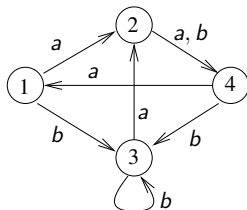
**end**

$T := S$ ;

**until**  $|\{T(x), x \in \mathcal{X}\}| = 1$ ;

**return**  $T(x_1)$  ( $x_1$  un élément dans  $\mathcal{X}$ )

**end**



	1	2	3	4
$T$	2	4	2	1
$S$	4	1	4	2

# Couplage depuis le passé : l'algorithme

Algorithme :

**Data :**  $(\mathcal{X}, A, \cdot, p)$ .

**Result :** Un état de  $\mathcal{X}$  distribué selon  $\pi$ .

**begin**

$n := 1$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$T(x) := x$ ;

**end**

**repeat**

    Tirer  $a$  selon  $p$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$S(x) := T(x \cdot a)$ ;

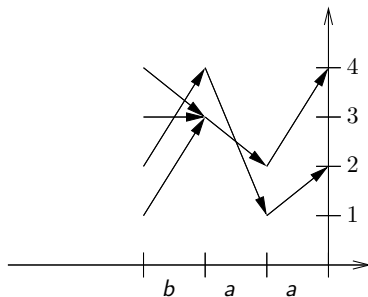
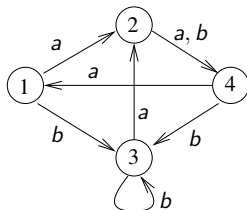
**end**

$T := S$ ;

**until**  $|\{T(x), x \in \mathcal{X}\}| = 1$ ;

**return**  $T(x_1)$  ( $x_1$  un élément dans  $\mathcal{X}$ )

**end**



	1	2	3	4
$T$	4	1	4	2
$S$	4	2	4	4

# Couplage depuis le passé : l'algorithme

Algorithme :

**Data :**  $(\mathcal{X}, A, \cdot, p)$ .

**Result :** Un état de  $\mathcal{X}$  distribué selon  $\pi$ .

**begin**

$n := 1$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$T(x) := x$ ;

**end**

**repeat**

    Tirer  $a$  selon  $p$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$S(x) := T(x \cdot a)$ ;

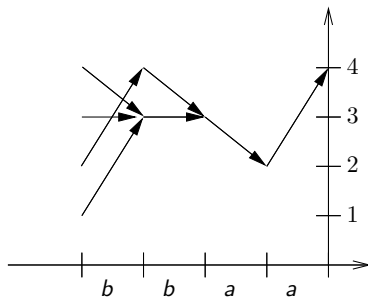
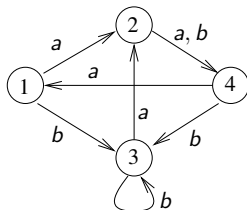
**end**

$T := S$ ;

**until**  $|\{T(x), x \in \mathcal{X}\}| = 1$ ;

**return**  $T(x_1)$  ( $x_1$  un élément dans  $\mathcal{X}$ )

**end**



	1	2	3	4
$T$	4	2	4	4
$S$	4	4	4	4

# Couplage depuis le passé : l'algorithme

Algorithme :

**Data :**  $(\mathcal{X}, A, \cdot, p)$ .

**Result :** Un état de  $\mathcal{X}$  distribué selon  $\pi$ .

**begin**

$n := 1$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$T(x) := x$ ;

**end**

**repeat**

        Tirer  $a$  selon  $p$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$S(x) := T(x \cdot a)$ ;

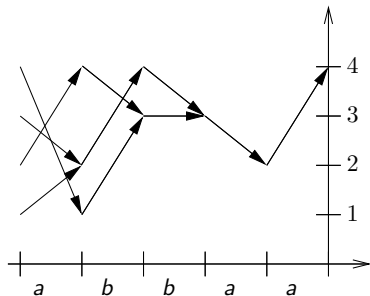
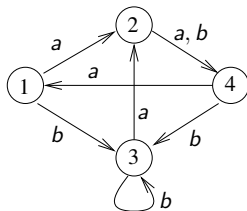
**end**

$T := S$ ;

**until**  $|\{T(x), x \in \mathcal{X}\}| = 1$ ;

**return**  $T(x_1)$  ( $x_1$  un élément dans  $\mathcal{X}$ )

**end**



	1	2	3	4
$T$	4	4	4	4
$S$	4	4	4	4

# Couplage depuis le passé : l'algorithme

Algorithme :

**Data** :  $(\mathcal{X}, A, \cdot, p)$ .

**Result** : Un état de  $\mathcal{X}$  distribué selon  $\pi$ .

**begin**

$n := 1$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$T(x) := x$ ;

**end**

**repeat**

        Tirer  $a$  selon  $p$ ;

**foreach** état  $x \in \mathcal{X}$  **do**

$S(x) := T(x \cdot a)$ ;

**end**

$T := S$ ;

**until**  $|\{T(x), x \in \mathcal{X}\}| = 1$ ;

**return**  $T(x_1)$  ( $x_1$  un élément dans  $\mathcal{X}$ )

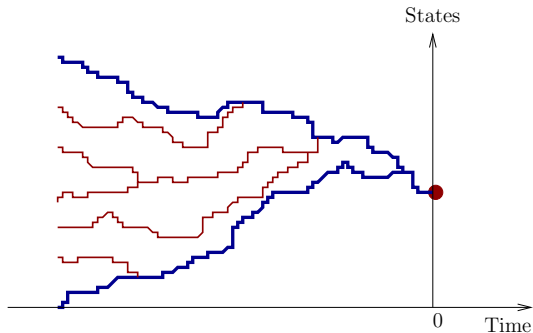
**end**

Inconvénient : La complexité  $O(\mathcal{C} \times |\mathcal{X}| \times \tau^b)$ , avec  $\mathcal{C}$  la complexité du calcul de la fonction de transition  $(\cdot)$

## SED monotone

On suppose que l'espace d'état est muni d'une relation d'ordre ( $\preceq$ ) et que la fonction de transition de l'automate est monotone :

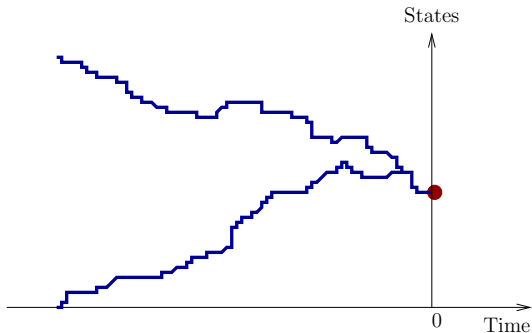
$$x \preceq y \Rightarrow \forall a \in A, x \cdot a \preceq y \cdot a.$$



# SED monotone

On suppose que l'espace d'état est muni d'une relation d'ordre ( $\preceq$ ) et que la fonction de transition de l'automate est monotone :

$$x \preceq y \Rightarrow \forall a \in A, x \cdot a \preceq y \cdot a.$$



## SED monotone

Soit  $E$  l'ensemble des éléments extrémaux de  $\mathcal{X}$  pour  $\preceq$ .

On note  $\bar{U} \triangleq \{x \in \mathcal{X} \text{ t.q. } \exists u, v \in U \mid u \preceq x \preceq v\}$ .

Par définition,  $\mathcal{X} = \bar{E}$ .



# SED monotone

Soit  $E$  l'ensemble des éléments extrémaux de  $\mathcal{X}$  pour  $\preceq$ .

On note  $\overline{U} \triangleq \{x \in \mathcal{X} \text{ t.q. } \exists u, v \in U \mid u \preceq x \preceq v\}$ .

Par définition,  $\mathcal{X} = \overline{E}$ .

Par monotonie,  $\overline{U} \cdot a \subseteq \overline{U \cdot a}$ ,

et donc, pour tout mot  $a_n, \dots, a_1$ ,

$$E \cdot a_{n \rightarrow 1} \subseteq \mathcal{X} \cdot a_{n \rightarrow 1} \subseteq \overline{E \cdot a_{n \rightarrow 1}}.$$

Toutes les trajectoires sont “capturées” entre les trajectoires extrémales.

# SED monotone

Soit  $E$  l'ensemble des éléments extrémaux de  $\mathcal{X}$  pour  $\preceq$ .

On note  $\overline{U} \triangleq \{x \in \mathcal{X} \text{ t.q. } \exists u, v \in U \mid u \preceq x \preceq v\}$ .

Par définition,  $\mathcal{X} = \overline{E}$ .

Par monotonie,  $\overline{U} \cdot a \subseteq \overline{U \cdot a}$ ,

et donc, pour tout mot  $a_n, \dots, a_1$ ,

$$E \cdot a_{n \rightarrow 1} \subseteq \mathcal{X} \cdot a_{n \rightarrow 1} \subseteq \overline{E \cdot a_{n \rightarrow 1}}.$$

Toutes les trajectoires sont “capturées” entre les trajectoires extrémales.

Si  $a_n, \dots, a_1$  est un mot couplant pour  $E$ ,

$$E \cdot a_{n \rightarrow 1} = \mathcal{X} \cdot a_{n \rightarrow 1} = \overline{E \cdot a_{n \rightarrow 1}}.$$

# Exemple

Événements :

a - arrivée (probabilité  $\lambda$ )

d - départ (probabilité  $\mu$ )

$$\lambda + \mu = 1$$



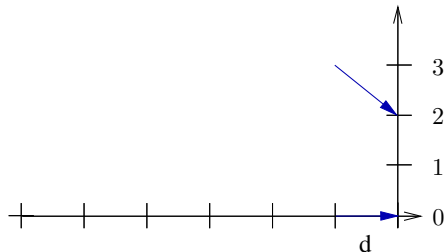
# Exemple

Événements :

a - arrivée (probabilité  $\lambda$ )

d - départ (probabilité  $\mu$ )

$$\lambda + \mu = 1$$



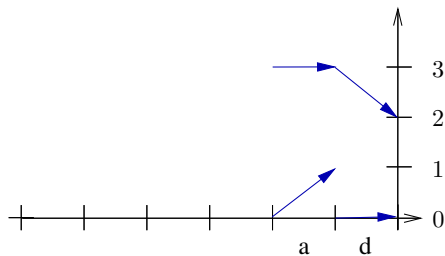
# Exemple

Événements :

a - arrivée (probabilité  $\lambda$ )

d - départ (probabilité  $\mu$ )

$$\lambda + \mu = 1$$



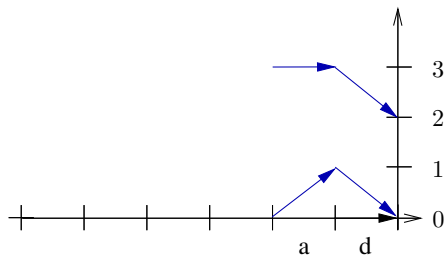
# Exemple

Événements :

a - arrivée (probabilité  $\lambda$ )

d - départ (probabilité  $\mu$ )

$$\lambda + \mu = 1$$



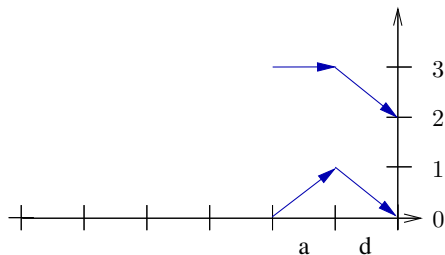
# Exemple

Événements :

a - arrivée (probabilité  $\lambda$ )

d - départ (probabilité  $\mu$ )

$$\lambda + \mu = 1$$







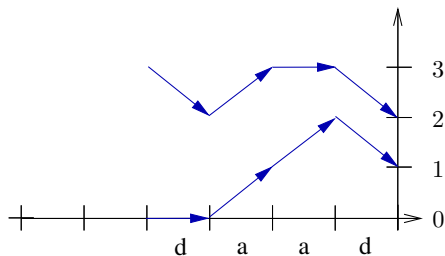
# Exemple

Événements :

a - arrivée (probabilité  $\lambda$ )

d - départ (probabilité  $\mu$ )

$$\lambda + \mu = 1$$



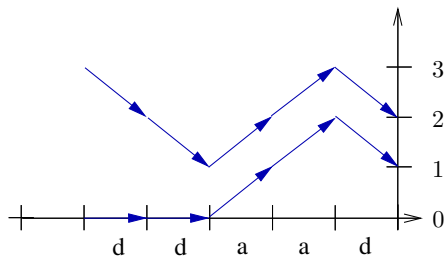
# Exemple

Événements :

a - arrivée (probabilité  $\lambda$ )

d - départ (probabilité  $\mu$ )

$$\lambda + \mu = 1$$



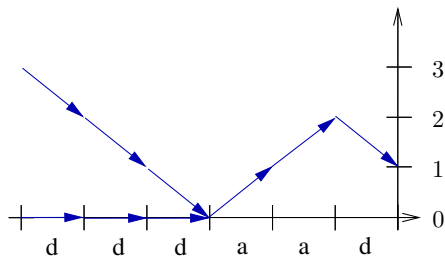
# Exemple

Événements :

a - arrivée (probabilité  $\lambda$ )

d - départ (probabilité  $\mu$ )

$$\lambda + \mu = 1$$



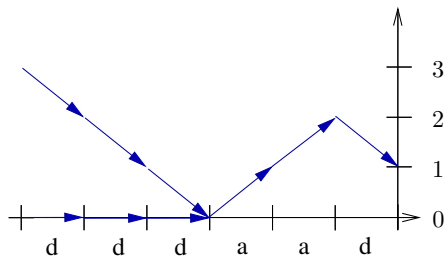
# Exemple

Événements :

a - arrivée (probabilité  $\lambda$ )

d - départ (probabilité  $\mu$ )

$$\lambda + \mu = 1$$



Inconvenient : Complexité  $O((\tau^b)^2 \times \mathcal{C} \times |E|)$   
(comparé à  $O(\tau^b \times \mathcal{C} \times |\mathcal{X}|)$ ).

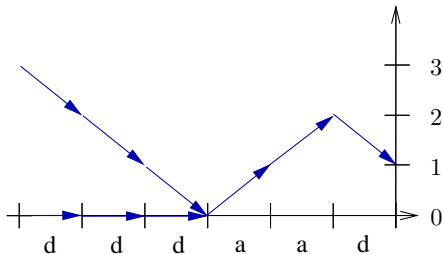
# Exemple

Événements :

a - arrivée (probabilité  $\lambda$ )

d - départ (probabilité  $\mu$ )

$$\lambda + \mu = 1$$



Inconvénient : Complexité  $O((\tau^b)^2 \times \mathcal{C} \times |E|)$   
(comparé à  $O(\tau^b \times \mathcal{C} \times |\mathcal{X}|)$ ).

Solution : **doublement de l'intervalle.**

# Algorithme de simulation parfaite monotone

**Data :**  $(\mathcal{X}, A, \cdot, p)$ ,  $E = \{ \text{états extrémaux} \}$ .

**Result :**  $x \in \mathcal{X}$  généré selon  $\pi$

**begin**

$n = 1$ ;

**repeat**

$S := E$ ;

**for**  $i = n$  **downto**  $\lfloor n/2 \rfloor + 1$  **do**

| Tirer  $a_i$  selon  $p$

**end**

**for**  $i = n$  **downto** 1 **do**

|  $S := S \cdot a_i$ ;

**end**

$n := 2n$ ;

**until**  $|S| = 1$ ;

**return**  $x$  unique état dans  $S$ ;

**end**

# Questions

## Question

*Quelle est la complexité de l'algorithme PSA-monotone ?*

# Questions

## Question

*Quelle est la complexité de l'algorithme PSA-monotone ?*

## Question

*L'algorithme PSA-monotone nécessite de garder en mémoire la suite  $a_1, a_2, \dots$  et de les réutiliser à chaque doublement de l'intervalle.*

*Que se passe-t-il si on utilise des nouveaux tirages (indépendants) à chaque fois ?*



# Questions

## Question

*Quelle est la complexité de l'algorithme PSA-monotone ?*

## Question

*L'algorithme PSA-monotone nécessite de garder en mémoire la suite  $a_1, a_2, \dots$  et de les réutiliser à chaque doublement de l'intervalle.*

*Que se passe-t-il si on utilise des nouveaux tirages (indépendants) à chaque fois ?*

## Question

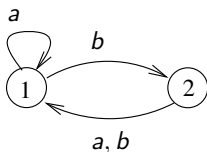
*On suppose que la fonction de transition de l'automate est vérifiée pour chaque  $a \in A$  :*

- ▶  $x \preceq y \Rightarrow x \cdot a \preceq y \cdot a$  (événement monotone) ou
- ▶  $x \preceq y \Rightarrow x \cdot a \succeq y \cdot a$  (événement anti-monotone).

*Montrer qu'on peut utiliser l'algorithme PSA-monotone pour échantillonner  $\pi$ .*

## Pourquoi réutiliser la même suite d'événements ?

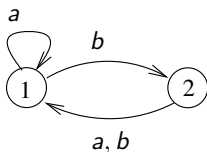
Que se passe-t-il si au lieu de garder  $a_1, \dots, a_n$ , on les retire indépendamment ?



$$p_a = p_b = 0.5$$

## Pourquoi réutiliser la même suite d'événements ?

Que se passe-t-il si au lieu de garder  $a_1, \dots, a_n$ , on les retire indépendamment ?

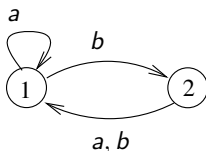


$$p_a = p_b = 0.5$$

Quelle est la probabilité que  $Y = 1$  ? On pose  $M$  la variable aléatoire du nombre de boucles effectués.

## Pourquoi réutiliser la même suite d'événements ?

Que se passe-t-il si au lieu de garder  $a_1, \dots, a_n$ , on les retire indépendamment ?



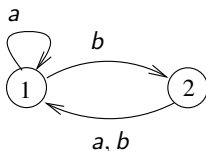
$$p_a = p_b = 0.5$$

Quelle est la probabilité que  $Y = 1$  ? On pose  $M$  la variable aléatoire du nombre de boucles effectués.

$$\mathbf{P}(Y = 1) = \sum_{m=1}^{\infty} \mathbf{P}(M = m, Y = 1)$$

## Pourquoi réutiliser la même suite d'événements ?

Que se passe-t-il si au lieu de garder  $a_1, \dots, a_n$ , on les retire indépendamment ?



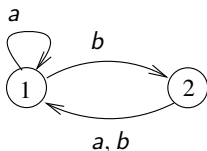
$$p_a = p_b = 0.5$$

Quelle est la probabilité que  $Y = 1$  ? On pose  $M$  la variable aléatoire du nombre de boucles effectués.

$$\begin{aligned} \mathbf{P}(Y = 1) &= \sum_{m=1}^{\infty} \mathbf{P}(M = m, Y = 1) \\ &\geq \mathbf{P}(M = 1, Y = 1) + \mathbf{P}(M = 2, Y = 1) \end{aligned}$$

## Pourquoi réutiliser la même suite d'événements ?

Que se passe-t-il si au lieu de garder  $a_1, \dots, a_n$ , on les retire indépendamment ?



$$p_a = p_b = 0.5$$

Quelle est la probabilité que  $Y = 1$  ? On pose  $M$  la variable aléatoire du nombre de boucles effectués.

$$\begin{aligned} \mathbf{P}(Y = 1) &= \sum_{m=1}^{\infty} \mathbf{P}(M = m, Y = 1) \\ &\geq \mathbf{P}(M = 1, Y = 1) + \mathbf{P}(M = 2, Y = 1) \\ &= \mathbf{P}(M = 1)P(Y = 1|M = 1) + P(M = 2)P(Y = 1|M = 2). \end{aligned}$$

## Pourquoi réutiliser la même suite d'événements ?

On sait que  $\mathbf{P}(M = 1) = 1/2$ , et  $\mathbf{P}(Y = 1|M = 1) = 1$  (le couplage a eu lieu en une étape, donc vient d'avoir lieu).

## Pourquoi réutiliser la même suite d'événements ?

On sait que  $\mathbf{P}(M = 1) = 1/2$ , et  $\mathbf{P}(Y = 1|M = 1) = 1$  (le couplage a eu lieu en une étape, donc vient d'avoir lieu).

De plus,

$$\mathbf{P}(M = 2) = \mathbf{P}(M > 1)\mathbf{P}(M = 2|M > 1) = \frac{1}{2}\left(1 - \left(\frac{1}{2}\right)^2\right) = 3/8$$

et  $\mathbf{P}(Y = 1|M = 2) =$



## Pourquoi réutiliser la même suite d'événements ?

On sait que  $\mathbf{P}(M = 1) = 1/2$ , et  $\mathbf{P}(Y = 1|M = 1) = 1$  (le couplage a eu lieu en une étape, donc vient d'avoir lieu).

De plus,

$$\mathbf{P}(M = 2) = \mathbf{P}(M > 1)\mathbf{P}(M = 2|M > 1) = \frac{1}{2}\left(1 - \left(\frac{1}{2}\right)^2\right) = 3/8$$

et  $\mathbf{P}(Y = 1|M = 2) =$

$$\mathbf{P}(Y = 1|M = 2, \text{ couple à la date } - 1)\mathbf{P}(\text{ couple à la date } - 1|M = 2)$$

## Pourquoi réutiliser la même suite d'événements ?

On sait que  $\mathbf{P}(M = 1) = 1/2$ , et  $\mathbf{P}(Y = 1|M = 1) = 1$  (le couplage a eu lieu en une étape, donc vient d'avoir lieu).

De plus,

$$\mathbf{P}(M = 2) = \mathbf{P}(M > 1)\mathbf{P}(M = 2|M > 1) = \frac{1}{2}\left(1 - \left(\frac{1}{2}\right)^2\right) = 3/8$$

et  $\mathbf{P}(Y = 1|M = 2) =$

$$\begin{aligned} &\mathbf{P}(Y = 1|M = 2, \text{ couple à la date } -1)P(\text{ couple à la date } -1|M = 2) \\ &+ \mathbf{P}(Y = 1|M = 2, \text{ couple à la date } -2)P(\text{ couple à la date } -2|M = 2) \end{aligned}$$

## Pourquoi réutiliser la même suite d'événements ?

On sait que  $\mathbf{P}(M = 1) = 1/2$ , et  $\mathbf{P}(Y = 1|M = 1) = 1$  (le couplage a eu lieu en une étape, donc vient d'avoir lieu).

De plus,

$$\mathbf{P}(M = 2) = \mathbf{P}(M > 1)\mathbf{P}(M = 2|M > 1) = \frac{1}{2}\left(1 - \left(\frac{1}{2}\right)^2\right) = 3/8$$

et  $\mathbf{P}(Y = 1|M = 2) =$

$$\begin{aligned} & \mathbf{P}(Y = 1|M = 2, \text{ couple à la date } -1)P(\text{ couple à la date } -1|M = 2) \\ & + \mathbf{P}(Y = 1|M = 2, \text{ couple à la date } -2)P(\text{ couple à la date } -2|M = 2) \\ = & \frac{1}{2} \frac{2}{3} + 1 \frac{1}{3} = \frac{2}{3}. \end{aligned}$$

## Pourquoi réutiliser la même suite d'événements ?

On sait que  $\mathbf{P}(M = 1) = 1/2$ , et  $\mathbf{P}(Y = 1|M = 1) = 1$  (le couplage a eu lieu en une étape, donc vient d'avoir lieu).

De plus,

$$\mathbf{P}(M = 2) = \mathbf{P}(M > 1)\mathbf{P}(M = 2|M > 1) = \frac{1}{2}\left(1 - \left(\frac{1}{2}\right)^2\right) = 3/8$$

et  $\mathbf{P}(Y = 1|M = 2) =$

$$\begin{aligned} & \mathbf{P}(Y = 1|M = 2, \text{ couple à la date } -1)P(\text{ couple à la date } -1|M = 2) \\ & + \mathbf{P}(Y = 1|M = 2, \text{ couple à la date } -2)P(\text{ couple à la date } -2|M = 2) \\ = & \frac{1}{2} \frac{2}{3} + 1 \frac{1}{3} = \frac{2}{3}. \end{aligned}$$

Donc  $\mathbf{P}(Y = 1) \geq \frac{1}{2} + \frac{3}{8} \frac{2}{3} = \frac{3}{4} > \frac{2}{3}$ .

# Algorithme de Propp et Wilson : simulation exacte

On se donne :

- ▶ une représentation fonctionnelle  $\phi : E \times F \rightarrow E$ ,
- ▶ une suite strictement croissante d'entiers  $N_1 < N_2 < \dots < N_m < \dots$
- ▶ et une suite de variables aléatoires i.i.d. sur  $F, U_{-1}, \dots, U_{-n}, \dots$

Soit  $x \in E$ , on note  $\phi_N(x)$  l'état obtenu en appliquant la suite  $U_{-N}, \dots, U_{-1}$  en partant de  $x$

$$\phi_N(x) = \phi(\dots(\phi(\phi(x, U_{-N}), U_{-N+1}), \dots), U_{-1}).$$

**début**

```
|  $m \leftarrow 0;$   
|  $E' \leftarrow E;$   
| tant que  $|E'| > 1$  faire  
|    $m \leftarrow m + 1;$   
|   pour chaque  $x \in E$  faire  
|     Calculer  $E' = \cup_{x \in E} \{\phi_{-N_m}(x)\};$   
|   fin  
|   retourner  $y$  l'unique élément de  $E'$ 
```

**fin**

**fin**

# Remarques

Quelques remarques/problèmes :

- ▶ Temps de couplage ?
- ▶ SED non-monotones ?
- ▶ Garder la suite aléatoire en mémoire ?

# Temps de couplage

En général un problème très difficile.

Quelques résultats théoriques dans des files d'attente :

- ▶ Une file M/M/1/C. Cas  $\lambda \neq \mu$  : temps de couplage  $\Theta(C)$ .  
Cas :  $\lambda = \mu$ ,  $O(C^2)$ .
- ▶ Réseaux de Jackson avec les capacités finies :

Pour un réseau acyclique de  $K$  files M/M/1/C, on peut montrer que  $\mathbf{E}\tau^b \leq \alpha(\lambda, \mu)KC^2$  [Dopper, Gaujal, Vincent, 2006], alors que la taille de l'espace d'état est  $N = C^K$ .

# Critère de Foster

Dans le cas d'un espace d'états  $\mathcal{X}$  dénombrable, une condition suffisante pour l'existence d'une distribution stationnaire.

**Théorème de Foster.** Soit  $\{X_t\}_{t \geq 0}$  une chaîne de Markov avec la matrice de transition  $P$  irréductible. Si il existe un  $\epsilon > 0$ , un ensemble fini  $F \subset \mathcal{X}$  et une fonction  $h : \mathcal{X} \rightarrow \mathbb{R}^+$  telle que :

$$\mathbf{E}[h(X_1) \mid X_0 = i] = \sum_{j \in \mathcal{X}} p_{i,j} h(j) < \infty, \forall i \in F,$$

et

$$\mathbf{E}[h(X_1) \mid X_0 = i] = \sum_{j \in \mathcal{X}} p_{i,j} h(j) \leq h(i) - \epsilon, \forall i \notin F,$$

alors  $\{X_t\}_{t \geq 0}$  est récurrente positive et possède une unique distribution stationnaire.

*Preuve* : Chapitre 5 de Pierre Brémaud. Markov Chains : Gibbs Fields, Monte Carlo Simulation, and Queues, Springer, 2008.



## Lien avec le temps de couplage

Notation :

$$\tau_F = \min\{t \geq 1 : X_t \in F\}.$$

**Lemme.** Sous les mêmes hypothèses que dans le théorème de Foster, pour tout  $i \notin F$ ,

$$\mathbf{E}[\tau_F \mid X_0 = i] \leq \frac{h(i)}{\epsilon}.$$

Dans le cas monotone, on peut obtenir une borne supérieure de temps de couplage en étudiant :

- ▶  $F = \{\perp\}$ , et  $i = \top$  ou
- ▶  $F = \{\top\}$ , et  $i = \perp$ .

Exemple : file avec capacité  $C$  et  $\lambda \neq \mu$  ( $\top = C$ ,  $\perp = 0$ ).

# Distance de variation

## Définition

Soient  $\mu_1$  et  $\mu_2$  deux distributions de probabilités sur un espace au plus dénombrable  $E$ . La distance de variation (totale) entre  $\mu_1$  et  $\mu_2$  est

$$\|\mu_1 - \mu_2\|_{\text{TV}} = \frac{1}{2} \sum_{x \in E} |\mu_1(x) - \mu_2(x)|.$$

# Distance de variation

## Définition

Soient  $\mu_1$  et  $\mu_2$  deux distributions de probabilités sur un espace au plus dénombrable  $E$ . La distance de variation (totale) entre  $\mu_1$  et  $\mu_2$  est

$$\|\mu_1 - \mu_2\|_{\text{TV}} = \frac{1}{2} \sum_{x \in E} |\mu_1(x) - \mu_2(x)|.$$

On remarque que  $0 \leq \|\mu_1 - \mu_2\|_{\text{TV}} \leq 1$ . On a  $\|\mu_1 - \mu_2\|_{\text{TV}} = 0$  si  $\mu_1 = \mu_2$  et  $\|\mu_1 - \mu_2\|_{\text{TV}} = 1$  si  $\mu_1$  et  $\mu_2$  ont des supports disjoints.

# Distance de variation

## Lemme (Définition équivalente)

Pour tout  $A \subseteq E$ , on pose  $\mu_i(A) = \sum_{x \in A} \mu_i(x)$ . Alors

$$\|\mu_1 - \mu_2\|_{\text{TV}} = \max_{A \subseteq E} |\mu_1(A) - \mu_2(A)|.$$

## Distance de variation

### Lemme (Définition équivalente)

Pour tout  $A \subseteq E$ , on pose  $\mu_i(A) = \sum_{x \in A} \mu_i(x)$ . Alors

$$\|\mu_1 - \mu_2\|_{\text{TV}} = \max_{A \subseteq E} |\mu_1(A) - \mu_2(A)|.$$

### Démonstration.

Soient  $E^+$  l'ensemble des éléments de  $E$  tels que  $\mu_1(x) \geq \mu_2(x)$  et  $E^-$  l'ensemble des éléments de  $E$  tels que  $\mu_1(x) < \mu_2(x)$ .

# Distance de variation

## Lemme (Définition équivalente)

Pour tout  $A \subseteq E$ , on pose  $\mu_i(A) = \sum_{x \in A} \mu_i(x)$ . Alors

$$\|\mu_1 - \mu_2\|_{\text{TV}} = \max_{A \subseteq E} |\mu_1(A) - \mu_2(A)|.$$

## Démonstration.

Soient  $E^+$  l'ensemble des éléments de  $E$  tels que  $\mu_1(x) \geq \mu_2(x)$  et  $E^-$  l'ensemble des éléments de  $E$  tels que  $\mu_1(x) < \mu_2(x)$ .

On a

$$\mu_1(E^+) - \mu_2(E^+) = \max_{A \subseteq E} (\mu_1(A) - \mu_2(A))$$

et

$$\mu_2(E^-) - \mu_1(E^-) = \max_{A \subseteq E} (\mu_2(A) - \mu_1(A)).$$

# Distance de variation

## Lemme (Définition équivalente)

Pour tout  $A \subseteq E$ , on pose  $\mu_i(A) = \sum_{x \in A} \mu_i(x)$ . Alors

$$\|\mu_1 - \mu_2\|_{\text{TV}} = \max_{A \subseteq E} |\mu_1(A) - \mu_2(A)|.$$

## Démonstration.

Soient  $E^+$  l'ensemble des éléments de  $E$  tels que  $\mu_1(x) \geq \mu_2(x)$  et  $E^-$  l'ensemble des éléments de  $E$  tels que  $\mu_1(x) < \mu_2(x)$ .

On a

$$\mu_1(E^+) - \mu_2(E^+) = \max_{A \subseteq E} (\mu_1(A) - \mu_2(A))$$

et

$$\mu_2(E^-) - \mu_1(E^-) = \max_{A \subseteq E} (\mu_2(A) - \mu_1(A)).$$

Or  $\mu_1(E^+) - \mu_2(E^+) = \mu_2(E^-) - \mu_1(E^-) \geq 0$ , donc

$$|\mu_1(E^+) - \mu_2(E^+)| + |\mu_1(E^-) - \mu_2(E^-)| = \sum_{x \in E} |\mu_1(x) - \mu_2(x)| = 2\|\mu_1 - \mu_2\|_{\text{TV}}.$$

# Distance de variation

## Exemple (Mélange de cartes)

On dispose d'un jeu de 52 cartes. Une carte est choisie indépendamment et uniformément, puis placée en haut de la pile de cartes. On peut représenter ce processus comme une chaîne de Markov sur les différents ordres des cartes. La distribution stationnaire de cette chaîne est la distribution uniforme sur tous ces ordres :

$$\pi_x = \frac{1}{52} \sum_{y \in N_x} \pi_y,$$

avec  $N_x$  l'ensemble des successeurs de  $x$  :  $|N_x| = 52$ . Si  $\mu_1$  est la distribution uniforme et  $\mu_2$  la distribution uniforme parmi les ordres qui ont l'as de pique au sommet de la pile, on a

$$\|\mu_1 - \mu_2\|_{\text{TV}} \geq 1 - \frac{1}{52} = \frac{51}{52}.$$

En effet, on prend l'ensemble  $A$  des tas ayant l'as de pique au sommet.



# Temps de mélange d'une chaîne de Markov

Soit  $\pi$  la distribution stationnaire d'une chaîne de Markov  $\{X_n\}$  sur l'espace d'états  $E$ . On note  $p_x^n$  la distribution de  $X_n$  sachant que l'état initial est  $x$ . On pose

$$\Delta_x(n) = \|p_x^n - \pi\|_{\text{TV}} \quad \text{et} \quad \Delta(n) = \max_x \Delta_x(n),$$

$$\tau_x(\epsilon) = \min\{n \mid \Delta_x(n) \leq \epsilon\} \quad \text{et} \quad \tau(\epsilon) = \max_{x \in E} \tau_x(\epsilon).$$

On appelle  $\tau(\epsilon)$  le **temps de mélange** de la chaîne de Markov.

# Temps de mélange d'une chaîne de Markov

Soit  $\pi$  la distribution stationnaire d'une chaîne de Markov  $\{X_n\}$  sur l'espace d'états  $E$ . On note  $p_x^n$  la distribution de  $X_n$  sachant que l'état initial est  $x$ . On pose

$$\Delta_x(n) = \|p_x^n - \pi\|_{\text{TV}} \quad \text{et} \quad \Delta(n) = \max_x \Delta_x(n),$$

$$\tau_x(\epsilon) = \min\{n \mid \Delta_x(n) \leq \epsilon\} \quad \text{et} \quad \tau(\epsilon) = \max_{x \in E} \tau_x(\epsilon).$$

On appelle  $\tau(\epsilon)$  le **temps de mélange** de la chaîne de Markov.

On dit qu'une chaîne **mélange rapidement** si  $\tau(\epsilon)$  est polynomial en  $\log(\frac{1}{\epsilon})$  et en la taille du problème.

# Temps de mélange d'une chaîne de Markov

Soit  $\pi$  la distribution stationnaire d'une chaîne de Markov  $\{X_n\}$  sur l'espace d'états  $E$ . On note  $p_x^n$  la distribution de  $X_n$  sachant que l'état initial est  $x$ . On pose

$$\Delta_x(n) = \|p_x^n - \pi\|_{\text{TV}} \quad \text{et} \quad \Delta(n) = \max_x \Delta_x(n),$$

$$\tau_x(\epsilon) = \min\{n \mid \Delta_x(n) \leq \epsilon\} \quad \text{et} \quad \tau(\epsilon) = \max_{x \in E} \tau_x(\epsilon).$$

On appelle  $\tau(\epsilon)$  le **temps de mélange** de la chaîne de Markov.

On dit qu'une chaîne **mélange rapidement** si  $\tau(\epsilon)$  est polynomial en  $\log(\frac{1}{\epsilon})$  et en la taille du problème.

La taille du problème n'est pas la taille de la chaîne de Markov !

Dans le cas du mélange du jeu de cartes, c'est le nombre de cartes.

# Couplage d'une chaîne de Markov

Une technique générale pour borner le temps de mélange.

## Définition

Un couplage d'une chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$  sur un espace d'états  $E$  est une chaîne de Markov  $Z_n = (X_n^1, X_n^2)$  sur  $E \times E$  telle que

$$\begin{cases} \mathbf{P}(X_{n+1}^1 = x' \mid Z_n = (x, y)) &= \mathbf{P}(X_{n+1} = x' \mid X_n = x) \\ \mathbf{P}(X_{n+1}^2 = y' \mid Z_n = (x, y)) &= \mathbf{P}(X_{n+1} = y' \mid X_n = y). \end{cases}$$

# Couplage d'une chaîne de Markov

Une technique générale pour borner le temps de mélange.

## Définition

Un couplage d'une chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$  sur un espace d'états  $E$  est une chaîne de Markov  $Z_n = (X_n^1, X_n^2)$  sur  $E \times E$  telle que

$$\begin{cases} \mathbf{P}(X_{n+1}^1 = x' \mid Z_n = (x, y)) &= \mathbf{P}(X_{n+1} = x' \mid X_n = x) \\ \mathbf{P}(X_{n+1}^2 = y' \mid Z_n = (x, y)) &= \mathbf{P}(X_{n+1} = y' \mid X_n = y). \end{cases}$$

Donc  $(X_n^1)$  et  $(X_n^2)$  se comportent exactement comme  $(X_n)$  en termes de probabilités de transition (elles ont la même matrice de transition). Par contre,  $(X_n^1)$  et  $(X_n^2)$  ne sont pas nécessairement indépendantes.

# Couplage d'une chaîne de Markov

Une technique générale pour borner le temps de mélange.

## Définition

Un couplage d'une chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$  sur un espace d'états  $E$  est une chaîne de Markov  $Z_n = (X_n^1, X_n^2)$  sur  $E \times E$  telle que

$$\begin{cases} \mathbf{P}(X_{n+1}^1 = x' \mid Z_n = (x, y)) &= \mathbf{P}(X_{n+1} = x' \mid X_n = x) \\ \mathbf{P}(X_{n+1}^2 = y' \mid Z_n = (x, y)) &= \mathbf{P}(X_{n+1} = y' \mid X_n = y). \end{cases}$$

Donc  $(X_n^1)$  et  $(X_n^2)$  se comportent exactement comme  $(X_n)$  en termes de probabilités de transition (elles ont la même matrice de transition). Par contre,  $(X_n^1)$  et  $(X_n^2)$  ne sont pas nécessairement indépendantes.

On s'intéresse aux couplages qui

- ▶ mènent à ce que  $(X_n^1)$  et  $(X_n^2)$  se rejoignent ( $\exists n$  tel que  $X_n^1 = X_n^2$ );

# Couplage d'une chaîne de Markov

Une technique générale pour borner le temps de mélange.

## Définition

Un couplage d'une chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$  sur un espace d'états  $E$  est une chaîne de Markov  $Z_n = (X_n^1, X_n^2)$  sur  $E \times E$  telle que

$$\begin{cases} \mathbf{P}(X_{n+1}^1 = x' \mid Z_n = (x, y)) &= \mathbf{P}(X_{n+1} = x' \mid X_n = x) \\ \mathbf{P}(X_{n+1}^2 = y' \mid Z_n = (x, y)) &= \mathbf{P}(X_{n+1} = y' \mid X_n = y). \end{cases}$$

Donc  $(X_n^1)$  et  $(X_n^2)$  se comportent exactement comme  $(X_n)$  en termes de probabilités de transition (elles ont la même matrice de transition). Par contre,  $(X_n^1)$  et  $(X_n^2)$  ne sont pas nécessairement indépendantes.

On s'intéresse aux couplages qui

- ▶ mènent à ce que  $(X_n^1)$  et  $(X_n^2)$  se rejoignent ( $\exists n$  tel que  $X_n^1 = X_n^2$ );
- ▶ une fois atteint ce temps où les chaînes se rejoignent, elles restent égales.

# Couplage d'une chaîne de Markov

## Théorème

Soit  $Z_n = (X_n, Y_n)$  un couplage d'une chaîne de Markov sur  $E$ .

Supposons que  $\exists T$  t.q.  $\forall x, y \in E, \mathbf{P}(X_T \neq Y_T \mid X_0 = x, Y_0 = y) \leq \epsilon$ .

Alors  $\tau(\epsilon) \leq T$ .



# Couplage d'une chaîne de Markov

## Théorème

Soit  $Z_n = (X_n, Y_n)$  un couplage d'une chaîne de Markov sur  $E$ .

Supposons que  $\exists T$  t.q.  $\forall x, y \in E, \mathbf{P}(X_T \neq Y_T \mid X_0 = x, Y_0 = y) \leq \epsilon$ .

Alors  $\tau(\epsilon) \leq T$ .

## Démonstration.

On choisit le couplage où  $Y_0 = \pi$  et  $X_0 = x \in E$ . On a bien

$$\mathbf{P}(X_T \neq Y_T) = \sum_y \mathbf{P}(X_T \neq Y_T \mid Y_0 = y)\pi(y) \leq \sum_y \epsilon\pi(y) \leq \epsilon.$$

# Couplage d'une chaîne de Markov

## Théorème

Soit  $Z_n = (X_n, Y_n)$  un couplage d'une chaîne de Markov sur  $E$ .

Supposons que  $\exists T$  t.q.  $\forall x, y \in E, \mathbf{P}(X_T \neq Y_T \mid X_0 = x, Y_0 = y) \leq \epsilon$ .

Alors  $\tau(\epsilon) \leq T$ .

## Démonstration.

On choisit le couplage où  $Y_0 = \pi$  et  $X_0 = x \in E$ . On a bien

$$\mathbf{P}(X_T \neq Y_T) = \sum_y \mathbf{P}(X_T \neq Y_T \mid Y_0 = y)\pi(y) \leq \sum_y \epsilon\pi(y) \leq \epsilon.$$

Soit  $A \subseteq E$ . On va d'abord borner  $|\mathbf{P}(X_T \in A) - \mathbf{P}(Y_T \in A)|$ .

# Couplage d'une chaîne de Markov

## Théorème

Soit  $Z_n = (X_n, Y_n)$  un couplage d'une chaîne de Markov sur  $E$ .

Supposons que  $\exists T$  t.q.  $\forall x, y \in E, \mathbf{P}(X_T \neq Y_T \mid X_0 = x, Y_0 = y) \leq \epsilon$ .

Alors  $\tau(\epsilon) \leq T$ .

## Démonstration.

On choisit le couplage où  $Y_0 = \pi$  et  $X_0 = x \in E$ . On a bien

$$\mathbf{P}(X_T \neq Y_T) = \sum_y \mathbf{P}(X_T \neq Y_T \mid Y_0 = y)\pi(y) \leq \sum_y \epsilon\pi(y) \leq \epsilon.$$

Soit  $A \subseteq E$ . On va d'abord borner  $|\mathbf{P}(X_T \in A) - \mathbf{P}(Y_T \in A)|$ .

$$\mathbf{P}(X_T \in A) \geq \mathbf{P}(X_T = Y_T \text{ et } Y_T \in A)$$

# Couplage d'une chaîne de Markov

## Théorème

Soit  $Z_n = (X_n, Y_n)$  un couplage d'une chaîne de Markov sur  $E$ .

Supposons que  $\exists T$  t.q.  $\forall x, y \in E, \mathbf{P}(X_T \neq Y_T \mid X_0 = x, Y_0 = y) \leq \epsilon$ .

Alors  $\tau(\epsilon) \leq T$ .

## Démonstration.

On choisit le couplage où  $Y_0 = \pi$  et  $X_0 = x \in E$ . On a bien

$$\mathbf{P}(X_T \neq Y_T) = \sum_y \mathbf{P}(X_T \neq Y_T \mid Y_0 = y)\pi(y) \leq \sum_y \epsilon\pi(y) \leq \epsilon.$$

Soit  $A \subseteq E$ . On va d'abord borner  $|\mathbf{P}(X_T \in A) - \mathbf{P}(Y_T \in A)|$ .

$$\begin{aligned} \mathbf{P}(X_T \in A) &\geq \mathbf{P}(X_T = Y_T \text{ et } Y_T \in A) \\ &\geq 1 - \mathbf{P}(X_T \neq Y_T \text{ ou } Y_T \notin A) \end{aligned}$$

# Couplage d'une chaîne de Markov

## Théorème

Soit  $Z_n = (X_n, Y_n)$  un couplage d'une chaîne de Markov sur  $E$ .

Supposons que  $\exists T$  t.q.  $\forall x, y \in E, \mathbf{P}(X_T \neq Y_T \mid X_0 = x, Y_0 = y) \leq \epsilon$ .

Alors  $\tau(\epsilon) \leq T$ .

## Démonstration.

On choisit le couplage où  $Y_0 = \pi$  et  $X_0 = x \in E$ . On a bien

$$\mathbf{P}(X_T \neq Y_T) = \sum_y \mathbf{P}(X_T \neq Y_T \mid Y_0 = y)\pi(y) \leq \sum_y \epsilon\pi(y) \leq \epsilon.$$

Soit  $A \subseteq E$ . On va d'abord borner  $|\mathbf{P}(X_T \in A) - \mathbf{P}(Y_T \in A)|$ .

$$\begin{aligned} \mathbf{P}(X_T \in A) &\geq \mathbf{P}(X_T = Y_T \text{ et } Y_T \in A) \\ &\geq 1 - \mathbf{P}(X_T \neq Y_T \text{ ou } Y_T \notin A) \\ &\geq 1 - \mathbf{P}(X_T \neq Y_T) - \mathbf{P}(Y_T \notin A) \end{aligned}$$

# Couplage d'une chaîne de Markov

## Théorème

Soit  $Z_n = (X_n, Y_n)$  un couplage d'une chaîne de Markov sur  $E$ .

Supposons que  $\exists T$  t.q.  $\forall x, y \in E, \mathbf{P}(X_T \neq Y_T \mid X_0 = x, Y_0 = y) \leq \epsilon$ .

Alors  $\tau(\epsilon) \leq T$ .

## Démonstration.

On choisit le couplage où  $Y_0 = \pi$  et  $X_0 = x \in E$ . On a bien

$$\mathbf{P}(X_T \neq Y_T) = \sum_y \mathbf{P}(X_T \neq Y_T \mid Y_0 = y)\pi(y) \leq \sum_y \epsilon\pi(y) \leq \epsilon.$$

Soit  $A \subseteq E$ . On va d'abord borner  $|\mathbf{P}(X_T \in A) - \mathbf{P}(Y_T \in A)|$ .

$$\begin{aligned} \mathbf{P}(X_T \in A) &\geq \mathbf{P}(X_T = Y_T \text{ et } Y_T \in A) \\ &\geq 1 - \mathbf{P}(X_T \neq Y_T \text{ ou } Y_T \notin A) \\ &\geq 1 - \mathbf{P}(X_T \neq Y_T) - \mathbf{P}(Y_T \notin A) \\ &\geq \mathbf{P}(Y_T \in A) - \epsilon = \pi(A) - \epsilon. \end{aligned}$$

# Couplage d'une chaîne de Markov

## Théorème

Soit  $Z_n = (X_n, Y_n)$  un couplage d'une chaîne de Markov sur  $E$ .

Supposons que  $\exists T$  t.q.  $\forall x, y \in E, \mathbf{P}(X_T \neq Y_T \mid X_0 = x, Y_0 = y) \leq \epsilon$ .

Alors  $\tau(\epsilon) \leq T$ .

## Démonstration.

On choisit le couplage où  $Y_0 = \pi$  et  $X_0 = x \in E$ . On a bien

$$\mathbf{P}(X_T \neq Y_T) = \sum_y \mathbf{P}(X_T \neq Y_T \mid Y_0 = y)\pi(y) \leq \sum_y \epsilon\pi(y) \leq \epsilon.$$

Soit  $A \subseteq E$ . On va d'abord borner  $|\mathbf{P}(X_T \in A) - \mathbf{P}(Y_T \in A)|$ .

$$\begin{aligned} \mathbf{P}(X_T \in A) &\geq \mathbf{P}(X_T = Y_T \text{ et } Y_T \in A) \\ &\geq 1 - \mathbf{P}(X_T \neq Y_T \text{ ou } Y_T \notin A) \\ &\geq 1 - \mathbf{P}(X_T \neq Y_T) - \mathbf{P}(Y_T \notin A) \\ &\geq \mathbf{P}(Y_T \in A) - \epsilon = \pi(A) - \epsilon. \end{aligned}$$

De même, si on remplace  $A$  par son complémentaire, on a

$$\mathbf{P}(X_T \notin A) \geq \mathbf{P}(Y_T \notin A) - \epsilon \text{ et } \mathbf{P}(X_T \in A) \leq \mathbf{P}(Y_T \in A) + \epsilon = \pi(A) + \epsilon.$$

# Couplage d'une chaîne de Markov

## Théorème

Soit  $Z_n = (X_n, Y_n)$  un couplage d'une chaîne de Markov sur  $E$ .

Supposons que  $\exists T$  t.q.  $\forall x, y \in E, \mathbf{P}(X_T \neq Y_T \mid X_0 = x, Y_0 = y) \leq \epsilon$ .

Alors  $\tau(\epsilon) \leq T$ .

## Démonstration.

On choisit le couplage où  $Y_0 = \pi$  et  $X_0 = x \in E$ . On a bien

$$\mathbf{P}(X_T \neq Y_T) = \sum_y \mathbf{P}(X_T \neq Y_T \mid Y_0 = y)\pi(y) \leq \sum_y \epsilon\pi(y) \leq \epsilon.$$

Soit  $A \subseteq E$ . On va d'abord borner  $|\mathbf{P}(X_T \in A) - \mathbf{P}(Y_T \in A)|$ .

$$\begin{aligned} \mathbf{P}(X_T \in A) &\geq \mathbf{P}(X_T = Y_T \text{ et } Y_T \in A) \\ &\geq 1 - \mathbf{P}(X_T \neq Y_T \text{ ou } Y_T \notin A) \\ &\geq 1 - \mathbf{P}(X_T \neq Y_T) - \mathbf{P}(Y_T \notin A) \\ &\geq \mathbf{P}(Y_T \in A) - \epsilon = \pi(A) - \epsilon. \end{aligned}$$

De même, si on remplace  $A$  par son complémentaire, on a

$\mathbf{P}(X_T \notin A) \geq \mathbf{P}(Y_T \notin A) - \epsilon$  et  $\mathbf{P}(X_T \in A) \leq \mathbf{P}(Y_T \in A) + \epsilon = \pi(A) + \epsilon$ . Donc

$$\max_A |\rho_x^n(A) - \pi(A)| \leq \epsilon.$$



## Exemple : Mélange de cartes

On souhaite construire un couplage *rapide*. On dispose de deux tas de cartes, pour chaque chaîne du couplage.

**1er essai.** On choisit des évolutions indépendantes. On a vu que les chaînes allaient coupler presque sûrement, mais l'évolution ensuite n'est pas la même. De plus, ce temps de couplage n'est pas efficace.

## Exemple : Mélange de cartes

On souhaite construire un couplage *rapide*. On dispose de deux tas de cartes, pour chaque chaîne du couplage.

**1er essai.** On choisit des évolutions indépendantes. On a vu que les chaînes allaient coupler presque sûrement, mais l'évolution ensuite n'est pas la même. De plus, ce temps de couplage n'est pas efficace.

**2ème essai.** On choisit une position  $j$  uniformément, et on met la  $j$ -ème carte au-dessus. Ce n'est pas efficace : la première carte est différente si les cartes en  $j$ -ème position sont différentes. Il ne va jamais y avoir couplage dès que les paquets initiaux sont différents.

## Exemple : Mélange de cartes

On souhaite construire un couplage *rapide*. On dispose de deux tas de cartes, pour chaque chaîne du couplage.

**1er essai.** On choisit des évolutions indépendantes. On a vu que les chaînes allaient coupler presque sûrement, mais l'évolution ensuite n'est pas la même. De plus, ce temps de couplage n'est pas efficace.

**2ème essai.** On choisit une position  $j$  uniformément, et on met la  $j$ -ème carte au-dessus. Ce n'est pas efficace : la première carte est différente si les cartes en  $j$ -ème position sont différentes. Il ne va jamais y avoir couplage dès que les paquets initiaux sont différents.

**3ème essai.** Pour la première chaîne, on choisit une position  $j$  uniformément. Soit  $c$  la valeur de la carte tirée. On met la carte qui a cette valeur sur le haut de la pile aussi pour la deuxième chaîne. On a bien un couplage puisque la valeur  $c$  est aussi choisie uniformément.

## Exemple : Mélange de cartes

Une fois la carte choisie et placée en haut du tas, cette carte va garder la même position dans les deux tas.

## Exemple : Mélange de cartes

Une fois la carte choisie et placée en haut du tas, cette carte va garder la même position dans les deux tas.

Il suffit donc d'avoir choisi toutes les cartes au moins une fois pour avoir le même ordre de cartes : ce problème se ramène à celui du [collecteur de coupons](#).

## Exemple : Mélange de cartes

Une fois la carte choisie et placée en haut du tas, cette carte va garder la même position dans les deux tas.

Il suffit donc d'avoir choisi toutes les cartes au moins une fois pour avoir le même ordre de cartes : ce problème se ramène à celui du [collecteur de coupons](#).

La probabilité qu'une carte  $c$  n'ait pas encore été choisie après  $n \ln n + cn$  étapes est

$$\left(1 - \frac{1}{n}\right)^{n \ln n + cn} \leq e^{-(\ln n + c)} = \frac{e^{-c}}{n}.$$

## Exemple : Mélange de cartes

Une fois la carte choisie et placée en haut du tas, cette carte va garder la même position dans les deux tas.

Il suffit donc d'avoir choisi toutes les cartes au moins une fois pour avoir le même ordre de cartes : ce problème se ramène à celui du [collecteur de coupons](#).

La probabilité qu'une carte  $c$  n'ait pas encore été choisie après  $n \ln n + cn$  étapes est

$$\left(1 - \frac{1}{n}\right)^{n \ln n + cn} \leq e^{-(\ln n + c)n} = \frac{e^{-c}}{n}.$$

Donc la probabilité qu'une carte quelconque n'est pas encore été choisie (et que les tas soient donc différents) est inférieure à  $e^{-c}$ . On peut donc choisir  $\epsilon = e^{-c}$  et le temps de mélange est

$$\tau(\epsilon) \leq n \ln n + n \ln \frac{1}{\epsilon}.$$

La chaîne mélange donc rapidement.

# Plan

Système à événements discrets

Echantillonnage de la distribution stationnaire

Temps de couplage

Annexe : Distance de variation et couplage de chaînes de Markov

**Annexe : Critère de Foster**

Annexe : Rappels et table d'alias



# Critère de Foster

## Théorème (Foster ; 1953)

Soit  $\{X_n\}$  une CMH irréductible, avec l'espace d'états  $\mathcal{E}$  dénombrable et la matrice de transition  $\mathbf{P}$ .

Supposons qu'il existe une fonction  $h : \mathcal{E} \rightarrow \mathbb{R}$  t.q.  $\inf_i h(i) > -\infty$ , un sous-ensemble fini  $F \subset \mathcal{E}$  et un  $\epsilon > 0$  tels que

$$\sum_{k \in \mathcal{E}} p_{ik} h(k) < \infty, \quad \forall i \in F,$$

$$\sum_{k \in \mathcal{E}} p_{ik} h(k) \leq h(i) - \epsilon, \quad \forall i \notin F.$$

Alors CMH est récurrente positive.

# Critère de Foster

## Lemme

Soit  $\{X_n\}$  une CMH irréductible, avec l'espace d'états  $\mathcal{E}$  dénombrable et la matrice de transition  $\mathbf{P}$ .

Soit  $F \subset \mathcal{E}$  un sous-ensemble fini, et  $\tau(F)$  le temps de retour dans  $F$ . Si  $\mathbf{E}_j(\tau(F)) < \infty, \forall j \in F$ , alors la chaîne est récurrente positive.

*Démonstration.* Soit  $i \in F$  et  $T_i$  le temps de retour à  $i$ . Notons  $\tau_1 = \tau(F), \tau_2, \dots$  les temps de retour successifs dans  $F$ .

Propriété de Markov forte implique que  $\{Y_n\}_{n \geq 0}$ , où  $Y_0 = X_0 = i$  et  $Y_n = X_{\tau_n}, n \geq 1$  est une CMH, avec l'espace d'états  $F$ .

CMH  $\{Y_n\}$  est

- ▶ irréductible, car  $\{X_n\}$  est irréductible ;
- ▶ récurrente positive, car  $F$  est fini ; en particulier,  $\mathbf{E}_i(\tilde{T}_i) < \infty$ , où  $\tilde{T}_i$  est le temps de retour à  $i$  pour la chaîne  $\{Y_n\}$ .

## Démonstration du lemme

On pose  $S_0 = \tau_1$ ,  $S_k = \tau_{k+1} - \tau_k$ ,  $k \geq 1$ . Alors  $T_i = \sum_{k=0}^{\infty} S_k \mathbf{1}_{k < \tilde{T}_i}$ , et

$$\mathbf{E}_i(T_i) = \sum_{k=0}^{\infty} \mathbf{E}_i(S_k \mathbf{1}_{k < \tilde{T}_i}).$$

On a  $\mathbf{E}_i(S_k \mathbf{1}_{k < \tilde{T}_i}) = \sum_{\ell \in F} \mathbf{E}_i(S_k \mathbf{1}_{k < \tilde{T}_i} \mathbf{1}_{X_{\tau_k} = \ell})$ .

Par la propriété de Markov forte pour  $\{X_n\}$  et le temps d'arrêt  $\tau_k$  ( $\{k < \tilde{T}_i\}$  est dans le passé de  $\{X_n\}$  au temps  $\tau_k$ ),

$$\begin{aligned} \mathbf{E}_i(S_k \mathbf{1}_{k < \tilde{T}_i} \mathbf{1}_{X_{\tau_k} = \ell}) &= \mathbf{E}_i(S_k \mid k < \tilde{T}_i, X_{\tau_k} = \ell) \mathbf{P}(k < \tilde{T}_i, X_{\tau_k} = \ell) \\ &= \mathbf{E}_i(S_k \mid X_{\tau_k} = \ell) \mathbf{P}(k < \tilde{T}_i, X_{\tau_k} = \ell) \end{aligned}$$

On a  $\mathbf{E}_i(S_k \mid X_{\tau_k} = \ell) = \mathbf{E}_\ell(\tau(F))$ , donc

$$\mathbf{E}_i(S_k \mathbf{1}_{k < \tilde{T}_i}) \leq (\max_{\ell \in F} \mathbf{E}_\ell(\tau(F))) \mathbf{P}(k < \tilde{T}_i) \text{ et}$$

$$\mathbf{E}_i(T_i) \leq (\max_{\ell \in F} \mathbf{E}_\ell(\tau(F))) \sum_{k=0}^{\infty} \mathbf{P}(k < \tilde{T}_i) = (\max_{\ell \in F} \mathbf{E}_\ell(\tau(F))) \mathbf{E}_i(\tilde{T}_i) < \infty.$$

## Démonstration du théorème

On suppose que  $h \geq 0$  (sans perte de généralité, en rajoutant une constante).

Soit  $\tau$  le temps de retour dans  $F$ , et  $Y_n = h(X_n)\mathbf{1}_{n < \tau}$ .

On a  $\mathbf{E}[h(X_{n+1}) | X_n = i] \leq h(i) - \epsilon$ ,  $\forall i \notin F$ . Pour  $i \notin F$ ,

$$\begin{aligned}\mathbf{E}_i[Y_{n+1} | X_0^n] &= \mathbf{E}_i[Y_{n+1}\mathbf{1}_{n < \tau} | X_0^n] + \mathbf{E}_i[Y_{n+1}\mathbf{1}_{n \geq \tau} | X_0^n] \\ &= \mathbf{E}_i[Y_{n+1}\mathbf{1}_{n < \tau} | X_0^n] \\ &\leq \mathbf{E}_i[h(X_{n+1})\mathbf{1}_{n < \tau} | X_0^n] \\ &= \mathbf{1}_{n < \tau} \mathbf{E}_i[h(X_{n+1}) | X_0^n] = \mathbf{1}_{n < \tau} \mathbf{E}_i[h(X_{n+1}) | X_n] \\ &\leq \mathbf{1}_{n < \tau} h(X_n) - \epsilon \mathbf{1}_{n < \tau} = Y_n - \epsilon \mathbf{1}_{n < \tau}\end{aligned}$$

En prenant l'esperance,

$$\mathbf{E}_i[Y_{n+1}] \leq \mathbf{E}_i[Y_n] - \epsilon \mathbf{P}_i(\tau > n)$$

et

$$0 \leq \mathbf{E}_i[Y_{n+1}] \leq \mathbf{E}_i[Y_0] - \epsilon \sum_{k=0}^n \mathbf{P}_i(\tau > k).$$

# Démonstration du théorème

On a  $Y_0 = h(i)$  et  $\sum_{k=0}^n \mathbf{P}_i(\tau > k) = \mathbf{E}_i[\tau]$ . Donc pour tout  $i \notin F$ ,

$$\mathbf{E}_i[\tau] \leq \epsilon^{-1} h(i).$$

Pour  $j \in F$ , méthode d'un pas en avant donne

$$\mathbf{E}_j[\tau] = 1 + \sum_{i \notin F} p_{ji} \mathbf{E}_i[\tau].$$

Donc

$$\mathbf{E}_j[\tau] \leq 1 + \epsilon^{-1} \sum_{i \notin F} p_{ji} h(i) < \infty$$

par hypothèse du théorème pour  $j \in F$ . On peut donc appliquer le lemme. □

## Corollaire (Pakes, 1969)

Soit  $\{X_n\}$  une CMH irréductible, avec l'espace d'états  $\mathcal{E} = \mathbb{N}$  telle que pour tout  $n \geq 0$  et  $i \in \mathcal{E}$ ,

$$\mathbf{E}[X_{n+1} - X_n \mid X_n = i] < \infty$$

et

$$\limsup_{i \rightarrow \infty} \mathbf{E}[X_{n+1} - X_n \mid X_n = i] < 0.$$

Alors  $\{X_n\}$  est récurrente positive.

*Démonstration.* Soit  $\limsup_{i \rightarrow \infty} \mathbf{E}[X_{n+1} - X_n \mid X_n = i] = -2\epsilon$ . Alors  $\epsilon > 0$ . Alors il existe  $i_0$  tel que pour tout  $i > i_0$ ,  $\mathbf{E}[X_{n+1} - X_n \mid X_n = i] < -\epsilon$ . Le corollaire suit du théorème de Foster pour  $h(i) = i$  et  $F = \{i; i \leq i_0\}$ . □

# Exemple

## Exemple (Marche aléatoire sur $\mathbb{N}$ )

Soit  $\{Z_n\}$  une suite i.i.d. des v.a. dans  $\mathbb{Z}$  intégrables et telles que  $\mathbf{E}[Z_1] < 0$ .

Soit

$$X_{n+1} = (X_n + Z_{n+1})^+,$$

et  $X_0$  indépendant de  $\{Z_n\}$ . Si  $\{X_n\}$  est irréductible, alors

$$\begin{aligned}\mathbf{E}[X_{n+1} - i \mid X_n = i] &= \mathbf{E}[(i + Z_{n+1})^+ - i] \\ &= \mathbf{E}[-i\mathbf{1}_{Z_{n+1} \leq -i} + Z_{n+1}\mathbf{1}_{Z_{n+1} > -i}] \leq \mathbf{E}[Z_1\mathbf{1}_{Z_1 > -i}]\end{aligned}$$

Par la convergence monotone,  $\lim_{i \rightarrow \infty} \mathbf{E}[Z_1\mathbf{1}_{Z_1 > -i}] = \mathbf{E}[Z_1] < 0$  et par le corollaire précédent, MCH  $\{X_n\}$  est récurrente positive.

# Critère négatif

## Théorème

Soit  $\{X_n\}$  une CMH irréductible, avec l'espace d'états  $\mathcal{E}$  dénombrable et la matrice de transition  $\mathbf{P}$ .

Supposons qu'il existe une fonction  $h : \mathcal{E} \rightarrow \mathbb{R}$  t.q.  $\inf_i h(i) > -\infty$ , un sous-ensemble fini  $F \subset \mathcal{E}$  tels que

$$\exists j \notin F \text{ tel que } h(j) > \max_{i \in F} h(i)$$

$$\sup_{i \in \mathcal{E}} \sum_{k \in \mathcal{E}} p_{ik} |h(k) - h(i)| < \infty,$$

$$\sum_{k \in \mathcal{E}} p_{ik} (h(k) - h(i)) \geq 0, \quad \forall i \notin F.$$

Alors CMH ne peut pas être récurrente positive.

**Exemple :** processus de naissance et de mort sur  $\mathbb{N}$ , avec  $p_i = p, \forall i$ ,  $q_i = q, i > 1$  et  $p \geq q$ .



# Démonstration

Soit  $\tau$  le temps de retour à  $F$ . On a

$h(X_\tau)\mathbf{1}_{\tau < \infty} = h(X_0) + \sum_{n=0}^{\infty} (h(X_{n+1}) - h(X_n))\mathbf{1}_{\tau > n}$ . Alors pour  $j \notin F$ ,

$$\begin{aligned}\sum_{n=0}^{\infty} \mathbf{E}_j[h(X_{n+1}) - h(X_n)|\mathbf{1}_{\tau > n}] &= \sum_{n=0}^{\infty} \mathbf{E}_j[\mathbf{E}_j[|h(X_{n+1}) - h(X_n)| \mid X_n^0]\mathbf{1}_{\tau > n}] \\ &= \sum_{n=0}^{\infty} \mathbf{E}_j[\mathbf{E}_j[|h(X_{n+1}) - h(X_n)| \mid X_n]\mathbf{1}_{\tau > n}] \\ &\leq K \sum_{n=0}^{\infty} \mathbf{P}(\tau > n).\end{aligned}$$

pour un  $K$  positif, par la condition 2. Si la chaîne était récurrente positive, alors  $\sum_{n=0}^{\infty} \mathbf{P}(\tau > n) = \mathbf{E}_j[\tau]$  et

$$\begin{aligned}\mathbf{E}_j[h(X_\tau)] &= \mathbf{E}_j[h(X_\tau)\mathbf{1}_{\tau < \infty}] = h(j) + \sum_{n=0}^{\infty} \mathbf{E}_j[(h(X_{n+1}) - h(X_n))\mathbf{1}_{\tau > n}] \\ &> h(j) > \max_{i \in F} h(i) \geq \mathbf{E}_j[h(X_\tau)],\end{aligned}$$

ce qui est une contradiction. □

# Protocole Aloha

Protocole utilisé pour la communication satellite.

Chaque utilisateur envoie un message via un seul canal de communication. Si plusieurs messages sont envoyés au même temps, il y a interférence et aucun message n'est transmis.

Les utilisateurs détectent s'il y a eu l'interférence, et dans ce cas, retransmettent le même message plus tard.

Protocole Aloha en temps discret

- ▶ Les messages sont envoyés à des temps discret ; la durée de transmission du message est plus petite que la durée du pas de temps ;
- ▶ Tous les messages **en attente** re-essaient avec une probabilité  $\nu \in (0, 1)$ , indépendamment les uns des autres (retransmission Bernoulli).
- ▶ Les **nouveau** messages transmettent toujours.

# Protocole Aloha

- ▶ Soit  $X_n$  le nombre de messages en attente au temps  $n$ .
- ▶ Notons par  $b_i(k)$  la probabilité que  $i$  messages re-essaient si  $X_n = k$ ,

$$b_i(k) = \binom{k}{i} \nu^i (1 - \nu)^{k-i}.$$

- ▶ Soit  $A_n$  le nombre de nouveaux messages au temps  $n$ . On suppose  $\{A_n\}$  i.i.d. avec  $\mathbf{P}(A_n = j) = a_j$ .  
 $\lambda = \mathbf{E}[A_n]$  est appelée l'intensité du trafic.
- ▶  $\{X_n\}$  est une CMH.

## Questions

- ▶ Quelle est sa matrice de transition ?
- ▶ Sous quelles conditions  $\{X_n\}$  est une CMH irréductible ?
- ▶ Est-elle récurrente positive ?

# Aloha modifié

Variante : probabilité de re-transmission  $\nu(k)$  si  $X_n = k$ .

**Question** : sous quelles conditions peut on stabiliser ce protocole ?

# Aloha modifié

Variante : probabilité de re-transmission  $\nu(k)$  si  $X_n = k$ .

**Question** : sous quelles conditions peut on stabiliser ce protocole ?

- ▶ Selon le lemme de Pakes, il suffit de trouver une fonction  $\nu$  telle que

$$\lambda \leq \lim_{i \rightarrow \infty} (b_1(i)a_0 + b_0(i)a_1) - \epsilon,$$

pour un  $\epsilon > 0$ .

- ▶ On va étudier la fonction

$$g_k(\nu) = (1 - \nu)^k a_1 + k\nu(1 - \nu)^{k-1} a_0,$$

car il suffit donc de montrer que  $\lambda \leq g_i(\nu(i)) - \epsilon$ . Pour  $k \geq 2$ ,

$$g'_k(\nu) = k(1 - \nu)^{k-2} [(a_0 - a_1) - \nu(ka_0 - a_1)].$$

## Aloha modifié

- ▶ On suppose que  $a_0 > a_1$ . Dans ce cas, pour  $k \geq 2$ , la dérivée est nulle pour

$$\nu(k) = \frac{a_0 - a_1}{ka_0 - a_1},$$

et on a la valeur maximale pour  $g_k$

$$g_k(\nu(k)) = a_0 \left( \frac{k-1}{k - a_1/a_0} \right)^{k-1}.$$

Donc  $\lim_{k \rightarrow \infty} g_k(\nu(k)) = a_0 \exp\left(\frac{a_1}{a_0} - 1\right)$ , et donc une condition suffisante pour la stabilité est

$$\lambda < a_0 \exp\left(\frac{a_1}{a_0} - 1\right).$$

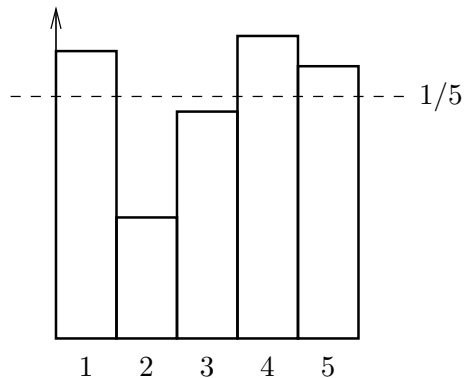
Par exemple, si on prend les arrivées qui suivent une loi de Poisson  $a_i = e^{-\lambda} \frac{\lambda^i}{i!}$ , cette condition devient  $\lambda < e^{-1}$ .

Notons que nous avons alors  $a_0 > a_1$ . On peut montrer que si  $a_0 \leq a_1$ , alors le CM ne peut pas être récurrente positive.



# Échantillonnage par table d'alias (Walker 1977)

← retour

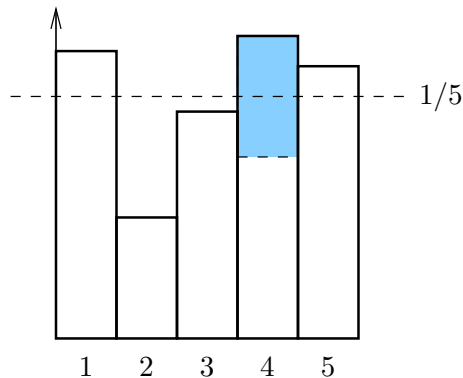


On commence par construire un histogramme de  $\mu$ .



# Échantillonnage par table d'alias (Walker 1977)

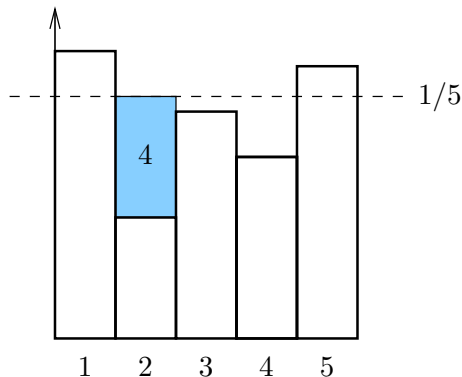
← retour



On remplit la colonne la plus basse jusqu'au niveau  $1/N$  en puisant dans la colonne la plus haute.

# Échantillonnage par table d'alias (Walker 1977)

← retour

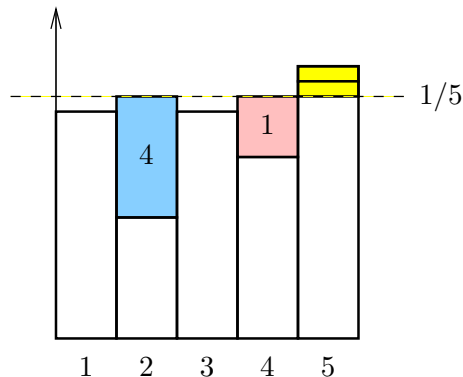


On remplit la colonne la plus basse jusqu'au niveau  $1/N$  en puisant dans la colonne la plus haute.



# Échantillonnage par table d'alias (Walker 1977)

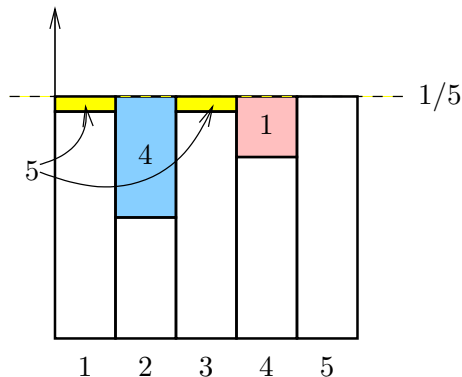
← retour



On recommence tant que toutes les colonnes ne sont pas exactement au niveau  $1/N$ . Au-delà du seuil  $S[i]$ , la colonne  $i$  correspond à la valeur  $V[i]$ .

# Échantillonnage par table d'alias (Walker 1977)

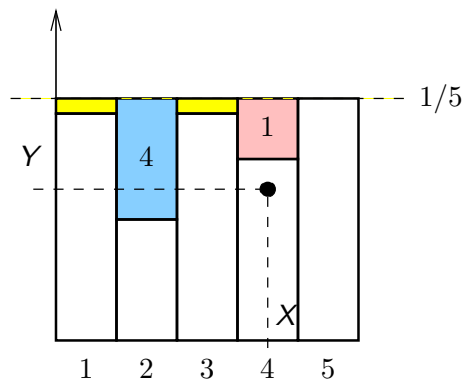
← retour



On recommence tant que toutes les colonnes ne sont pas exactement au niveau  $1/N$ . Au-delà du seuil  $S[i]$ , la colonne  $i$  correspond à la valeur  $V[i]$ .

# Échantillonnage par table d'alias (Walker 1977)

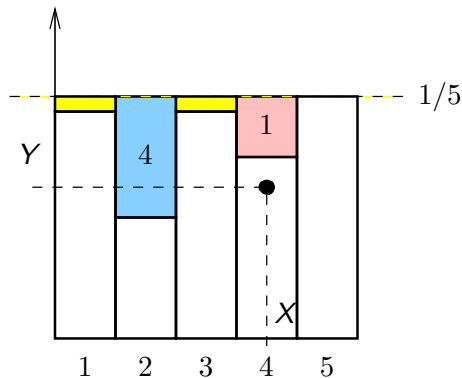
← retour



On tire  $X$  et  $Y$  deux variables uniformes, indépendantes :  $X$  dans  $\{1, \dots, N\}$ ,  $Y$  dans  $[0, 1/N]$ . Par construction des tables  $S$  et  $V$ , la quantité  $X\mathbf{1}_{Y < S[X]} + V[X]\mathbf{1}_{Y \geq S[X]}$  est distribuée selon  $\mu$ .

# Échantillonnage par table d'alias (Walker 1977)

← retour



Une fois les tableaux  $S$  et  $V$  construits, le tirage se fait en temps constant.