

# MIPS: Processus stochastiques

Pr. Lacroix

SEATECH  
Université de Toulon

Promo 2018

## 1 Introduction

- Définitions

## 2 Processus stationnaires

- Définitions
- Théorème ergodique
- Trajectoires

## 3 Markov

- Noyau et chaîne
- Mesure invariante
- Classification des états
- Existence de lois
- Convergence
- Vitesse
- Pagerank

- Hastings-Metropolis
- Mesure de Gibbs
- Recuit simulé
- Voyageur de commerce

## 4 Martingales

- Définitions
- Convergence
- Temps d'arrêt
- Robbins-Monro

## 5 Branchement

- Galton-Watson
- Fctn génératrice
- Géométrie
- Cas général
- Asymptotique

Certains phénomènes aléatoires évoluent au cours du temps :

- la cote de popularité d'un politique ;
- le taux de globules rouges d'un individu ;
- la fortune d'un joueur au casino ;
- le classement d'une grande école d'ingénieurs
- ...

La théorie des probabilités permet de modéliser et de comprendre dans certains cas ces fluctuations temporelles. C'est l'objet de l'étude des processus stochastiques.

## Définition

*Soit  $(T, \mathcal{T})$  un espace mesurable et  $(\Omega, \mathcal{B}, \mu)$  un espace probabilisé. Un processus stochastique indexé par  $T$  est une application mesurable*

$$X : \Omega \times T \rightarrow E, (\omega, t) \mapsto X_t(\omega) := X(\omega, t),$$

*où  $(E, \mathcal{E})$  est un troisième espace mesurable.*

*On appelle  $E$  l'espace des états.*

- *Si  $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}_1)$ , on parle de processus réel.*
- *Si  $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}_n)$ , on parle de processus multidimensionnel.*
- *Si  $E \subset \mathbb{Z}$ , on parle de processus à espace d'états discret.*

## Définition

*Pour  $\omega \in \Omega$  fixé, on appelle  $(X_t(\omega))_{t \in T}$  la trajectoire de  $\omega$ .*

*On parle de processus continu si  $T = \mathbb{R}$ , discret si  $T \subset \mathbb{Z}$  est infini, et on parle de famille à un paramètre si  $T$  est fini.*

*Pour  $t \in T$  fixé,  $X_t$  est une VA.*

On interprète souvent  $t$  comme le temps mais ce n'est pas systématique. On parlera de séries chronologiques si  $t$  prend des valeurs discrètes équidistantes.

## Définition

*La loi du processus est la donnée de toutes les marginales finies*

$\mathbb{P}_{(X_{t_1}, \dots, X_{t_p})}$ ,  $t_i \in T$ ,  $p \geq 1$ .

*Si  $T \subset \mathbb{R}$ , l'accroissement du processus sur l'intervalle  $[t_i, t_j[$  est  $X_{t_j} - X_{t_i}$  si  $t_i < t_j$ .*

### Définition (Processus équivalents)

*On dit que deux processus  $X$  et  $\tilde{X}$  sont équivalents si leurs lois sont identiques.*

### Définition (Moments)

*On suppose que  $X_t \in \mathcal{L}^2$  pour chaque  $t \in T$ . Alors la fonction de covariance du processus est définie par*

$$C(t, s) = \text{Cov}(X_t, X_s), \quad t, s \in T.$$

*La fonction d'autocorrélation est définie par*

$$\rho(t, s) = \frac{\text{Cov}(t, s)}{\sigma_t \sigma_s}, \quad s, t \in T,$$

*où  $\sigma_t$  est l'écart type de  $X_t$ .*

## Définition (Continuités)

*Un processus est continu en probabilités si quel que soit  $\varepsilon > 0$ ,*

$$\mathbb{P}(|X_{t+h} - X_t| \geq \varepsilon) \rightarrow 0 \text{ si } h \rightarrow 0.$$

*Un processus aléatoire est continu en moyenne quadratique si*

$$\mathbb{E}((X_{t+h} - X_t)^2) \rightarrow 0 \text{ ; si } h \rightarrow 0.$$

*La continuité en moyenne quadratique implique la continuité de la moyenne  $t \mapsto \mathbb{E}(X_t)$ , de la variance, et de la covariance.*

Ces deux conditions n'impliquent pas la continuité des trajectoires.



## Définition (Accroissements indépendants)

Un processus est  $(X_t)_{t \in T}$  avec  $T \subset \mathbb{R}$  est dit à **accroissements indépendants** si quels que soient  $t_1 < t_2 < \dots < t_k$ , les variables  $U_i = X_{t_{i+1}} - X_{t_i}$  sont indépendantes.

Un exemple typique est le cas où  $T = \mathbb{N}$  et

$X_n = X_0 + Y_1 + \dots + Y_n$ , avec  $(Y_n)_{n \geq 1}$  une suite indépendante.

En temps continu, le plus célèbre est le processus de

**Wiener-Lévy** : c'est un processus  $(X_t)_{t \geq 0}$  à accroissements indépendants tel que :

- si  $s < t$ ,  $\mathbb{P}_{X_t - X_s} = \mathcal{N}(0, t - s)$  ;
- $\mathbb{P}_{X_t - X_s} = \mathbb{P}_{X_{t-s}}$  (stationnarité) ;
- $t \mapsto X_t$  est presque sûrement continue.

Ces processus interviennent en calcul stochastique (intégrale d'Ito, EDPS, EDPSR).

### Définition (Espace des trajectoires)

*Si  $(X_t)_{t \in T}$  est un processus, l'espace des trajectoires est défini comme l'ensemble  $\mathcal{E} = \{(X_t(\omega))_{t \in T} : \omega \in \Omega\}$ . Il est muni de la plus petite tribu rendant mesurables toutes les applications  $(X_t(\omega))_{t \in T} \mapsto (X_t(\omega))_{t \in I}$  pour  $I$  fini  $\subset T$ . La loi du processus est alors la loi image de  $\mathbb{P}$  par  $\omega \mapsto (X_t(\omega))_{t \in T}$ , et elle est caractérisée par les lois finidimensionnelles des Variables  $(X_t(\omega))_{t \in I}$ ,  $I \subset T$  fini.*

### Définition (Stationnarité stricte)

*Un processus est strictement stationnaire si sa loi est invariante par translation, i.e.  $\forall t_i, h$ ,*

$$\mathbb{P}(X_{t_1}, \dots, X_{t_p}) = \mathbb{P}(X_{t_1+h}, \dots, X_{t_p+h}).$$

### Définition (Stationnarité faible)

*Le processus est faiblement stationnaire si les moyennes et les variances sont constantes, et si  $\text{Cov}(X_t, X_s) = \phi(|t - s|)$ .*

*Un processus strictement stationnaire est faiblement stationnaire.*

## Définition (Accroissements stationnaires)

*Un processus est à accroissements stationnaires si  $(X_{t+h} - X_t)_{t \in T}$  est stationnaire quel que soit  $h$  (au sens strict ou faible, donc).*

## Exemple

Soit  $X_t = at + b + \epsilon_t$  où  $(\epsilon_t)_{t \in \mathbb{R}}$  est iid centrée de variance  $\sigma^2$ . Alors  $(X_t)_{t \in \mathbb{R}}$  n'est pas stationnaire, mais  $(Y_t)_{t \in \mathbb{R}}$  est faiblement stationnaire.

Soit  $(X_t)_{t \geq 1}$  défini par  $\mathbb{P}(X_t = 0) = 1 - 1/t$  et  $\mathbb{P}(X_t = \pm\sqrt{t}) = 1/2t$ . Il est faiblement stationnaire.

Pour les processus stationnaires au sens strict il existe un résultat de convergence :

### Théorème (Théorème Ergodique)

*Soit  $(X_n)_{n \geq 1}$  une suite stationnaire de var, de loi commune intégrable. Alors il existe une var intégrable  $Y$  telle que*

$$\frac{1}{N} \sum_{k=1}^N X_k \xrightarrow{N \rightarrow \infty} Y \quad \mathbb{P} - p.s..$$

*De plus,  $\mathbb{E}(X_1) = \mathbb{E}(Y)$ .*

On dit que la suite est ergodique si quel que soit  $A \in \mathcal{B}_1$ , invariant au sens où  $\forall k \geq 1, X_k^{-1}(A) = X_{k+1}^{-1}(A)$ ,  $\mathbb{P}_{X_1}(A) = 0$  ou  $\mathbb{P}_{X_1}(A) = 1$ .

Alors si la suite est ergodique,  $Y = \mathbb{E}(X_1)$   $\mathbb{P}$ -p.s..

### Remarque

*Le théorème ergodique entraîne la loi forte des grands nombres. L'ergodicité est valable dans ce cas là et se démontre par la loi 0 – 1.*

Sous certaines hypothèses assez générales, un processus stationnaire peut être considéré comme étant de la forme  $(X_n)_{n \in \mathbb{Z}}$  avec  $(\Omega, \mathcal{B})$  un espace métrique compact muni de sa tribu borélienne, une application continue  $\phi : \Omega \rightarrow \Omega$ , et  $\mathbb{P}$  une loi telle que  $\mathbb{P}_\phi = \mathbb{P}$ . Alors  $X_{n+1} = X_n \circ \phi$ . On parle alors de système dynamique  $(\Omega, \mathcal{B}, \mathbb{P}, \phi)$ , et on dit qu'il est ergodique si  $\phi^{-1}(A) = A \Rightarrow \mathbb{P}(A)\mathbb{P}(A^c) = 0$ .

Dans ce cas le théorème ergodique se traduit par un énoncé du type  $\forall f : \Omega \rightarrow \mathbb{R} \in \mathcal{L}^1(\mathbb{P})$ ,

$$\frac{1}{N} \sum_{k=0}^{N-1} f(\phi^k \omega) \rightarrow \mathbb{E}(f) \quad \mathbb{P} - p.s..$$

### Exemple

Si  $(\varepsilon_i(x))_{i \geq 1}$  désigne la suite des chiffres dans le développement en base 10 d'un réel  $x \in [0, 1]$ , alors les suites  $((\varepsilon_i, \dots, \varepsilon_{i+p})_{i \geq 0}$  sont équadistribuées sur  $\{0, 1, \dots, 9\}^{p+1}$ , stationnaires, ergodiques, mais non indépendantes si  $p \geq 1$ .



Soit  $S$  fini ou dénombrable.

### Définition (Noyau de transition)

*Un noyau de transition sur  $S$  est une application  $p : S \times S \rightarrow \mathbb{R}^+$  telle que pour tout  $x \in S$ ,  $p(x, \cdot)$  est une probabilité sur  $S$ .*

### Définition

*Une chaîne de Markov d'espace d'états  $S$  est une suite de VA  $(X_n)_{n \geq 0}$ ,  $X_n : \Omega \rightarrow S$ , telle qu'il existe une suite de noyaux de transition  $(p_n)_{n \geq 0}$ , et une probabilité  $\nu$  sur  $S$ , telles que pour tout  $n \geq 0$  et  $x_0, \dots, x_n \in S$ ,*

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \nu(x_0) \prod_{i=0}^{n-1} p_i(x_i, x_{i+1}).$$

*On appelle  $\nu$  la loi initiale de la chaîne et  $(p_n)_{n \geq 0}$  la famille de noyaux de transition.*

Il découle de cette définition la propriété fondamentale suivante :

$$\mathbb{P}(X_{m+1} = x_{m+1}, \dots, X_n = x_n | X_0 = x_0, \dots, X_m = x_m) =$$

$$\mathbb{P}(X_{m+1} = x_{m+1}, \dots, X_n = x_n | X_m = x_m) .$$

Nous nous intéressons ci-après au cas où  $p_n = p$ ,  $n \geq 0$ . On appelle alors  $p$  la **matrice de transition de la chaîne**. On parle alors de chaîne de Markov homogène.

### Exemple (Marche aléatoire simple)

La marche aléatoire symétrique sur  $\mathbb{Z}$  : soit  $Z = (Z_n)_{n \geq 1}$  une suite iid de var telle que  $\mathbb{P}(Z_n = \pm 1) = \frac{1}{2}$ , et  $X_0$  une var à valeurs dans  $\mathbb{Z}$  indépendante de  $(Z_n)_{n \geq 1}$ . La marche aléatoire simple est alors définie par  $Y = (Y_n)_{n \geq 0}$ ,  $Y_n = X_0 + \sum_{k \leq n} Z_k$ . C'est une chaîne de Markov homogène sur  $\mathbb{Z}$  de noyau  $p(x, y) = 0$  si  $|x - y| \neq 1$  et  $p(x, y) = \frac{1}{2}$  sinon, et de loi initiale  $\mathbb{P}_{X_0}$ .

### Exemple (Stock)

Soit  $X_n$  l'état d'un stock à l'instant  $n$ ,  $D_{n+1}$  la demande aléatoire formulée par les clients à l'instant  $n+1$ , et  $q$  la production fixe de pièces entre deux instants consécutifs. Alors l'état du stock à l'instant suivant est  $X_{n+1} = (X_n + q - D_{n+1})^+$ . On suppose  $(D_n)_{n \geq 0}$  iid. Alors  $(X_n)_{n \geq 0}$  est une chaîne de Markov de loi initiale  $\mathbb{P}_{X_0}$  et de noyau

$$p(x, y) = \mathbb{P}(D = k) \text{ si } y = x + q - k > 0, \quad p(x, 0) = \mathbb{P}(D \geq x + q).$$

### Exemple (Processus de branchement)

Soit  $X_n$  la taille d'une population à la génération  $n$ . On note  $\xi_i^n$  le nombre aléatoire des descendants de chaque individu  $i$  de cette génération,  $1 \leq i \leq X_n$ . Alors  $X_{n+1} = \sum_{i=1}^{X_n} \xi_i^n$ . On suppose  $(\xi_i^n)_{i \leq X_n, n \geq 0}$  iid. Alors  $(X_n)_{n \geq 0}$  est un chaîne homogène sur  $\mathbb{N}$  de noyau  $p(0,0) = 1$  et  $p(x,y) = \mathbb{P}(\sum_{i=1}^x \xi_i^0 = y)$  si  $x \geq 1$ .

Nous y reviendrons.

### Exemple (Ruine du joueur)

Deux joueurs  $A, B$  de fortunes initiales  $a, b$  jouent à pile ou face, en jouant 1 euro à chaque étape.  $B$  perd avec probabilité  $p$  et gagne avec probabilité  $1 - p$ . Les lancers sont indépendants et le jeu s'arrête lorsqu'un joueur est ruiné.

Soit  $(\xi_n)_{n \geq 0}$  iid,  $\xi_n \in \pm 1$ ,  $\mathbb{P}(\xi_n = 1) = p$ . La fortune de  $A$  est alors donnée par

$$X_0 = a, X_n = X_{n-1} + \xi_n \mathbf{1}_{0 < X_{n-1} < a+b}, \quad n \geq 1.$$

Alors  $(X_n)_{n \geq 0}$  est une chaîne de Markov d'état initial  $a$  sur  $\{0, \dots, a+b\}$ , et de noyau

$$p(0,0) = p(a+b, a+b) = 1, \quad p(n, n+1) = 1 - p(n, n-1) = p \text{ sinon.}$$

### Exemple (Processus de renouvellement)

Soit  $(D_i)_{i \geq 1}$  iid à valeurs dans  $\mathbb{N}^*$ ,  $D_i$  représente la durée de vie d'un  $i^{\text{ème}}$  composant installé après que le  $i - 1^{\text{ème}}$  soit tombé en panne. On note  $X_n$  l'âge du composant à la date  $n$ ,

$$X_n = n - \theta_{n-1}, \quad \theta_{n-1} := \sup\{D_1 + \dots + D_k : D_1 + \dots + D_k \leq n\}, \quad \theta_0 = 0$$

La suite  $(X_n)_{n \geq 1}$  est Markov homogène de noyau

$$p(x, x+1) = \mathbb{P}(D_1 > x+1 | D_1 > x),$$

$$p(x, 0) = 1 - p(x, x+1), \quad p(0, 0) = 0.$$

### Exemple (File d'attente)

Une file d'attente permet de traiter une tâche à chaque instant  $n$ . On note  $\xi_n$  le nombre entier aléatoire de tâches à traiter augmentant la file à l'instant  $n$ . On suppose  $(\xi_n)_{n \geq 1}$  iid. Si  $X_n$  est le nombre de tâches à traiter à l'instant  $n$ , alors

$$X_n = (X_{n-1} - 1)^+ + \xi_n, \quad X_1 = \xi_1.$$

La suite  $(X_n)_{n \geq 1}$  est alors Markov sur  $\mathbb{N}$  de noyau

$$p(x, y) = \mathbb{P}(\xi_1 = y - (x - 1)^+), \quad x, y \in \mathbb{N}.$$



### Définition (Mesure invariante)

*Une probabilité  $\nu$  sur l'espace d'états  $E$  d'une chaîne de Markov homogène  $(X_n)_{n \geq 0}$  de noyau  $p$  est dite invariante si  $\nu p = \nu$  (produit matriciel).*

### Exemple

La marche aléatoire symétrique n'admet pas de loi invariante.

## Théorème

*Si  $E$  est fini, il existe au moins une loi invariante.*

Une chaîne est dite réversible par rapport à une loi  $\nu$  si pour tous  $x, y \in E$ ,

$$\nu(x)p(x, y) = \nu(y)p(y, x).$$

Si tel est le cas,  $\nu$  est nécessairement invariante.

On suppose que la loi initiale de la chaîne est  $\mathbb{P}_{X_0} = \delta_x$  pour un état donné  $x \in E$ . On définit alors le premier retour à l'état  $x \in E$  par

$$R_x = \inf\{n \geq 1 : X_n = x\}.$$

C'est un temps d'arrêt.

### Définition

*Un état  $x \in E$  est dit récurrent si  $\mathbb{P}(R_x < +\infty | X_0 = x) = 1$ , sinon il est dit transitoire. S'il est récurrent, il est*

- *récurrent positif si  $\mathbb{E}(R_x | X_0 = x) < +\infty$  ;*
- *récurrent nul sinon.*

*Si  $E$  est fini, tout état récurrent est récurrent positif.*

On notera  $\mathbb{P}_x := \mathbb{P}(\cdot | X_0 = x)$ .

Soit  $N^x := \sum_{n \geq 0} \mathbf{1}_{X_n=x}$  le nombre de visites de l'état  $x$  par la chaîne. Alors  $\mathbb{P}_x(N^x > k) = \mathbb{P}_x(R_x < \infty)^k$ , et donc

- $x$  est récurrent ssi  $\mathbb{P}_x(N^x = \infty) = 1$  ssi  $\mathbb{E}(N^x | X_0 = x) = \infty$  ;
- $x$  est transitoire ssi  $\mathbb{P}_x(N^x = \infty) = 0$ .

Soit  $\mu_x(y) = \mathbb{E}(\sum_{n=0}^{R_x-1} \mathbf{1}_{X_n=y}) = \sum_{n \geq 0} \mathbb{P}_x(R_x > n, X_n = y)$ ,  
 $x \in E$ .

## Théorème

- $\mu_x p = \mu_x$  ssi  $x$  est récurrent ;
- $\mu_x$  est finie ssi  $x$  est récurrent positif. Alors  $\nu_x := \mu_x / \mathbb{E}(R_x | X_0 = x)$  est une loi invariante pour la chaîne.

## Définition

*Soient deux états  $x, y \in E$  :*

- *on dit que  $x$  communique avec  $y$ , et on note  $x \rightarrow y$ , si il existe  $n \geq 1$  et des états  $x_0 = x, x_1, \dots, x_n = y$  tels que  $p(x_0, x_1) \dots p(x_{n-1}, x_n) > 0$ .*
- *on note  $x \leftrightarrow y$  si  $x \rightarrow y$  et  $y \rightarrow x$ .*
- *une classe  $E_0 \subset E$  est dite irréductible si  $x \leftrightarrow y$  pour tous  $x, y \in E_0$ .*
- *une chaîne est irréductible si  $E$  l'est.*
- *une classe  $E_0 \subset E$  est fermée si  $x \in E_0$  et  $x \rightarrow y \Rightarrow y \in E_0$ .*

## Proposition

- si  $x \rightarrow y$  et  $x$  est récurrent, alors  $y$  est récurrent et  $N^y = +\infty$   $\mathbb{P}_x$ -p.s..
- si  $x \leftrightarrow y$  alors ils sont simultanément soit transitoires soit récurrents.
- les états qui ne communiquent pas avec eux-mêmes sont transitoires.
- les classes d'équivalence pour  $\leftrightarrow$  des états tels que  $x \rightarrow x$  sont des classes irréductibles fermées.
- tous les autres états sont transitoires.

Donc soit une chaîne démarre d'un état récurrent et alors elle peut être considérée comme une chaîne irréductible restreinte à la classe d'équivalence de l'état de départ.

Soit elle part d'un état transitoire et atteint un état récurrent en temps fini.

Soit elle va d'états transitoires en états transitoires (un nombre fini de fois pour chaque) et part vers l'infini. On dit qu'elle est transiente.



## Théorème

*Pour une chaîne de Markov irréductible  $X = (X_n)_{n \geq 0}$ , il y a équivalence entre :*

- *$X$  est récurrente positive ;*
- *$X$  admet une loi invariante  $\nu$ .*

*Si tel est le cas,  $\nu$  est unique et décrite par*

$$\nu(x) = \frac{1}{\mathbb{E}(R_x | X_0 = x)}, \quad x \in E.$$

Il s'en suit que toute chaîne irréductible à espace d'états fini est récurrente.

### Exemple (Application à la marche aléatoire)

Soit  $\{e_i : 1 \leq i \leq n\}$  la base canonique de  $\mathbb{R}^n$  et posons

$$X_n = X_{n-1} + \xi_n, \quad n \geq 1,$$

avec  $(\xi_n)_{n \geq 1}$  iid de loi uniforme sur  $\{\pm e_i : 1 \leq i \leq n\}$ . Cette chaîne est irréductible sur  $\mathbb{Z}^n$ . Alors on montre que cette marche est récurrente nulle ssi  $n \leq 2$ , sinon elle est transiente.

### Exemple (Ruine du joueur)

Les états possibles sont  $\{0, \dots, N\}$ ,  $N$  représentant la fortune totale des deux joueurs. La classe  $E_0 = \{1, \dots, N-1\}$  est irréductible mais non fermée, les états 0 et  $N$  sont absorbants.

### Exemple (Processus de branchement)

L'état  $\{0\}$  est absorbant, et forme une classe fermée récurrente. Si  $\mathbb{P}(\xi_1 = 0) > 0$ , tout état est transitoire car il communique avec 0 qui est absorbant. Si  $\mathbb{P}(\xi_1 = 0) + \mathbb{P}(\xi_1 = 1) < 1$ , alors  $\mathbb{N}^*$  est irréductible mais non fermée. Alors soit la population s'éteint, soit elle explose.

### Exemple (Processus de renouvellement)

Si  $\mathbb{P}(D_1 = x) > 0$  pour tout  $x \geq 1$ , la chaîne est irréductible, et 0 est récurrent positif ssi  $\mathbb{E}(D_1) < +\infty$ .

### Théorème (Théorème ergodique pour chaînes de Markov)

*Si  $X$  est irréductible récurrente positive sur  $E$ , de noyau  $p$ , et d'unique loi invariante  $\nu$ , alors pour toute  $g : E \times E \rightarrow \mathbb{R}$  positive ou telle que  $\mathbb{E}_\nu(|g(X_0, X_1)|) < +\infty$ ,*

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k) &\rightarrow_{p.s.} \mathbb{E}_\nu(g(X_0, X_1)) \\ &= \sum_{x \in E} \nu(x) \sum_{y \in E} p(x, y) g(x, y) \end{aligned}$$

*quelle que soit la loi initiale  $\pi_0$  de  $X_0$ .*

Il existe aussi un théorème limite centrale pour chaînes de Markov...

Pour tout  $x \in E$ , on note  $p(x) = \text{PGCD}(I(x))$  où  $I(x) = \{n \geq 1 : p^n(x, x) > 0\}$ . Si la chaîne de Markov est irréductible, alors  $p(x)$  est constante, notée  $\mathbf{p}_X$ .

### Proposition

*On dit que la chaîne est apériodique si  $\mathbf{p}_X = 1$ . Ceci équivaut à ce que pour tout  $x \in E$ , il existe  $n(x) \geq 1$  tel que  $p^n(x, x) > 0$  si  $n \geq n(x)$ . Si  $E$  est fini,  $n(x)$  ne dépend pas de  $x$ .*

### Théorème

*Si  $X$  est récurrente positive d'unique loi invariante  $\nu$ , et apériodique, alors*

$$\mathbb{P}(X_n = x) \rightarrow \nu(x), \quad x \in E.$$

### Proposition (Condition de Doeblin)

*Supposons  $E$  fini,  $X$  irréductible et apériodique. Alors  $p$  vérifie la condition de Doeblin : il existe  $k \geq 0$ ,  $\varepsilon > 0$  et  $\delta$  une loi sur  $E$  tels que*

$$p^k(x, y) \geq \varepsilon \delta(y), \quad x, y \in E.$$

*Alors il existe une unique loi invariante  $\nu \geq \varepsilon \delta$  et elle vérifie*

$$\sup_{x \in E} \sum_{y \in E} |p^n(x, y) - \nu(y)| \leq 2(1 - \varepsilon)^{\lfloor \frac{n}{k} \rfloor}.$$

# L'algorithme Pagerank de google

Soit  $E = \{1, \dots, N\}$  l'ensemble des pages web, chaque page  $i$  contient des liens  $I_i = \{i_k : 1 \leq k \leq d_i\} \subset E$  vers d'autres pages. Pour les pages sans lien,  $I_i := E$ . On assigne un poids uniforme pour le déplacement d'une page à une autre :

$$L(i, j) = (1/d_i) \mathbf{1}_{j \in I_i}, \quad i, j \in E.$$

Ceci modélise un navigateur sur le web surfant totalement au hasard. On lui associe une chaîne de Markov homogène  $(X_n)$  où  $X_n$  est la page visitée à la date  $n$ .

Le principe de Pagerank est d'attribuer un rang  $r_i$  à chaque page  $i \in E$  en fonction du nombre de liens pointant vers  $i$ , et du rang de la page contenant ce lien :

$$r_i = \sum_{j:i \in I_j} \frac{r_j}{d_j} \implies r = rL.$$

On rechercherait donc  $r$  comme une loi invariante pour la chaîne de transition  $L$ . Plusieurs problèmes se posent (convergence, unicité...).

Brin et Page ont introduit la matrice de Google  $G$  définie par

$$G_{i,j} = \alpha L_{i,j} + (1 - \alpha) \frac{1}{N}, \quad i, j \in E,$$

où  $\alpha$  est un paramètre de damping (amortissement).  $G$  vérifie la condition de Doeblin. Le choix de  $\alpha$  proche de 0 accélère la convergence mais gomme la vraie nature du web. La vraie valeur de  $\alpha$  est secrète proche de 0.85.



# Algorithme de Hastings-Metropolis

On veut simuler une v.a.  $Z$  de loi  $\nu$  sur  $E$ . On va pour cela interpréter  $\nu$  comme loi invariante d'une chaîne de Markov. On choisit pour cela une matrice stochastique  $Q$  sur  $E$ , dite de *Proposition*, irréductible et vérifiant

$$Q(x, y) > 0 \Leftrightarrow Q(y, x) > 0.$$

Puis  $h : \mathbb{R}^{*,+} \rightarrow ]0, 1]$  telle que  $h(y) = yh(1/y)$  ( $h(y) = \min(1, y)$  par exemple). On définit la probabilité de rejet de la valeur courante :

$$R(x, y) = h\left(\frac{\nu(y)Q(y, x)}{\nu(x)Q(x, y)}\right), \quad x, y \in E, x \neq y, Q(x, y) \neq 0.$$

On choisit  $X_0$  de loi arbitraire, puis  $X_{n-1}$  étant construite,

- on simule  $Y_n$  selon  $Q(X_{n-1}, \cdot)$ ;
- on accepte  $Y_n$  (et on rejette  $X_{n-1}$  avec probabilité  $R(X_{n-1}, Y_n)$ ), en simulant  $U_n$  de loi  $\mathcal{U}([0, 1])$  indépendante de  $(X_{n-1}, Y_n)$  et on pose

$$X_n := Y_n \mathbf{1}_{\{U_n \leq R(X_{n-1}, Y_n)\}} + X_{n-1} \mathbf{1}_{\{U_n > R(X_{n-1}, Y_n)\}}.$$

La matrice de transition de la chaîne  $X$  est donnée par

$$P(x, y) = R(x, y)Q(x, y) \text{ si } x \neq y, \quad P(x, x) = 1 - \sum_{y \neq x} P(x, y).$$

Alors  $P$  est stochastique irréductible, réversible par rapport à  $\nu$ , apériodique si  $Q$  l'est ou  $h < 1$ .

# Limite thermodynamique

Soit  $T > 0$  une constante,  $E$  fini, et  $V : E \rightarrow \mathbb{R}$ . La mesure de Gibbs associée à  $T$  et  $V$  est définie par

$$\Gamma_T^V(x) := \frac{1}{Z_T} \exp\left(-\frac{V(x)}{T}\right), \quad x \in E, \quad Z_T = \sum_{y \in E} \exp\left(-\frac{V(y)}{T}\right)$$

$T$  est la température,  $V(x)$  représente l'énergie du système dans la configuration  $x$ , et  $Z_T$  est appelée la fonction de partition.

La mesure  $\Gamma_T^V$  maximise en un sens l'entropie  $-\sum_{x \in E} \nu(x) \ln \nu(x)$  parmi les lois  $\nu$  d'énergie moyenne donnée en fonction de  $T$ .

Dans la pratique la simulation de la loi de Gibbs pour  $E$  grand se fait avec l'algorithme de Hastings-Metropolis, avec la probabilité de rejet simplement donnée par

$$R(x, y) = \min\left(1, \frac{\exp(-(V(y) - V(x))/T) Q(y, x)}{Q(x, y)}\right), \quad x, y \in E.$$

## Exemple (Le modèle d'Ising)

On considère  $N$  particules positionnées sur une grille régulière, chacune affectée d'un spin d'orientation  $\pm 1$ . L'espace d'états est donc  $E = \{\pm 1\}^N$ . On définit l'énergie du système pour une configuration  $x \in E$  par

$$V(x) = -\frac{1}{2} \sum_{i,j=1}^N J_{i,j} x(i) x(j) - \sum_{k=1}^N x(k) B_k,$$

où  $B = (B_k)_{1 \leq k \leq N}$  modélise l'effet d'un champ magnétique extérieur, et  $J$  modélise l'interaction entre les particules : on la choisit symétrique à diagonale nulle, positive pour les modèles ferromagnétiques.

On choisit souvent comme matrice de proposition la matrice qui fait transiter de  $x$  uniformément vers un état différent en exactement un seul spin.

# Recuit simulé

On cherche des minima globaux  $\in V_{\min} \subset E$  de  $V : E \rightarrow \mathbb{R}$ ,  $E$  fini lorsque les minima locaux peuvent conduire à piéger les méthodes habituelles (puits de potentiel).

L'idée repose sur l'observation que

$$\lim_{T \rightarrow 0^+} \Gamma_T^V(x) = \frac{\mathbf{1}_{V_{\min}}(x)}{\text{Card}(V_{\min})}.$$

On va choisir une suite  $(T_n)_{n \geq 1}$  telle que  $T_n \rightarrow 0$  et à l'étape  $n$  utiliser l'algorithme de Hastings-Metropolis de température  $T_n$ .

Ceci conduit à une chaîne de Markov non homogène  $(X_n)$  de loi initiale  $\pi_0$ , de noyaux  $(p_n)$ , de lois marginales  $\pi_n$  (loi de  $X_n$ ), la loi de Gibbs  $\Gamma_{T_n}^V$  étant celle d'une chaîne de Markov fictive de transition  $p_n$ . Alors

$$\pi_n = \pi_0 p_1 \dots p_n, \quad \Gamma_{T_n}^V = \Gamma_{T_n}^V p_n.$$

Il faut bien contrôler la vitesse  $T_n \rightarrow 0$  pour obtenir la convergence de la chaîne vers  $V_{\min}$ .

Étant donnée  $Q$  symétrique, notons

$$p_n(x, y) = q(x, y) \exp(-(V(y) - V(x))^+ / T_n) \text{ si } x \neq y.$$

### Théorème

*Supposons que  $Q$  satisfasse la condition de Doeblin sur  $E$ , et soit  $T_n = (b \ln n)^{-1}$ . Alors il existe  $b_0 > 0$  tel que pour tout  $b \leq b_0$ ,  $\eta > 0$  et  $\pi_0$ ,*

$$\lim_{n \rightarrow \infty} \| \pi_n - \Gamma_{T_n}^V \|_{\ell^1} = 0 \text{ et } \lim_n \mathbb{P}(V(X_n) \geq \eta + \min_E V) = 0.$$

(Ici,  $\| u \|_{\ell^p} = (\sum_i |u(i)|^p)^{1/p}$ ).

# Le voyageur de commerce

Un Voyageur part d'une ville 0, doit visiter  $N$  villes une seule fois, et cherche à minimiser la distance parcourue selon l'ordre choisi pour les visites, avant de revenir à son point de départ. Ici  $E$  est l'ensemble des permutations  $\sigma$  de  $\{1, \dots, N\}$ , et en notant  $\sigma(0) = \sigma(N+1) = 0$ , il cherche à minimiser

$$V(\sigma) = \sum_{i=0}^N \text{dist}(\sigma(i), \sigma(i+1)).$$

On dira que deux trajets  $\sigma$  et  $\beta \in E$  sont voisins s'ils sont distincts et si le second provient de l'échange d'exactly deux villes dans le premier. On note  $\sigma \equiv \beta$ . On choisit alors

$$Q(\sigma, \beta) = \frac{2}{N^2} \mathbf{1}_{\sigma \equiv \beta} + \frac{1}{N} \mathbf{1}_{\sigma = \beta}.$$

De meilleurs choix de  $Q$  existent.

Une filtration est une suite croissante (au sens de l'inclusion) de tribus  $(\mathcal{F}_n)_{n \geq 0}$ .

Une suite de var  $(X_n)_{n \geq 0}$  est une martingale (resp. sous-martingale) relativement à la filtration  $(\mathcal{F}_n)_{n \geq 0}$  si

- $X_n$  est  $\mathcal{F}_n$ -mesurable ;
- $\mathbb{E}(|X_n|) < +\infty$  ;
- $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$  (resp.  $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$ ).

Si tel est le cas, alors la même suite est nécessairement une martingale relativement à la filtration naturelle  $(\mathcal{G}_n)_{n \geq 0}$ , où  $\mathcal{G}_n = \sigma(X_0, \dots, X_n)$ .

### Définition

*Une suite  $(Y_n)_{n \geq 1}$  est dite prévisible relativement à la filtration  $(\mathcal{F}_n)_{n \geq 0}$  si pour chaque  $n$ ,  $Y_n$  est  $\mathcal{F}_{n-1}$ -mesurable.*



## Exemple

- Si  $(Y_i)_{i \geq 1}$  est iid dans  $\mathcal{L}^1$  centrée, alors  $X_n = Y_1 + \dots + Y_n$  définit une martingale.
- Si  $Z$  est intégrable et si  $(\mathcal{F}_n)_{n \geq 1}$  est une suite croissante de tribus, alors  $X_n = \mathbb{E}(Z|\mathcal{F}_n)$  définit une martingale.
- Si  $(X_n)$  est une martingale,  $(|X_n|)$  est une sous-martingale.

## Théorème (Doob)

*Si  $(X_n)$  est une sous-martingale telle que  $K = \sup_n \mathbb{E}(X_n^+) < +\infty$ , alors il existe  $X \in \mathcal{L}^1$  telle que*

$$X_n \rightarrow X \text{ p.s..}$$

## Théorème (Inégalité maximale de Doob)

*Pour une sous-martingale,*

$$\mathbb{P}(\max_{i \leq n} X_i^+ \geq \alpha) \leq \frac{1}{\alpha} \mathbb{E}(X_n^+).$$

### Exemple

Soit  $(X_n)_{n \geq 1}$  iid telle que  $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 2) = \frac{1}{2}$ . Alors  $Z_n = \prod_{i=1}^n X_i$  est une martingale. Montrer qu'elle converge presque sûrement mais pas dans  $\mathcal{L}^1$ .

## Définition (Temps d'arrêt)

*Un temps d'arrêt relativement à la filtration  $(\mathcal{F}_n)_{n \geq 0}$  est une application  $\tau : \Omega \rightarrow \mathbb{N}$  mesurable telle que  $\{\tau = n\} \in \mathcal{F}_n$  pour chaque  $n$ .*

## Exemple

Si  $(X_n)_{n \geq 0}$  est une suite de VA à valeurs dans  $\mathbb{R}^n$  et  $B \in \mathcal{B}_n$ , alors  $T = \inf\{n : X_n \in B\}$  définit un temps d'arrêt relativement à la filtration  $(\sigma(X_0, \dots, X_n))_{n \geq 0}$ .

## Théorème (Théorème d'arrêt de Doob)

*Si  $(X_n)_{n \geq 0}$  et  $T$  sont respectivement une martingale et un temps d'arrêt relativement à la filtration  $(\mathcal{F}_n)_{n \geq 0}$  alors*

$$\mathbb{E}(X_T) = \mathbb{E}(X_0).$$

# Algorithme de Robbins-Monro

Rappelons un procédé de recherche de racine dans  $\mathbb{R}^d$  :

## Proposition

*Soient  $\theta^*$  et  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  continue, bornée, nulle en  $\theta^*$ , et séparante en  $\theta^*$  (c'est à dire que  $\langle \theta - \theta^*, f(\theta) \rangle > 0$  si  $\theta \neq \theta^*$ ). Soit  $\theta_0$  arbitraire et  $\theta_n = \theta_{n-1} - \gamma_n f(\theta_{n-1})$ ,  $n \geq 1$ . Si  $\gamma_n > 0$ ,  $\sum_n \gamma_n = +\infty$ ,  $\sum_n \gamma_n^2 < +\infty$ , alors  $\theta_n \rightarrow \theta^*$ .*

Supposons que  $f(\theta) = \mathbb{E}(F(\theta, X))$ , avec  $F$  bornée. Soit  $(X_i)_{i \geq 1}$  iid de même loi que  $X$ . Soit  $\mathcal{F}_i = \sigma(X_j : j \leq i)$ . L'algorithme de Robins-Monro est le suivant :

- $Y_{n+1} = F(\theta_n, X_{n+1})$ ;
- $\xi_{n+1} = Y_{n+1} - f(\theta_n)$ ;
- $\theta_{n+1} = \theta_n - \gamma_{n+1} Y_{n+1}$ .

### Proposition

*Si  $\gamma_n > 0$ ,  $\sum_n \gamma_n = +\infty$ ,  $\sum_n \gamma_n^2 < +\infty$ , alors  $\theta_n \rightarrow \theta^*$  p.s..*

Si  $(\xi_i^n)_{i \geq 1, n \geq 0}$  est une suite iid de loi commune celle de  $\xi \in \mathbb{N}$ , le processus de Galton - Watson associé est défini par

$$Z_0 = 1 \text{ et } Z_{n+1} := \sum_{i=1}^{Z_n} \xi_i^n.$$

On note  $p_j = \mathbb{P}(\xi = j)$ ,  $j \geq 0$ . Alors  $(Z_n)_{n \geq 0}$  est une chaîne homogène de noyau

$$p(i, j) = \mathbb{P}(Z_{n+1} = j | Z_n = i) = \sum_{j_1 + \dots + j_i = j} p_{j_1} \dots p_{j_i}.$$

On va étudier ce processus grâce aux fonctions génératrices.  
Remarquons que si  $p_0 + p_1 = 1$ , alors soit  $p_0 = 0$  et  $Z_n = 1$ , soit  $p_0 > 0$  et alors  $Z_n \rightarrow 0$  p.s..

## Définition

*On appelle fonction génératrice d'une var  $\xi \in \mathbb{N}$  la fonction  $G_\xi(s) = \mathbb{E}(s^\xi)$ . Ses dérivées successives en 1 permettent de retrouver les moments de la variable  $\xi$ .*



## Proposition

$$G_{Z_n}(s) = \underbrace{G_\xi \circ \dots \circ G_\xi}_{n \text{ fois}}(s).$$

- $\mathbb{E}(Z_n) = G'_{Z_n}(1) = \mathbb{E}(\xi)^n$  ;
- $G_\xi$  est strictement convexe croissante ;
- $G_\xi(0) = p_0$ ,  $G_\xi(1) = 1$  ;
- si  $\mathbb{E}(\xi) \leq 1$ ,  $G_\xi$  n'a pas de point fixe sur  $[0, 1[$  ;
- si  $\mathbb{E}(\xi) > 1$ ,  $G_\xi$  admet un unique point fixe sur  $[0, 1[$  ;
- si  $\pi = \min\{0 \leq t \leq 1 : G_\xi(t) = t\}$ , et  $0 \leq s < \pi < t < 1$ , alors  $G_{Z_n}(s) \uparrow \pi$  et  $G_{Z_n}(t) \downarrow \pi$  ;
- $G_{Z_n}$  est dérivable sur  $[0, 1[$  et converge simplement ;
- $G'_{Z_n}(s) \geq G'_\xi(s)^n$ ,  $G'_{Z_n}(t) \geq G'_\xi(t)^n$ .

Si l'on considère  $p_n = (1 - p)p^n$  pour  $0 < p < 1$ ,  
 $G_\xi(s) = (1 - p)/(1 - ps)$  et on calcule que  $\pi = \min\{1, (1 - p)/p\}$ .  
Également on remarque que

$$\frac{G_\xi(s) - (1 - p)/p}{G_\xi(s) - 1} = \frac{1 - ps - (1 - p)/p}{p} \frac{1}{s - 1}$$

ce qui conduit par récurrence à une expression de  $G_{Z_n}$  qui permet de calculer  $G_{Z_n}(0)$  et de conclure que

- si  $\mathbb{E}(\xi) \leq 1$  ( $p \leq \frac{1}{2}$ ) alors  
 $\mathbb{P}(\text{extinction}) = \mathbb{P}(\cup_n \uparrow \{Z_n = 0\}) = 1 = \pi$ ;
- si  $\mathbb{E}(\xi) > 1$ ,  $\mathbb{P}(\text{extinction}) = 1/\mathbb{E}(\xi) = (1 - p)/p = \pi$ .

On s'intéresse à

$\mathbb{P}(\text{extinction}) = \mathbb{P}(\cup_n \uparrow \{Z_n = 0\}) = \lim_n \uparrow G_{Z_n}(0)$ . Soit  
 $\{\text{explosion}\} = \{Z_n \rightarrow +\infty\}$ . Alors

### Proposition

Si  $\pi$  est le plus petit point fixe de  $G$  sur  $[0, 1]$ , alors

$$\mathbb{P}(\text{extinction}) = 1 - \mathbb{P}(\text{explosion}) = \pi.$$

- *sous-critique* :  $\mathbb{E}(\xi) < 1 \Rightarrow \text{extinction p.s.}$ ,  
 $\mathbb{P}(\limsup\{Z_n = 0\}) = 0$ , et  $\mathbb{P}(\{Z_n \neq 0\}) \leq \mathbb{E}(\xi)^n$ .
- *critique* :  $\mathbb{E}(\xi) = 1 \Rightarrow \text{extinction p.s.}$ . Si  $\mathbb{E}(\xi^2) < \infty$ ,  
 $\lim_n n\mathbb{P}(\{Z_n \neq 0\}) = 2/\text{Var}(\xi)$ .
- *sur-critique* :  $\mathbb{E}(\xi) > 1 \Rightarrow \text{extinction avec probabilité } \pi$ . Et  
 $a > 1 \Rightarrow \mathbb{P}(\liminf\{Z_n \notin [1, a]\}) = 1$ , et

$$1 - \mathbb{P}(Z_n > a | \text{non extinction}) \leq \frac{\pi^{1-a}}{1 - \pi} G'_\xi(\pi)^n, \quad n \geq 1.$$

On observe que  $M_n = Z_n / \mathbb{E}(\xi)^n$  est une martingale positive partant de  $M_0 = 1$ . Il existe donc  $M_\infty \in \mathcal{L}^1$  telle que  $M_n \rightarrow M_\infty$  p.s.. On peut obtenir les informations suivantes si  $\mathbb{E}(\xi^2) < +\infty$  et  $m := \mathbb{E}(\xi) > 1$  :

- $M_n \rightarrow M_\infty$  dans  $\mathcal{L}^2$  ;
- $\mathbb{E}(M_\infty) = 1$ ,  $\text{Var}(M_\infty) = \frac{\text{Var}(\xi)}{m^2 - m}$  ;
- $\mathbb{P}(M_\infty = 0) = \pi$ .