

# **Modélisation Stochastique I**

**Fabien Panloup**

Polycopié de cours

Master 1 Data Science - Université d'Angers

Année universitaire 2022-2023

# Table des matières

<b>1</b>	<b>Rappels et compléments</b>	<b>5</b>
1.1	Rappels de probabilités . . . . .	5
1.1.1	Variables aléatoires réelles . . . . .	5
1.1.2	Espérance/Moments . . . . .	6
1.1.3	Caractérisation de la loi/Indépendance . . . . .	8
1.1.4	Quantiles . . . . .	10
1.1.5	Vecteurs aléatoires réels . . . . .	10
1.2	Convergence de suites aléatoires . . . . .	12
1.2.1	Deux théorèmes fondamentaux . . . . .	15
1.2.2	Préservation des convergences . . . . .	16
1.3	Espérance conditionnelle/Loi conditionnelle . . . . .	17
1.3.1	Un exemple . . . . .	17
1.3.2	Définition/Propriétés . . . . .	18
1.3.3	Loi conditionnelle . . . . .	20
<b>2</b>	<b>Chaînes de Markov</b>	<b>23</b>
2.1	Un exemple . . . . .	23
2.2	Introduction aux chaînes de Markov . . . . .	25
2.3	Classification des chaînes de Markov . . . . .	27
2.4	Probabilité invariante . . . . .	33
2.5	Convergence vers la probabilité invariante . . . . .	38
<b>3</b>	<b>Simulation et Méthodes de Monte-Carlo</b>	<b>41</b>
3.1	Simulation de variables aléatoires réelles . . . . .	41
3.1.1	Génération de nombres pseudo-aléatoires . . . . .	41
3.1.2	Simulation de variables aléatoires discrètes . . . . .	43
3.1.3	Méthode par inversion . . . . .	43
3.1.4	Méthode du rejet . . . . .	45
3.1.5	Simulation de variables aléatoires gaussiennes . . . . .	48
3.1.6	Annexe : quelques rappels sur les vecteurs gaussiens . . . . .	49
3.2	Méthodes de Monte-Carlo . . . . .	52
3.2.1	Principe . . . . .	52
3.2.2	Contrôle de l'erreur . . . . .	53
3.2.3	Réduction de variance . . . . .	55
3.3	Ouverture : Méthodes MCMC . . . . .	60

# Chapitre 1

## Rappels et compléments

### 1.1 Rappels de probabilités

Nous supposons connus la théorie élémentaire de la mesure, le calcul des probabilités, ses axiomes et ses définitions. Nous encourageons les étudiant(e)s à se rapporter aux cours correspondants de Licence. L'objectif de ces rappels est de se focaliser uniquement sur les notions qui joueront un rôle important dans les cours de statistique et probabilités du master.

#### 1.1.1 Variables aléatoires réelles

**Définition 1.1.1.** Une variable aléatoire réelle (v.a.r.) sur un espace probabilisable  $(\Omega, \mathcal{A})$  est une fonction mesurable  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  où  $\mathcal{B}(\mathbb{R})$  est la tribu borélienne de  $\mathbb{R}$ . Pour tout  $\omega \in \Omega$ ,  $X(\omega)$  est une réalisation de  $X$ . L'ensemble  $E = X(\Omega)$  des réalisations possibles de  $X$  est son univers image.

**Définition 1.1.2.** La fonction de répartition d'une v.a.r.  $X$  est la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  définie par

$$\forall t \in \mathbb{R}, \quad F_X(t) = \mathbb{P}(X \leq t).$$

La fonction de répartition  $F_X$  est croissante, càdlàg (continue à droite et admettant une limite à gauche), et satisfait

$$\lim_{t \rightarrow -\infty} F_X(t) = 0 \quad \text{et} \quad \lim_{t \rightarrow +\infty} F_X(t) = 1.$$

Bien qu'il existe d'autres cas de figure, on rencontre généralement  $X$  dans l'une des configurations suivantes :

- $E$  est fini ou dénombrable, on qualifie la variable de *discrète*. La fonction  $f_X : E \rightarrow [0, 1]$  définie par

$$\forall k \in E, \quad f_X(k) = \mathbb{P}(X = k)$$

est sa densité par rapport à la mesure de comptage. *Exemples.* Bernoulli, binomiale, Poisson, etc.

- Il existe une fonction intégrable  $f_X : \mathbb{R} \rightarrow \mathbb{R}^+$  telle que

$$\forall t \in \mathbb{R}, \quad F_X(t) = \int_{-\infty}^t f_X(s) ds.$$

Alors,  $X$  admet la densité  $f_X$  par rapport à la mesure de Lebesgue. On qualifiera (abusivement) cette variable de *continue* et dans ce cas, l'ensemble  $E$  est non dénombrable. En pratique, on aura

$$\forall t \in C'_F, \quad f_X(t) = F'_X(t)$$

où  $C'_F$  est l'ensemble de dérivabilité de  $F_X$ . On remarquera que  $f_X(t) = 0$  dès que  $t \notin E$ .  
*Exemples.* Normale, exponentielle, uniforme continue, etc.

Si  $X_1 \in E$  est discrète et  $X_2 \in \mathbb{R}$  continue, on a pour tout  $A \subset \mathbb{R}$ ,

$$\mathbb{P}(X_1 \in A) = \sum_{k \in A} \mathbb{P}(X_1 = k) \quad \text{et} \quad \mathbb{P}(X_2 \in A) = \int_A f_{X_2}(s) ds.$$

Pour  $A = ]-\infty, t]$ , on retrouve  $F_X(t)$  et pour  $A = \mathbb{R}$ , on sait que ces sommes valent 1.

### 1.1.2 Espérance/Moments

D'une manière générale, la *formule du transfert* appliquée à toute fonction  $h : \mathbb{R} \rightarrow \mathbb{C}$  mesurable établit que, si l'intégrale existe,

$$\mathbb{E}[h(X_1)] = \sum_{k \in E} h(k) \mathbb{P}(X_1 = k) \quad \text{et} \quad \mathbb{E}[h(X_2)] = \int_{\mathbb{R}} h(s) f_{X_2}(s) ds.$$

*Exemple.* Soit  $A \subset \mathbb{R}$ . Montrer que  $\mathbb{E}[\mathbf{1}_{\{X \in A\}}] = \mathbb{P}(X \in A)$ .

$$\mathbb{E}[\mathbf{1}_{\{X \in A\}}] \stackrel{\text{discret}}{=} \sum_{k \in \mathbb{Z}} \mathbf{1}_{\{k \in A\}} \mathbb{P}(X = k) = \sum_{k \in A} \mathbb{P}(X = k) = \mathbb{P}(X \in A)$$

Cas continu  $\rightarrow$  Exercice.

**Définition 1.1.3.** La *fonction caractéristique* d'une v.a.r.  $X$  est définie par

$$\Phi_X(t) = \mathbb{E}[e^{itX}], \quad t \in \mathbb{R}.$$

*Exemple.* Déterminez la fonction caractéristique d'une variable suivant la loi  $\mathcal{N}(0, 1)$ .

Si  $X \sim \mathcal{N}(0, 1)$ , alors  $f_X(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \quad x \in \mathbb{R}$

$$\Phi_X(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{its} \frac{e^{-s^2/2}}{\sqrt{2\pi}} ds = \int_{\mathbb{R}} \frac{e^{-(s-it)^2}}{\sqrt{2\pi}} \cdot e^{\frac{(it)^2}{2}} ds$$

$$\frac{-s^2 + it}{2} = -\frac{1}{2}(s^2 - 2it)$$

$$\Phi_X(t) = e^{-t^2/2} \int_{\mathbb{R}} \underbrace{\frac{e^{-z^2/2}}{\sqrt{2\pi}}}_{=1} dz \quad \text{avec} \quad z = s - it$$

$$\Phi_X(t) = e^{-t^2/2}$$

**Définition 1.1.4.** Le *moment d'ordre*  $r > 0$  des v.a.r.  $X_1$  discrète et  $X_2$  continue est défini par

$$\mathbb{E}[X_1^r] = \sum_{k \in E} k^r \mathbb{P}(X_1 = k) \quad \text{et} \quad \mathbb{E}[X_2^r] = \int_{\mathbb{R}} s^r f_{X_2}(s) ds.$$

Le moment d'ordre  $r$  existe si et seulement si  $\mathbb{E}[|X|^r] < +\infty$ . On voit que le moment d'ordre  $r$  correspond à  $h(x) = x^r$ . L'espérance de  $X$  est son moment d'ordre 1 ( $h(x) = x$ ), sa variance est liée à son moment d'ordre 2 par la relation

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

On a ici respectivement  $h(x) = x$  et  $h(x) = (x - \mathbb{E}[X])^2$ . L'écart-type de  $X$  vaut  $\sigma(X) = \sqrt{\mathbb{V}(X)}$  et, quant à la covariance et la corrélation entre deux v.a.r.  $X$  et  $Y$ , on a

$$\mathbb{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \quad \text{et} \quad \text{Cor}(X, Y) = \frac{\mathbb{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

On voit en particulier que  $\mathbb{Cov}(X, Y) = \mathbb{Cov}(Y, X)$  et que  $\mathbb{Cov}(X, X) = \mathbb{V}(X)$ . La corrélation quantifie le degré de *dépendance linéaire* entre  $X$  et  $Y$  dans le sens où  $\text{Cor}(X, Y) = \pm 1$  est équivalent à l'existence d'une relation de la forme  $Y = \alpha X + \beta$ . Elle ne doit pas être confondue avec la notion d'*indépendance* :  $X \perp Y \Rightarrow \text{Cor}(X, Y) = 0$  mais la réciproque est généralement fautive. La formule du transfert et la linéarité de l'intégrale nous donnent directement les propriétés de l'espérance et de la variance/covariance proposées dans le formulaire. Il est alors facile d'obtenir les généralisations

$$\mathbb{E}\left[\sum_{k=1}^n X_k\right] = \sum_{k=1}^n \mathbb{E}[X_k] \quad \text{et} \quad \mathbb{V}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \sum_{\ell=1}^n \mathbb{Cov}(X_k, X_\ell).$$

On s'intéresse maintenant à deux inégalités célèbres que l'on rencontrera par la suite.

**Proposition 1.1.5** (*Inégalité de Bienaymé-Tchebychev*). Soit  $X$  une v.a.r. de variance finie. Alors, pour tout  $\varepsilon > 0$ ,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

**Preuve :**

Etape 1 :  $Y$  v.a. positive  $a > 0$ . Mg  $\mathbb{P}(Y \geq a) \leq \frac{1}{a} \mathbb{E}[Y]$  (Inégalité de Markov)

$$\mathbb{E}[Y] = \mathbb{E}[Y \mathbb{1}_{\{Y \geq a\}}] + \mathbb{E}[Y \mathbb{1}_{\{Y < a\}}] \geq \mathbb{E}[Y \mathbb{1}_{\{Y \geq a\}}] \geq \mathbb{E}[a \mathbb{1}_{\{Y \geq a\}}] \geq a \mathbb{P}(Y \geq a)$$

Etape 2 :  $\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) = \mathbb{P}(\underbrace{|X - \mathbb{E}[X]|^2}_{Y} \geq \underbrace{\varepsilon^2}_a) \leq \frac{1}{a} \mathbb{E}[Y] = \frac{1}{\varepsilon^2} \mathbb{E}[|X - \mathbb{E}[X]|^2] = \frac{\mathbb{V}(X)}{\varepsilon^2}$

Markov

**Proposition 1.1.6** (*Inégalité de Hölder*). Soient  $1 < p, q < +\infty$  deux réels tels que  $\frac{1}{p} + \frac{1}{q} = 1$  et  $X$  et  $Y$  des v.a.r. telles que  $\mathbb{E}[|X|^p] < +\infty$  et  $\mathbb{E}[|Y|^q] < +\infty$ . Alors,

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} (\mathbb{E}[|Y|^q])^{\frac{1}{q}}.$$

Preuve: Indices pour la démo:

- ①  $a^{1/p} b^{1/q} \leq \frac{1}{p} a + \frac{1}{q} b$   $a, b \geq 0$  passer en log. et utiliser sa concavité
- ② En déduire Hölder lorsque  $\mathbb{E}[X^p] = \mathbb{E}[Y^q] = 1$
- ③ Cas général.

Remarques:

→ si  $p=q=2$   $\mathbb{E}[XY] \leq \mathbb{E}[X^2]^{1/2} \mathbb{E}[Y^2]^{1/2}$  (Cauchy-Schwarz)

→ Posons  $\langle X, Y \rangle = \mathbb{E}[XY]$  produit scalaire sur  $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ , i.e. l'espace des v.a.  $X$  tq  $\mathbb{E}[X^2] < +\infty$

$\|X\| = \sqrt{\langle X, X \rangle} = \mathbb{E}[X^2]^{1/2}$  D'où l'inégalité de Cauchy-Schwarz:

$$|\langle X, Y \rangle| \leq \|X\|_2 \|Y\|_2$$

Comme corollaire immédiat ( $p = q = 2$ ), on en tire l'inégalité de Cauchy-Schwarz,

$$\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2]} \sqrt{\mathbb{E}[Y^2]} \quad \text{ou, formulée autrement,} \quad |\text{Cov}(X, Y)| \leq \sqrt{\mathbb{V}(X)} \sqrt{\mathbb{V}(Y)}.$$

Cette deuxième écriture permet de démontrer que  $|\text{Cor}(X, Y)| \leq 1$ .

### 1.1.3 Caractérisation de la loi/Indépendance

**Définition 1.1.7.** Deux variables aléatoires  $X$  et  $Y$  ont même loi si pour toute fonction  $f$  mesurable et positive,

$$\left. \begin{array}{l} \cdot F_X = F_Y \\ \cdot \phi_X = \phi_Y \\ \cdot f_X = f_Y \end{array} \right\} \Leftrightarrow \mathbb{E}[f(X)] = \mathbb{E}[f(Y)].$$

**Important.** Dans le cas de variables discrètes à valeurs dans un ensemble (fini ou dénombrable)  $E$ , il suffit de vérifier que cette égalité est vraie pour les fonctions  $f(x) = 1_{x=k}$  pour tout  $k \in E$ , i.e. que  $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$  pour tout  $k \in E$ .

Dans le cas des variables continues, il suffit que l'égalité soit vraie pour une classe de fonctions suffisamment riche. Par exemple, deux variables aléatoires  $X$  et  $Y$  ont même loi si  $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$ ,

- pour toute fonction  $f$  de la forme  $f(x) = 1_{x \in A}$  où  $A$  est un borélien (Dans ce cas, on écrira plus classiquement  $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ ), ou,
- pour toute fonction  $f$  continue bornée, ou même,
- pour toute fonction  $f \in \mathcal{C}^\infty$  à support compact (i.e. nulle en dehors d'un intervalle fermé borné).

*Exemple. Montrez que si  $X \sim \mathcal{E}(\lambda)$ , alors  $\lambda X$  suit une loi  $\mathcal{E}(1)$ .*

Soit  $f: \mathbb{R} \rightarrow \mathbb{R}$  mesurable positive

$$\mathbb{E}[f(\lambda X)] = \int_0^{+\infty} f(\lambda x) \lambda e^{-\lambda x} dx = \int_0^{+\infty} f(y) e^{-y} dy = \mathbb{E}[f(Y)] \quad \text{où } Y \sim \mathcal{E}(1)$$

$y = \lambda x$  d'où  $dy = \lambda dx$

Donc  $\lambda X \sim Y$

En pratique, on peut aussi montrer l'égalité en loi à l'aide des fonctions de répartition ou caractéristiques :

**Proposition 1.1.8.** *Deux variables aléatoires  $X$  et  $Y$  ont même loi si elles ont la même fonction de répartition ou si elles ont la même fonction caractéristique.*

En d'autres termes, ces fonctions contiennent suffisamment d'information pour caractériser la loi.

**Définition 1.1.9.** *Deux variables aléatoires  $X$  et  $Y$  sont indépendantes si pour tout  $f, g$  positives,*

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

Dans la définition ci-dessus, comme précédemment, on considère seulement les fonctions positives afin d'assurer la bonne définition des espérances. Néanmoins, si cette propriété est vraie, alors elle s'étend à toutes les fonctions telles que  $f(X)$ ,  $g(Y)$  et  $f(X)g(Y)$  sont intégrables, i.e. telles que  $\mathbb{E}[|f(X)|]$ ,  $\mathbb{E}[|g(Y)|]$  et  $\mathbb{E}[|f(X)g(Y)|]$  sont finies.

*Exemple. Exprimez la fonction caractéristique de  $X + Y$  en fonction de celles de  $X$  et de  $Y$  lorsque ces variables sont indépendantes.*

$$\Phi_X(t) = \mathbb{E}[e^{itX}] \quad \text{et} \quad \Phi_Y(t) = \mathbb{E}[e^{itY}]$$

$$\Phi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX} e^{itY}] \stackrel{\substack{\uparrow \\ X \perp Y}}{=} \mathbb{E}[e^{itX}] \mathbb{E}[e^{itY}] = \Phi_X(t) \Phi_Y(t)$$

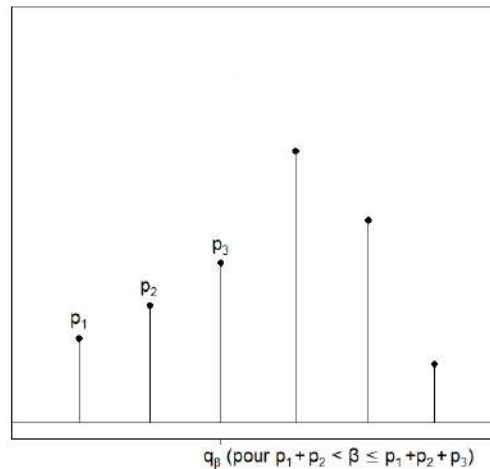
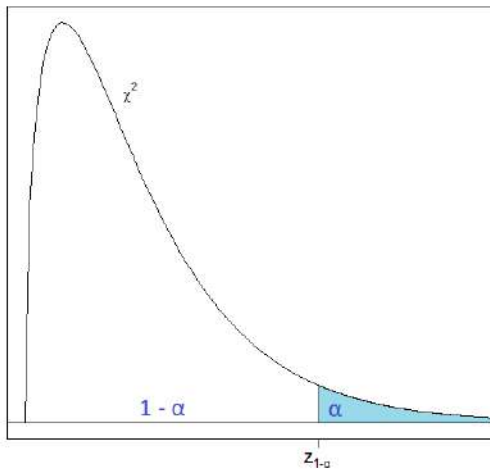
### 1.1.4 Quantiles

**Définition 1.1.10.** La fonction quantile  $Q_X : [0, 1] \rightarrow E$  de la v.a.r.  $X$  est définie par

$$Q_X(\beta) = \inf_{x \in E} \{x \mid F_X(x) \geq \beta\}.$$

$Q_X(\beta)$  est le quantile d'ordre  $\beta$  de la loi de  $X$ .

Lorsque  $F_X$  est strictement croissante continue, il s'agit simplement de son inverse  $F_X^{-1}$  sinon il s'agit d'une inverse généralisée à gauche, parfois notée  $F_X^-$  ou  $F_X^{\leftarrow}$ . Lorsque  $\beta = \frac{1}{2}$ , le quantile obtenu est une *médiane* de  $X$ .



*Exemples.* Soit  $\alpha \in [0, 1]$ . Montrer que, pour les lois  $\mathcal{N}(0, 1)$  et  $t(n)$ , on a  $Q_X(1 - \alpha) = -Q_X(\alpha)$ .

$$\begin{aligned} \text{Si } X \text{ symétrique } &\Leftrightarrow X \sim -X \Rightarrow F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(-X \leq t) \\ &= \mathbb{P}(X \geq -t) = 1 - F_X(-t) \\ \Rightarrow F_X(-Q_X(\alpha)) &= 1 - F_X(Q_X(\alpha)) = 1 - \alpha \\ \Rightarrow -Q_X(\alpha) &\text{ est le quantile d'ordre de } 1 - \alpha \end{aligned}$$

En statistique, le quantile apparaît dans la construction d'intervalles de confiance. Le plus utilisé est sans doute  $u_{0.975} \approx 1.9600$  vérifiant, pour  $X \sim \mathcal{N}(0, 1)$ ,

$$\mathbb{P}(|X| \geq u_{0.975}) = \mathbb{P}(X \leq -u_{0.975}) + \mathbb{P}(X \geq u_{0.975}) = 0.05.$$

### 1.1.5 Vecteurs aléatoires réels

Pour conclure cette section, on introduit rapidement la notion de *vecteur aléatoire réel*, c'est-à-dire d'un vecteur  $\underline{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$  dont chaque composante est une variable aléatoire



réelle. On lui associe un vecteur d'espérance et une matrice de covariance de la façon suivante (sous réserve d'existence),

$$\mathbb{E}[\underline{X}] = (\mathbb{E}[X_k])_{1 \leq k \leq n} \quad \text{et} \quad \mathbb{V}(\underline{X}) = \overset{\text{matrice variance-covariance}}{(\text{Cov}(X_k, X_\ell))_{1 \leq k, \ell \leq n}}$$

où l'on voit donc que l'espérance du vecteur est le vecteur des espérances et la variance du vecteur est la matrice (symétrique et semi-définie positive) des covariances. Le formulaire nous donne quelques propriétés utiles de l'espérance et de la variance/covariance dans le cas vectoriel.

Construire l'espérance et la variance du vecteur  $(X, Y)$  où  $\mathbb{P}(X = 0) = \frac{1}{2}$ ,  $\mathbb{P}(X = 1) = \frac{1}{3}$ ,  $\mathbb{P}(X = 2) = \frac{1}{6}$  et  $Y = \sqrt{X}$ .

$\mathbb{E}[X] = 0 \times \frac{1}{2} + 1 \times \frac{1}{3} + 2 \times \frac{1}{6} = \frac{2}{3}$	$\mathbb{E}[Y] = \sqrt{0} \times \frac{1}{2} + \sqrt{1} \times \frac{1}{3} + \sqrt{2} \times \frac{1}{6} = (2 + \sqrt{2})/6$
$\Rightarrow \mathbb{E}[(X, Y)] = \left[ \frac{2}{3}, (2 + \sqrt{2})/6 \right]$	$\mathbb{V}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ $= 0^2 \times \frac{1}{2} + 1^2 \times \frac{1}{3} + 2^2 \times \frac{1}{6} - \frac{4}{9}$ $= \frac{1}{3} - \frac{4}{9} = \frac{5}{9}$
$\Rightarrow \mathbb{V}[(X, Y)] = \begin{pmatrix} \text{Cov}(X, X) & \text{Cov}(Y, X) \\ \text{Cov}(X, Y) & \text{Cov}(Y, Y) \end{pmatrix}$ $= \begin{pmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \text{Var}(Y) \end{pmatrix}$	$\mathbb{V}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \mathbb{E}[X] - \frac{(2 + \sqrt{2})^2}{6}$ $= \frac{1}{2} - \sqrt{2}/3$
	$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y] = \frac{1 + 2\sqrt{2}}{9}$

**Propriété 1.1.11.** Soit un ensemble de v.a.r. indépendantes et identiquement distribuées (i.i.d.)  $X_1, \dots, X_n$  de même loi qu'une v.a.r.  $X$ . La loi du vecteur aléatoire  $\underline{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$  satisfait

$$\forall B_1, \dots, B_n \subset \mathbb{R}, \quad \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n \mathbb{P}(X \in B_i).$$

Si  $X$  est discrète, la densité du vecteur prendra la forme **la vraisemblance**

$$\forall \underline{k} = (k_1, \dots, k_n) \in E^n, \quad \ell_X(\underline{k}) = \mathbb{P}(\underline{X} = \underline{k}) = \prod_{i=1}^n \mathbb{P}(X = k_i)$$

alors que, si  $X$  est continue, la densité s'écrit simplement

$$\forall \underline{x} = (x_1, \dots, x_n) \in E^n, \quad \ell_X(\underline{x}) = \prod_{i=1}^n f_X(x_i).$$

*Exemple.* Donner la loi, l'espérance et la variance d'un tel vecteur pour  $X \sim \mathcal{B}(p)$  et pour  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

$\underline{X} = (X_1, \dots, X_n)$  où les  $X_i$  sont i.i.d de loi  $\mathcal{B}(p)$

Dans ce cas,  $\forall (x_1, \dots, x_n) \in \{0, 1\}^n$

$$\begin{aligned} \mathbb{P}(\underline{X} = (x_1, \dots, x_n)) &= \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \left( = \mathbb{P}\left(\bigcap_{i=1}^n \{X_i = x_i\}\right) \right) \stackrel{\perp}{=} \prod_{i=1}^n \mathbb{P}(X_i = x_i) \\ &= \prod_{i=1}^n \mathbb{P}(X = x_i) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{S_n} (1-p)^{1-S_n}, \quad S_n = \sum_{i=1}^n x_i \end{aligned}$$

$\underline{X} = (X_1, \dots, X_n)$  avec les  $X_i \perp$  et de loi  $\mathcal{N}(\mu, \sigma^2)$

$X \sim \mathcal{N}(\mu, \sigma^2)$  si  $f_X(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \quad x \in \mathbb{R}.$

$$f_{\underline{X}}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n \frac{e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left(\sum_{i=1}^n -\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

On pourra noter pour terminer que la formule du transfert se généralise également aux vecteurs aléatoires, à l'aide d'une intégration multiple et de la densité du vecteur.

## 1.2 Convergence de suites aléatoires

Dans ce qui suit,  $((U_n)_{n \geq 1}, U)$  est une suite de v.a.r. définies sur un même espace. On note  $F_{U_n}$  la fonction de répartition de  $U_n$  pour  $n \geq 1$ ,  $F_U$  celle de  $U$  et  $C_U$  le domaine de continuité de  $F_U$ .

**Définition 1.2.1.** On dit que  $(U_n)_{n \geq 1}$  **converge en probabilité** vers  $U$  si, pour tout  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|U_n - U| \geq \varepsilon) = 0.$$

On note alors  $U_n \xrightarrow{\mathbb{P}} U$ .

**Définition 1.2.2.** On dit que  $(U_n)_{n \geq 1}$  **converge dans  $L^2$**  vers  $U$  si  $U_n \in L^2$  pour tout  $n \geq 1$ , si  $U \in L^2$  et si

$$\lim_{n \rightarrow +\infty} \mathbb{E}[(U_n - U)^2] = 0.$$

On note alors  $U_n \xrightarrow{L^2} U$ .

Dire que la v.a.r.  $X$  est dans  $L^2$  revient à dire que  $\mathbb{E}[X^2] < +\infty$ .

**Définition 1.2.3.** On dit que  $(U_n)_{n \geq 1}$  **converge presque sûrement** vers  $U$  s'il existe un ensemble  $\Omega^* \subset \Omega$  avec  $\mathbb{P}(\Omega^*) = 1$  tel que

$$\forall \omega \in \Omega^*, \quad \lim_{n \rightarrow +\infty} U_n(\omega) = U(\omega).$$

On note alors  $U_n \xrightarrow{\text{p.s.}} U$ .

Cela traduit la convergence de (presque) toutes les réalisations de la suite, on peut donc interpréter cette convergence et la convergence des suites réelles de manière analogue.

**Définition 1.2.4.** On dit que  $(U_n)_{n \geq 1}$  converge en loi vers  $U$  si pour toute fonction  $h$  continue et bornée,

$$\lim_{n \rightarrow +\infty} \mathbb{E}[h(U_n)] = \mathbb{E}[h(U)].$$

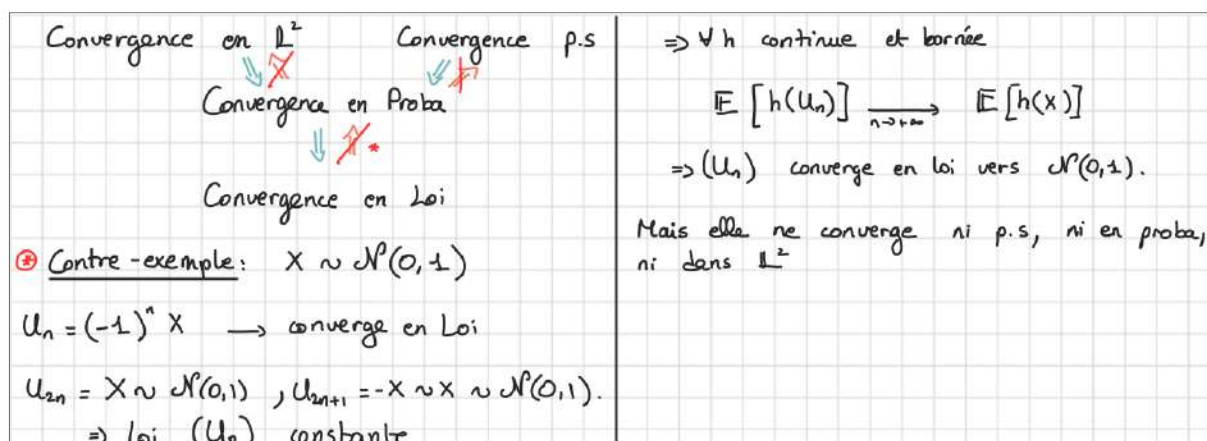
On note alors  $U_n \xrightarrow{\mathcal{L}} U$ .

**Remarque 1.2.1.** De manière équivalente,  $(U_n)_{n \geq 1}$  converge en loi vers  $U$  si pour tout  $t \in C_U$ ,

$$\lim_{n \rightarrow +\infty} F_{U_n}(t) = F_U(t),$$

ce qui se ramène à la convergence simple sur  $C_U$  de la suite  $(F_{U_n})_{n \geq 1}$  vers une fonction de répartition.

**Proposition 1.2.5** (Arbre des implications). Entre ces modes de convergence, on a les implications suivantes :



Prouvons que la convergence p.s. implique la convergence en probabilité.

$$\varepsilon > 0 \quad \mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{E}[Z_n] \quad \text{avec} \quad Z_n = \mathbb{1}_{\{|X_n - X| \geq \varepsilon\}}$$

$$(i) \quad X_n \xrightarrow{p.s.} X \Rightarrow \exists n_0(\omega) \in \mathbb{N} \text{ tq } \forall n \geq n_0(\omega), |X_n(\omega) - X(\omega)| \leq \varepsilon$$

$$\Rightarrow \exists n_0(\omega) \in \mathbb{N} \text{ tq } \forall n \geq n_0(\omega), Z_n(\omega) = 0 \Rightarrow Z_n \xrightarrow{p.s.} 0$$

$$(ii) \quad \text{Domination : } |Z_n| \leq 1 \in L^1(\mathbb{P}) \Leftrightarrow \mathbb{E}[1] = 1 < +\infty$$

$$\text{Par CVD} \quad \mathbb{E}[Z_n] \rightarrow 0 \quad X_n \xrightarrow{\mathbb{P}} X$$

démo sur Moodle

**Proposition 1.2.6** (Lemme de Scheffé, admis). Si  $(f_{U_n})_{n \geq 1}$  forme une suite de densités définies sur  $\mathbb{R}$  par rapport à la mesure de Lebesgue, et si l'on a la convergence presque partout de  $f_{U_n}$  vers une densité  $f_U$ , alors

$$U_n \xrightarrow{\mathcal{L}} U.$$

Dans le cas discret, le résultat est également vrai et le critère de convergence en loi s'écrit plus simplement

$$\forall k \in E, \quad \lim_{n \rightarrow +\infty} \mathbb{P}(U_n = k) = \mathbb{P}(U = k).$$

Exemple. Montrer que si  $U_n \sim \mathcal{B}(n, \frac{\lambda}{n})$  pour un paramètre  $\lambda > 0$ , alors  $U_n \xrightarrow{\mathcal{L}} U \sim \mathcal{P}(\lambda)$ .

k fixé. On veut mq  $\lim_{n \rightarrow +\infty} \mathbb{P}(U_n = k) = \mathbb{P}(U = k)$

$$\begin{aligned} \mathbb{P}(U_n = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n^k}{k!} \underbrace{\left(1 - \frac{\lambda}{n}\right) \left(1 - \frac{\lambda}{n}\right) \dots \left(1 - \frac{\lambda}{n}\right)}_{\substack{\sim 1 \\ n \rightarrow +\infty}} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \underset{n \rightarrow +\infty}{\sim} \frac{n^k}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \end{aligned}$$

$$\left(1 - \frac{\lambda}{n}\right)^{n-k} = \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \xrightarrow{n \rightarrow +\infty} e^{-\lambda} \cdot 1$$

$$\mathbb{P}(U_n = k) \underset{n \rightarrow +\infty}{\sim} \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{Donc } U_n \xrightarrow{\mathcal{L}} U \sim \mathcal{P}(\lambda)$$

Exemple. Montrer que si  $U_n \sim t(n)$ , alors  $U_n \xrightarrow{\mathcal{L}} U \sim \mathcal{N}(0, 1)$ .

Si  $U_n \sim t(n)$  alors  $f_{u_n}(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$

①  $\Gamma(n) = (n-1)!$   $\forall x \in \mathbb{R} \quad \Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$

On peut mq  $\Gamma(x+1) \underset{t \rightarrow \infty}{\sim} \sqrt{2\pi x} \left(\frac{x}{e}\right)^x$  (mais on l'admet)

On regarde  $\frac{\Gamma(x+3/2)}{\Gamma(x+1)} \underset{t \rightarrow \infty}{\sim} \frac{\sqrt{2\pi(x+1/2)}}{\sqrt{2\pi x}} \times \left(\frac{x+1/2}{e}\right)^{x+1/2} \times \left(\frac{e}{x}\right)^x \underset{t \rightarrow \infty}{\sim} \dots \underset{t \rightarrow \infty}{\sim} e^{-1/2} \underbrace{\left(1 + \frac{1}{2x}\right)^2}_{\sim e^{1/2}} \underbrace{\sqrt{x+1/2}}_{\sim \sqrt{x}}$

Ainsi  $\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \underset{t \rightarrow \infty}{\sim} \sqrt{\frac{n}{2}} \underset{t \rightarrow \infty}{\sim} \sqrt{\frac{n}{2}}$

$\log(②) = -\frac{n+1}{2} \log\left(1 + \frac{t^2}{n}\right) \underset{t \rightarrow \infty}{\sim} -\frac{(n+1)}{2} \times \frac{t^2}{n} \underset{t \rightarrow \infty}{\sim} -\frac{t^2}{2}$

Finalement  $f_{u_n}(t) \underset{t \rightarrow \infty}{\sim} \frac{1}{\sqrt{2\pi}} \int \frac{n}{2} e^{-t^2/2}$

$\lim_{n \rightarrow +\infty} f_{u_n}(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad \forall t \in \mathbb{R}$   
densité de  $\mathcal{N}(0, 1)$

Scheffé

$\Rightarrow U_n \xrightarrow{\mathcal{L}} U \sim \mathcal{N}(0, 1)$

Le résultat ci-dessus est souvent utilisé en statistique lorsque l'on considère l'approximation gaussienne des quantiles de Student pour  $n$  suffisamment grand (typiquement,  $t_{0.975}(n) \approx 1.96$ ).

## Deux résultats classiques de théorie de la mesure:

### Proposition 1.2.7. .

- (a) **(Convergence monotone)** Soit  $(X_n)$  une suite de variables aléatoires positives telle que  $(X_n)$  croît p.s. vers  $X$ . Alors,  $(\mathbb{E}[X_n])_{n \geq 0}$  (croît et) converge vers  $\mathbb{E}[X]$ .
- (b) **(Convergence dominée)** Soit  $(X_n)$  une suite de v.a.r. telles que,  $|X_n(\omega)| \leq V(\omega) \in L^1$  et  $X_n(\omega) \rightarrow X(\omega)$  p.s., alors,

$$\mathbb{E}[X_n] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[X].$$

## 1.2.1 Deux théorèmes fondamentaux

**Théorème 1.2.2** (Lois des grands nombres). *On a les deux lois des grands nombres suivantes :*

- **Loi faible (LfGN)** Soit  $(X_n)_{n \geq 1}$  une suite de v.a.r. décorréées, de même espérance  $\mu$  et de même variance  $\sigma^2$  finie. Alors,

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\mathbb{P}} \mu.$$

- **Loi forte (LFGN)** Soit  $(X_n)_{n \geq 1}$  une suite de v.a.r. de même loi et indépendantes, d'espérance  $\mu$  finie. Alors,

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{p.s.}} \mu.$$

<p><b>Preuve: Inégalité de Markov :</b>          Soit <math>Z \geq 0</math> (<math>Z</math> v.a. <math>\mathbb{P}(Z=k) = 1</math> avec <math>k \in \mathbb{R}</math>)          Alors <math>\forall \varepsilon &gt; 0 \quad \mathbb{P}(Z \geq \varepsilon) \leq \frac{\mathbb{E}(Z)}{\varepsilon}</math></p> <p>On a <math>\mathbb{E}[\bar{X}_n] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k] = \frac{n\mu}{n} = \mu</math>  <small>linéarité</small></p> <p><math>\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}</math>  <small>v.a.r. décorréées.</small></p> <p>Soit <math>\varepsilon &gt; 0</math>  <math>\mathbb{P}( \bar{X}_n - \mu  \geq \varepsilon) = \mathbb{P}( \bar{X}_n - \mu ^2 \geq \varepsilon^2)</math>  <small>il y a une équivalence grâce à la valeur absolue</small>  <math>\leq \frac{\mathbb{E}( \bar{X}_n - \mu ^2)}{\varepsilon^2} = \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} \leq \frac{\sigma^2}{n\varepsilon^2}</math></p>	<p>En passant à la limite <math>n \rightarrow +\infty</math>  <math>0 \leq \lim_{n \rightarrow +\infty} \mathbb{P}( \bar{X}_n - \mu  \geq \varepsilon) \leq 0</math></p> <p>d'où <math>\bar{X}_n \xrightarrow{\mathbb{P}} \mu</math></p>
--	--

**Théorème 1.2.3** (Théorème central limite (TCL)). Soit  $(X_n)_{n \geq 1}$  une suite de v.a.r. de même loi et indépendantes, d'espérance  $\mu$  et de variance  $\sigma^2 > 0$  finie. Alors,

$$\frac{\Sigma_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{où} \quad \Sigma_n = \sum_{k=1}^n X_k.$$

En statistique, on rencontre plus fréquemment le TCL sous la forme

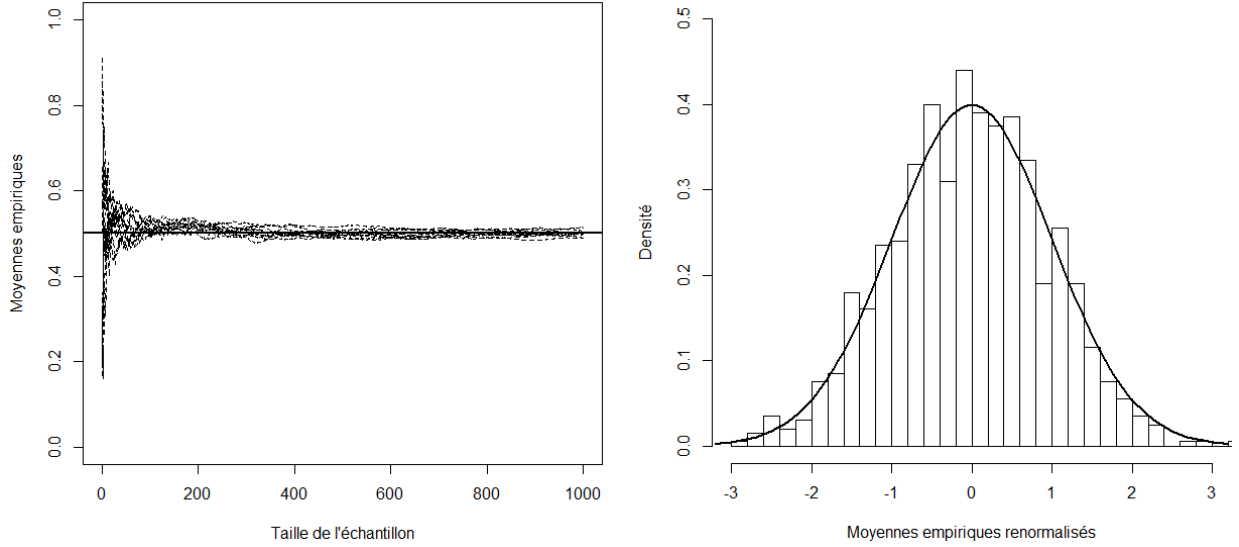
$$Z_n = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

ou encore, pour  $a \leq b$ , sous la forme

$$\lim_{n \rightarrow +\infty} \mathbb{P}(a \leq Z_n \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{s^2}{2}} ds.$$

Ci-dessous, deux graphes illustrant la LGN et le TCL par quelques simulations.

- Pour la LGN : on a simulé  $N = 10$  échantillons de loi  $\mathcal{U}([0, 1])$  de taille  $n = 1000$ , puis on a représenté l'évolution de la suite  $(\bar{X}_k)_{1 \leq k \leq n}$  pour chaque échantillon.
- Pour le TCL : on a simulé  $N = 1000$  échantillons de loi  $\mathcal{U}([0, 1])$  de taille  $n = 1000$ , puis on a représenté l'histogramme des  $N$  valeurs de  $Z_n$  obtenues (ici,  $\mu = \frac{1}{2}$  et  $\sigma^2 = \frac{1}{12}$ ). La densité  $\mathcal{N}(0, 1)$  est superposée.



### 1.2.2 Préservation des convergences

**Proposition 1.2.8** (Théorème de continuité de Mann-Wald (CMT)). Soient  $((X_n)_{n \geq 1}, X) \in \mathbb{R}^p$  une suite de vecteurs aléatoires et  $h : \mathbb{R}^p \rightarrow \mathbb{R}^q$  une fonction dont l'ensemble  $D_h$  des discontinuités est tel que  $\mathbb{P}(X \in D_h) = 0$ . Alors,

$$X_n \xrightarrow{*} X \implies h(X_n) \xrightarrow{*} h(X)$$

où  $\xrightarrow{*}$  désigne soit  $\xrightarrow{\text{p.s.}}$ , soit  $\xrightarrow{\mathbb{P}}$ , soit  $\xrightarrow{\mathcal{L}}$ .

Dans nos applications, on aura généralement  $p = q = 1$  mais il est important de savoir que le résultat est vrai pour les vecteurs également, ne serait-ce que pour établir les corollaires suivants.

**Corollaire 1.2.4.** Si  $U_n \xrightarrow{\text{p.s.}} U$  et  $V_n \xrightarrow{\text{p.s.}} V$ , alors

$$U_n + V_n \xrightarrow{\text{p.s.}} U + V \quad \text{et} \quad U_n V_n \xrightarrow{\text{p.s.}} UV.$$

Il en va de même si l'on remplace  $\xrightarrow{\text{p.s.}}$  par  $\xrightarrow{\mathbb{P}}$ .

Attention à l'erreur classique : avec  $\xrightarrow{\mathcal{L}}$ , le résultat est potentiellement faux. On a besoin pour cela de l'hypothèse plus forte de convergence en loi du vecteur  $(U_n, V_n)$  vers le couple  $(U, V)$ . Tout cela nous conduit à cet autre résultat fondamental de préservation des convergences.



**Proposition 1.2.9** (Lemme de Slutsky). Soient  $((U_n)_{n \geq 1}, U) \in \mathbb{R}^p$  et  $(V_n)_{n \geq 1} \in \mathbb{R}^q$  deux suites de vecteurs aléatoires. Alors, si les suites peuvent être additionnées et/ou multipliées,

$$U_n \xrightarrow{\mathcal{L}} U \quad \text{et} \quad V_n \xrightarrow{\mathbb{P}} c \implies U_n + V_n \xrightarrow{\mathcal{L}} c + U \quad \text{et/ou} \quad U_n V_n \xrightarrow{\mathcal{L}} cU$$

où  $c \in \mathbb{R}^q$  est une constante.

*Exemple.* Soient  $\mu \in \mathbb{R}$  et  $\sigma > 0$  des paramètres. Supposons que  $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ , que  $S_n^2 \xrightarrow{\text{p.s.}} \sigma^2$  et que  $S_n^2 > 0$ . On pose de plus  $S_n^{*2} = \frac{n}{n-1} S_n^2$ . Donner la limite de

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \quad \text{et celle de} \quad \sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^{*2}}}.$$

$U_n = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$	$\Rightarrow \frac{\sigma}{\sqrt{S_n^2}} \xrightarrow{\mathbb{P}} 1$
$S_n^2 \xrightarrow{\text{p.s.}} \sigma^2 \Rightarrow S_n^2 \xrightarrow{\mathbb{P}} \sigma^2$	$\text{Donc } \sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$
$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \cdot \frac{\sigma}{\sqrt{S_n^2}} = U_n \cdot \frac{\sigma}{\sqrt{S_n^2}}$	$\uparrow$ par Slutsky
$h(S_n^2) = \frac{\sigma}{\sqrt{S_n^2}} \quad \text{et on pose } h(x) = \frac{\sigma}{\sqrt{x}} \quad x \in \mathbb{R}_+^*$	
$h(S_n^2) \xrightarrow{\text{p.s.}} h(\sigma^2)$ par continuité	
$\text{donc } \frac{\sigma}{\sqrt{S_n^2}} \xrightarrow{\text{p.s.}} \frac{\sigma}{\sqrt{\sigma^2}} = 1$	
	$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^{*2}}} = \underbrace{\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}}}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)} \times \underbrace{\sqrt{\frac{S_n^2}{S_n^{*2}}}}_{\sqrt{\frac{n-1}{n}} \rightarrow 1} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$
	$\frac{S_n^2}{S_n^{*2}} = \frac{S_n^2}{\frac{n}{n-1} S_n^2} = \frac{n-1}{n}$

## 1.3 Espérance conditionnelle/Loi conditionnelle

### 1.3.1 Un exemple

On lance successivement deux dés et on regarde le produit des deux résultats. Ce jeu peut se modéliser de la manière suivante. On considère un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ . On définit deux variables aléatoires indépendantes  $X$  et  $Z$  à valeurs dans  $\{1, \dots, 6\}$  et on s'intéresse à  $Y = XZ$ . Supposons qu'au premier lancer, on ait obtenu la valeur  $X(\omega) = x \in \{1, \dots, 6\}$ . Connaissant cette information, la moyenne du résultat  $Y = XZ$  est :

$$\mathbb{E}[Y | \{X = x\}] = \frac{\mathbb{E}[XZ 1_{\{X=x\}}]}{\mathbb{P}(X = x)} = x \mathbb{E}[Z].$$

Plutôt que de conditionner par chaque événement, il est alors plus adapté de définir une variable aléatoire  $U$  par  $U(\omega) = X(\omega) \mathbb{E}[Y]$ , qui donne l'espérance du résultat  $Y$  connaissant  $X(\omega)$ . Cette variable aléatoire est appelée l'*espérance conditionnelle* de  $Y$  sachant  $X$ , c'est le résultat moyen attendu connaissant l'information  $X(\omega)$ . On notera dans la suite  $U := \mathbb{E}[Y|X]$ .

On peut remarquer que tel qu'il a été défini dans l'exemple,  $U$  satisfait les deux propriétés suivantes :

1.  $U$  est une fonction de  $X$ .
2. Pour tout  $x \in \{1, 2, 3, 4, 5, 6\}$ ,  $\mathbb{E}[Y1_{X=x}] = \mathbb{E}[U1_{X=x}]$ .

C'est en fait de cette manière que l'on va définir l'espérance conditionnelle de façon plus générale dans la suite.

**Remarque 1.3.1.** Si vous lisez un livre de probabilités standard, l'espérance conditionnelle sera généralement définie relativement à une tribu  $\mathcal{B}$ , i.e. à l'information apportée par les événements appartenant à  $\mathcal{B}$ . Pour simplifier, on se limitera ici à  $\mathcal{B} = \sigma(X)$ , i.e. au cas où la tribu est celle engendrée par la variable  $X$  (à comprendre comme l'information fournie par la variable aléatoire  $X$ ). On fera également quelques abus afin d'éviter d'introduire trop de théorie.

### 1.3.2 Définition/Propriétés

**Définition 1.3.1** (et proposition). Soit  $X$  et  $Y$  deux variables aléatoires. Alors, si  $Y$  est intégrable, il existe une fonction  $\phi$  (mesurable) tel que  $U = \phi(X)$  satisfait

$$\mathbb{E}[Yh(X)] = \mathbb{E}[Uh(X)] \quad \text{pour toute fonction } h \text{ mesurable bornée.} \quad (1.3.1)$$

Cette variable  $U$  est unique au sens suivant : s'il existe  $\tilde{U} = \tilde{\phi}(X)$  tel que (1.3.1) est vraie alors  $U = \tilde{U}$  presque sûrement.  $U$  est alors appelée l'espérance conditionnelle de  $Y$  sachant  $X$  et notée :  $U = \mathbb{E}[Y|X]$ .

**Proposition 1.3.2.** Supposons que  $X$  soit une variable aléatoire discrète à valeurs dans  $\mathcal{X}$ . Alors

$$\mathbb{E}[Y|X] = \phi(X) \quad \text{avec} \quad \phi(x) = \mathbb{E}[Y|X=x] = \frac{\mathbb{E}[Y1_{X=x}]}{\mathbb{P}(X=x)}.$$

Preuve : Soit  $h$  une fonction mesurable et bornée. Posons  $U = \phi(X)$

$$\begin{aligned} \mathbb{E}[U \cdot h(X)] &= \mathbb{E}[\phi(X) \cdot h(X)] = \sum_{x \in \mathcal{X}} \phi(x) h(x) \mathbb{P}(X=x) = \sum_{x \in \mathcal{X}} \mathbb{E}[Y \cdot 1_{\{X=x\}}] h(x) \\ &= \sum_{x \in \mathcal{X}} \mathbb{E}[Y \cdot h(X) \cdot 1_{\{X=x\}}] = \mathbb{E}\left[Y \cdot h(X) \sum_{x \in \mathcal{X}} 1_{\{X=x\}}\right] = \mathbb{E}[Y \cdot h(X)] \end{aligned}$$

Donc  $\mathbb{E}[U \cdot h(X)] = \mathbb{E}[Y \cdot h(X)]$

Malheureusement, la formule ci-dessus, bien plus simple que la définition n'a un sens que lorsque que  $\mathbb{P}(X=x) \neq 0$ . La définition générale devient nécessaire lorsque l'on considère des variables continues. Néanmoins, c'est bien la compréhension qu'il faudra conserver de l'espérance conditionnelle dans le cas général. L'espérance conditionnelle est le résultat moyen obtenu en remplaçant la variable aléatoire  $X$  par sa valeur  $x$  lorsque celle-ci est connue. Pour cette raison, on notera  $\phi(x) = \mathbb{E}[Y|X=x]$  dans le cas général.



**Remarque 1.3.2.** Notons que si l'on considère une fonction  $f$ , alors quitte à poser  $\tilde{Y} = f(Y)$ , on peut définir  $\mathbb{E}[f(Y)|X]$ . Par exemple, si l'on pose  $f(y) = 1_{y \in A}$ , et que  $X$  est discrète, alors

$$\mathbb{P}(Y \in A|X) = \phi(X) \quad \text{avec} \quad \phi(x) = \mathbb{P}(Y \in A|X = x).$$

On retrouve alors la probabilité conditionnelle (classique), de l'évènement  $\{Y \in A\}$  conditionnellement à l'évènement  $\{X = x\}$ .

Voici quelques propriétés utiles.

**Proposition 1.3.3.** (a) *L'application  $Y \mapsto \mathbb{E}[Y|X]$  est linéaire.*

(b)  $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$  (L'espérance de l'espérance conditionnelle est égale à l'espérance).

(c) Si  $Y \geq 0$  p.s., alors  $\mathbb{E}[Y|X] \geq 0$ .

(d) Si  $Y$  est  $\sigma(X)$ -mesurable, alors  $\mathbb{E}[Y|X] = Y$ . (En fait, si  $Y$  est  $\sigma(X)$ -mesurable, alors il existe  $\phi$  telle que  $Y = \phi(X)$ ).

(e) Si  $X$  et  $Y$  sont indépendantes, alors  $\mathbb{E}[Y|X] = \mathbb{E}[Y]$ .

Les propriétés (c) et (d) sont des cas particuliers de la propriété suivante (très utile) :

(f) Si  $Y$  est indépendant de  $X$ , alors pour toute fonction  $h$  (telle que  $h(X, Y)$  est intégrable),

$$\mathbb{E}[h(X, Y)|X] = H(X) \quad \text{avec} \quad H(x) = \int_{\mathbb{R}} h(x, y) \mathbb{P}_Y(dy). \quad (1.3.2)$$

(g) Soit  $X_1, X_2, Y$  trois variables aléatoires:

$$\mathbb{E}[Y|X_1] = \mathbb{E}[\mathbb{E}[Y|(X_1, X_2)]|X_1].$$

$\mathbb{P}_Y$  : loi de  $Y$  : "mesure image de  $P$  par  $Y$ "  
*mesure de probabilité sur  $\mathbb{R}$*   
 $\mathbb{P}_Y = \begin{cases} P(Y=y) & \text{si discret} \\ f_Y(y) & \text{si continu} \end{cases}$

**Remarque 1.3.3.** La dernière propriété fait intervenir l'espérance conditionnelle relativement à un couple de variables aléatoires, notion que l'on n'a pas vraiment définie précédemment, mais que l'on peut facilement comprendre comme le résultat moyen obtenu pour  $Y$  connaissant  $X_1$  et  $X_2$ . Cette propriété peut se comprendre comme une propriété de projection. Pour projeter sur le sous-espace le plus petit, on peut d'abord projeter sur l'espace plus gros associé à  $(X_1, X_2)$  puis reprojeter sur l'espace plus petit associé à  $X_1$  (Cette heuristique peut être rendue rigoureuse).

Interpréter les propriétés (d) et (e) en termes d'information :

(d)  $Y = \phi(X) \Rightarrow$  "si je connais  $X$ , je connais  $Y$ "

Dans ce cas la meilleure estimation de  $Y$  sachant  $X$ , c'est  $Y$ .

(e)  $Y \perp\!\!\!\perp X \Rightarrow$  "Si je connais  $X$ , je n'ai pas d'information sur  $Y$ ."

Dans ce cas la meilleure estimation de  $Y$  sachant  $X$ , c'est  $\mathbb{E}[Y]$

Démontrez la propriété (b) lorsque  $X$  est discrète.

Utilisez la propriété (f) pour calculer  $\mathbb{E}[\cos(X+Y)|X]$  lorsque  $X$  et  $Y$  sont indépendantes et que  $Y \sim \mathcal{U}([0, 1])$ .

$$\begin{aligned}\mathbb{E}[\cos(X+Y)|X] &= H(X) \quad \text{avec} \quad H(x) = \mathbb{E}[\cos(X+Y)] \\ H(x) &= \int_0^1 \cos(x+y) \frac{1}{1-0} \mathbb{1}_{\{0 \leq y \leq 1\}} dy = \int_0^1 \cos(x+y) dy = \left[ \sin(x+y) \right]_0^1 \\ &= \sin(x+1) - \sin(x) = \cos\left(2x + \frac{1}{2}\right) \sin\left(\frac{1}{2}\right) \\ \mathbb{E}[\cos(X+Y)|X] &= 2 \cos\left(2x + \frac{1}{2}\right) \sin\left(\frac{1}{2}\right)\end{aligned}$$

### 1.3.3 Loi conditionnelle

On termine la notion de loi conditionnelle qui généralise la notion d'espérance conditionnelle. On commence par le cas simple où  $X$  est discrète.

**Définition 1.3.4.** Soit  $x \in \mathbb{R}$  tel que  $\mathbb{P}(X = x) > 0$ . La loi conditionnelle de  $Y$  relativement à l'événement  $\{X = x\}$  est alors la loi de probabilité  $\mathbb{P}_{Y|X=x}$  définie par : pour tout Borélien  $A$  de  $\mathbb{R}$ ,

$$\mathbb{P}_{Y|X=x}(A) = \mathbb{P}(Y \in A | X = x).$$

Plus généralement, on appelle loi conditionnelle de  $Y$  sachant  $X$ , la famille  $(\mathbb{P}_{Y|X=x})_{x \in \mathcal{X}}$ .

Déterminez la loi conditionnelle de  $Y$  sachant  $X$  de l'exemple introductif.  $Y = XZ$

$Y$  discret  $\Rightarrow$  on regarde simplement les événements  $A = \{y\}$  avec  $y \in \mathcal{Y}$

On a  $\mathbb{P}(Y=y | X=x) = \begin{cases} 1/6 & \text{si } y = kx, \quad k = 1, \dots, 6 \\ 0 & \text{sinon} \end{cases}$

$\Rightarrow \mathcal{L}(Y | X=x) = \mathcal{U}(\{kx, k=1, \dots, 6\})$

Comme l'espérance conditionnelle, cette définition se généralise aux variables continues. On ne détaillera pas ici la construction en rappelant que l'intuition du discret reste valable. Par exemple, si l'on écrit que  $\mathcal{L}(Y|X=x) = \mathcal{N}(x, \sigma^2)$ , cela signifie que lorsque  $X$  est connue et vaut  $x$ , alors la loi de  $Y$  conditionnellement à cette information est une loi normale de moyenne  $x$  et de variance  $\sigma^2$ .

Terminons par une propriété dans le cas où  $(X, Y)$  est un couple de variables continues :

**Proposition 1.3.5.** Supposons que  $(X, Y)$  est un couple de variables continues de densité  $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ . Alors, pour tout  $x \in \mathbb{R}$  tel que  $f_X(x) \neq 0$ , la loi conditionnelle de  $Y$  sachant  $X = x$  est une loi continue de densité  $f_{Y|X=x}$  définie par :

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}. \quad f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy$$

$f_{Y|X}$  est appelée la densité conditionnelle de  $Y$  sachant  $X$ . De plus,

$$\mathbb{E}[Y|X] = \phi(X) \quad \text{avec} \quad \phi(x) = \int y f_{Y|X=x}(y) dy.$$

La dernière propriété s'interprète simplement comme suit : l'espérance conditionnelle de  $Y$  sachant  $X$  est l'espérance de la loi conditionnelle de  $Y$  sachant  $X$ .

#### Proposition 1.3.5 = Interprétation

- $\mathbb{P}(X \in dx) = f_X(x) dx$   $dx = [x-E, x+E] \quad E > 0$   
↑ largeur de l'intervalle
- $\mathbb{P}((X, Y) \in dx dy) = f_{X,Y}(x, y) \cdot dx \cdot dy$
- $\mathbb{P}(Y \in dy | X \in dx) = \frac{\mathbb{P}(Y \in dy, X \in dx)}{\mathbb{P}(X \in dx)} = \frac{f_{X,Y}(x, y) \cdot dx \cdot dy}{f_X(x) \cdot dx}$   

$$= \frac{f_{X,Y}(x, y)}{f_X(x)} dy$$

Si je veux montrer que  $Z$  a une densité  $f$ , je prend  $h$  muette, et :

$$\mathbb{E}[h(Z)] = \dots = \int h(z) f(z) dz$$

Preuve : Soit  $h$  une fonction muette,  $g$  une fonction muette

$$\begin{aligned} \mathbb{E}[h(Y) \cdot g(X)] &= \iint h(y) g(x) f_{X,Y}(x, y) dx dy \\ &= \int \left( \int h(y) \frac{f_{X,Y}(x, y)}{f_X(x)} dy \right) g(x) f_X(x) dx \\ &= \int \psi_h(x) g(x) f_X(x) dx = \mathbb{E}[\psi_h(X) g(X)] \end{aligned}$$

$$\Rightarrow \psi_h(x) = \mathbb{E}[h(Y) | X=x]$$

## Chapitre 2

# Chaînes de Markov

Les <sup>CM</sup>chaînes de Markov, qui constituent de nos jours un inépuisable domaine de recherche, sont très souvent utilisées pour modéliser des phénomènes aléatoires évoluant au cours du temps et comportant une forme de dépendance. Dans ce qui va suivre, nous ne verrons que des chaînes de Markov à valeurs dans un ensemble unidimensionnel au plus dénombrable. Mentionnons tout de même qu'il existe une théorie générale de ces chaînes à valeurs dans des espaces plus riches comme  $\mathbb{R}^d$ .

### 2.1 Un exemple

Considérons deux urnes  $U_1$  et  $U_2$ . Au départ (instant initial) l'urne  $U_1$  contient trois boules blanches et l'urne  $U_2$  trois boules noires. À chaque seconde on extrait une boule prise au hasard dans chaque urne et on les échange. On étudie l'évolution de ce système. À l'occasion de cet exemple, nous présenterons les notations ainsi que le vocabulaire employé.

Nous dirons que le système est dans l'état  $i$  ( $i = 0, \dots, 3$ ) si l'urne  $U_1$  contient  $i$  boules noires. Dans cet exemple l'espace  $E = \{0, 1, 2, 3\}$  est **l'espace des états**. Il est fini de cardinal 4. Le système évolue chaque seconde, nous dirons que le temps est **discret**.

Au temps 0 le système est dans l'état 0, au temps 1 il est dans l'état 1 mais au temps 2 il peut occuper les états 0, 1 ou 2. L'évolution de ce système est donc **aléatoire**, nous appellerons  $X_n$  la variable aléatoire qui prend la valeur  $i$  si le système est dans l'état  $i$  au temps  $n$ . Notre but est d'étudier la suite  $(X_n)$  de variables aléatoires. C'est cette suite que nous appelons Chaîne de Markov.

Nous noterons  $P(i, j) = \mathbb{P}(X_{n+1} = j | X_n = i)$  pour  $i = 0, \dots, 3$  et  $j = 0, \dots, 3$  (probabilité que le système soit en  $j$  au temps  $n+1$  sachant qu'il est en  $i$  au temps  $n$ ). Ces probabilités seront appelées **probabilités de transition** dans la suite. Il est ici facile de calculer la **matrice**  $P = (P(i, j))_{i,j}$ ; on a :

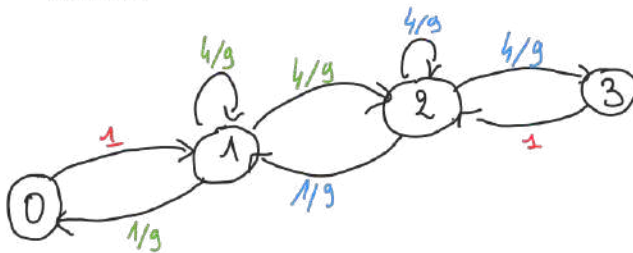
Matrice de transition.

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/9 & 4/9 & 4/9 & 0 \\ 0 & 4/9 & 4/9 & 1/9 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

1. On note que le système est sans mémoire.
2.  $P$  (dite **matrice de transition** de la chaîne) ne dépend pas de  $n$  : la chaîne est dite **homogène**.

On peut aussi représenter le système par le graphe suivant, où chaque sommet correspond à un état, et chaque arête orientée à la probabilité de transition entre deux états :

Graphe. orienté



### Une relation fondamentale

D'après la formule des probabilités totales :

$$\begin{aligned} p_n(i) &= \mathbb{P}(X_n = i) = \sum_{j \in E} \mathbb{P}(X_n = i | X_{n-1} = j) \mathbb{P}(X_{n-1} = j) = \sum_{j \in E} P(j, i) \mathbb{P}(X_{n-1} = j) \\ &= \sum_{j \in E} P(j, i) p_{n-1}(j) = (p_{n-1} P)(i) \Rightarrow p_n = p_{n-1} P = p_0 P^n \end{aligned} \quad (2.1.1)$$

Si on note  $p_n(i) = \mathbb{P}(X_n = i)$  et  $p_n = [p_n(0), \dots, p_n(3)]$ , on remarque que l'opération ci-dessus s'écrit matriciellement :  $p_n = p_{n-1} P$ .  $\mu := p_0$  s'appelle la **loi initiale** de la chaîne. Il s'ensuit que  $p_n = \mu P^n$ . Le calcul de la loi de  $X_n$  peut donc se ramener théoriquement au calcul de  $P^n$ .

Le calcul des valeurs propres de  $P$  donne  $Sp(P) = \left\{1, -\frac{1}{3}, \frac{1}{3}, -\frac{1}{9}\right\}$ . On note que les valeurs propres sont de module inférieur à 1. On peut alors vérifier par calcul que :

$$\lim_{n \rightarrow \infty} p_n = \lim_{n \rightarrow \infty} \mu P^n = \left(\frac{1}{20}, \frac{9}{20}, \frac{9}{20}, \frac{1}{20}\right)$$

On peut en déduire que le nb moyen de boules noires dans l'urne 1 est asymptotiquement égal à :  $0 \times \frac{1}{20} + 1 \times \frac{9}{20} + 2 \times \frac{9}{20} + 3 \times \frac{1}{20} = \frac{3}{2}$

Ceci mérite une petite explication physique. Quoi de plus "naturel" que de penser que le processus mélange les boules entre les deux urnes, et que cela revient à compter le nombre de boules blanches lors d'un tirage sans remise de trois boules prises dans une urne composée de trois boules blanches et de trois boules noires. Or dans ce cas, en appelant  $Z$  la variable aléatoire associée au nombre de boules blanches extraites, on obtient :

$$\mathbb{P}(Z = 0) = \frac{1}{20}, \mathbb{P}(Z = 1) = \frac{9}{20}, \mathbb{P}(Z = 2) = \frac{9}{20}, \mathbb{P}(Z = 3) = \frac{1}{20}.$$

La loi  $\pi = (\frac{1}{20}, \frac{9}{20}, \frac{9}{20}, \frac{1}{20})$  s'appelle **loi stationnaire**, mesure invariante, loi d'équilibre ou loi limite de la chaîne. Notons que si  $p_0 = \pi$  on a :  $\pi P = \pi$ .

Donc  $\pi$  est un vecteur propre à gauche de  $P$  pour la valeur propre 1.

Dans la suite, on va reprendre une à une les notions décrites dans cette introduction.

## 2.2 Introduction aux chaînes de Markov

Dans la suite, l'ensemble  $E$  est appelé espace des états et est supposé au plus dénombrable, c'est-à-dire fini ou dénombrable (en bijection avec  $\mathbb{N}$ ).

**Définition 2.2.1.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a. à valeurs dans  $E$ . Cette suite est une chaîne de Markov (homogène) si pour tous  $n \in \mathbb{N}$  et  $x_0, x_1, \dots, x_{n-1}, x, y \in E$ ,

- (i)  $\mathbb{P}(X_{n+1} = y \mid X_n = x; X_{n-1} = x_{n-1}; \dots; X_0 = x_0) = \mathbb{P}(X_{n+1} = y \mid X_n = x)$  le futur ne dépend du passé que par le présent
- (ii) la probabilité  $\mathbb{P}(X_{n+1} = y \mid X_n = x)$  ne dépend pas de  $n$ . homogénéité  $\mathbb{P}(X_{n+1} = y \mid X_0, \dots, X_n) = \mathbb{P}(X_{n+1} = y \mid X_n)$

Dans ce cas on note  $P(x, y) = \mathbb{P}(X_{n+1} = y \mid X_n = x)$  : c'est la probabilité de transition de l'état  $x$  à l'état  $y$ .

En d'autres termes, si  $X_n$  représente le présent, les  $X_k$  le passé lorsque  $k \in \{0, \dots, n-1\}$ , et le futur quand  $k \geq n+1$ , on dit que "le futur est indépendant du passé, conditionnellement au présent". Par ailleurs, la loi de  $X_0$  est appelée loi initiale de la chaîne.

Dans la suite de ce cours, la suite  $(X_n)_{n \in \mathbb{N}}$  désigne une chaîne de Markov à valeurs dans  $E$  et de matrice de transition  $P$ . Notons que  $P = (P(x, y))_{x, y \in E}$  définit une matrice carrée de taille  $\text{Card}(E)$  (éventuellement infinie si  $E$  n'est pas de cardinal fini) dont chacun des éléments est un nombre compris entre 0 et 1 et vérifiant

$$\sum_{y \in E} P(x, y) = 1, \quad x \in E,$$

somme sur la ligne  $\rightarrow$   $\sum_{y \in E}$   $\uparrow$  ligne colonne

**Définition 2.2.2.** Une matrice carrée  $P$  (ayant un nombre fini ou infini dénombrable de colonnes) telle que

1. Pour tous  $x, y \in E$ ,  $0 \leq P(x, y) \leq 1$ ,



$\lambda = 1$  tjs valeur propre? En effet  $u = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$   $P(u)_i = \sum_{j \in E} P_{i,j} u_j = \sum_{j \in E} P_{i,j} = 1 \Rightarrow Pu = 1u$   
 $\Rightarrow u$  vecteur propre associé à la valeur propre  $\lambda = 1$ .  
 $\forall \lambda$  valeur propre,  $|\lambda| \leq 1$ ? Soit  $u \neq 0$  tq  $Pu = \lambda u \Rightarrow \|Pu\|_\infty = |\lambda| \|u\|_\infty$  Et  $\|u\|_\infty = \max_{j \in E} |u_j|$   
Or  $\forall i \in E$   $(Pu)_i = \sum_{j \in E} P_{i,j} u_j \Rightarrow |(Pu)_i| \leq \sum_{j \in E} P_{i,j} |u_j| \leq \sum_{j \in E} P_{i,j} \|u\|_\infty = \|u\|_\infty \cdot 1$ .  
 $\Rightarrow \|u\|_\infty \geq |\lambda| \|u\|_\infty \Rightarrow |\lambda| \leq 1$  (car  $\|u\|_\infty \neq 0$ )

26

## CHAPITRE 2. CHAÎNES DE MARKOV

$$2. \forall x \in E, \sum_{y \in E} P(x, y) = 1,$$

est appelée matrice stochastique. Toute valeur propre d'une matrice stochastique est nécessairement de module inférieur ou égal à 1. De plus,  $\lambda = 1$  est toujours valeur propre, associée au vecteur propre  $v_1 = [1, \dots, 1]^t$ . **preuve**

Dans la pratique, on représente la chaîne de Markov par un graphe orienté et valué dont les noeuds sont les états de  $E$ . Un noeud  $x$  est relié à un noeud  $y$  par une arête orientée  $(xy)$  si  $P(x, y) > 0$ , et l'arête reçoit la valeur  $P(x, y)$ .

**Proposition 2.2.3.** Notons  $\mu$  la loi initiale de la chaîne. Alors pour tous  $n, k \in \mathbb{N}$  et tous  $x_0, \dots, x_{n+k} \in E$ ,

- (i)  $\mathbb{P}(X_n = x_n; X_{n-1} = x_{n-1}; \dots; X_0 = x_0) = \underbrace{\mu(x_0)}_{P(x_0=x_0)} P(x_0, x_1) \dots P(x_{n-1}, x_n) \rightarrow (x_0, \dots, x_n) \text{ est appelé chemin}$   
(ii)  $\mathbb{P}(X_{n+k} = x_{n+k}; \dots; X_{n+1} = x_{n+1} | X_n = x_n; \dots; X_0 = x_0) = P(x_n, x_{n+1}) \dots P(x_{n+k-1}, x_{n+k})$ .

**Preuve :**

(i)  $\mathbb{P}(X_n = x_n; \dots; X_0 = x_0) = \underbrace{P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0)}_{P(x_{n-1}, x_n) \text{ car c'est une CM}} \cdot \mathbb{P}(X_{n-1} = x_{n-1}, \dots, X_0 = x_0)$   
 $= P(x_{n-1}, x_n) \dots P(x_1, x_2) \cdot \underbrace{P(X_1 = x_1, X_0 = x_0)}_{P(x_0, x_1) P(x_0 = x_0)}$

(ii) (Avec  $k=2$ ).  
 $\mathbb{P}(X_{n+2} = x_{n+2}, X_{n+1} = x_{n+1} | X_n = x_n, \dots) \stackrel{CM}{=} \mathbb{P}(X_{n+2} = x_{n+2}, X_{n+1} = x_{n+1} | X_n = x_n)$   
 $= \mathbb{P}_{\{X_n = x_n\}}(X_{n+2} = x_{n+2}, X_{n+1} = x_{n+1}) \quad \text{ou} \quad P_n(\cdot) = P(\cdot | A)$   
 $= \mathbb{P}_{\{X_n = x_n\}}(X_{n+2} = x_{n+2} | X_{n+1} = x_{n+1}) \cdot \mathbb{P}_{\{X_n = x_n\}}(X_{n+1} = x_{n+1})$   
 $= P(x_{n+1}, x_{n+2}) P(x_n, x_{n+1})$

A présent, on définit par récurrence les puissances  $n$ -ièmes de la matrice de transition  $P$ : pour tout  $n \in \mathbb{N}_*$ ,

$$P^n(x, y) = \sum_{z \in E} P(x, z) P^{n-1}(z, y) = \sum_{z \in E} P^{n-1}(x, z) P(z, y), \quad x, y \in E,$$

La probabilité d'aller de  $x$  à  $y$  en  $n$  étapes

avec par convention  $P^0(x, y) = 1_{\{x=y\}}$ . On montre facilement par récurrence que pour tout  $n \in \mathbb{N}$ , la matrice  $P^n$  est stochastique, c'est-à-dire que chaque élément appartient à l'intervalle  $[0, 1]$  et

$$\sum_{y \in E} P^n(x, y) = 1, \quad x \in E.$$

L'intérêt d'introduire ces puissances successives de la matrice  $P$  réside dans le résultat suivant.

**Proposition 2.2.4.** Étant donné  $n \in \mathbb{N}$ , la loi conditionnelle de  $X_n$  sachant  $X_0$  est donnée par la matrice  $P^n$ , i.e.

$$\mathbb{P}(X_n = y | X_0 = x) = P^n(x, y), \quad x, y \in E.$$

On notera parfois dans la suite  $\mathbb{P}_x$  et  $\mathbb{E}_x$  respectivement la probabilité et l'espérance sachant  $X_0 = x$ .

Preuve par récurrence: (pour aller de 0 à  $n+1$ )

•  $n=1$   $\mathbb{P}(X_1=y | X_0=x) \stackrel{\text{def}}{=} P(x,y) = P^1(x,y)$  évident

•  $n \Rightarrow n+1$ :

$$\begin{aligned} \mathbb{P}(X_{n+1}=y | X_0=x) &= \frac{1}{\mathbb{P}(X_0=x)} \mathbb{P}(X_{n+1}=y, X_0=x) = \\ &= \frac{1}{\mathbb{P}(X_0=x)} \sum_{z \in E} \mathbb{P}(X_{n+1}=y, X_n=z, X_0=x) \\ &= \frac{1}{\mathbb{P}(X_0=x)} \sum_{z \in E} \underbrace{\mathbb{P}(X_{n+1}=y | X_n=z, X_0=x)}_{P(z,y)} \underbrace{\mathbb{P}(X_n=z | X_0=x)}_{P^n(x,z) \text{ par HR}} \cdot \cancel{\mathbb{P}(X_0=x)} \\ &= \sum_{z \in E} P^n(x,z) P(z,y) = P^{n+1}(x,y) \end{aligned}$$

Formule générale:  $P^n(x,y) = \sum_{x_1, \dots, x_{n-1} \in E} P(x, x_1) \dots P(x_{n-1}, y) \rightarrow \left\{ \begin{array}{l} \text{on somme sur tous} \\ \text{les chemins possibles} \end{array} \right.$

Étant donnée une probabilité  $\pi$  sur  $E$ , on définit la quantité

$$\pi P(y) = \sum_{x \in E} \pi(x) P(x, y), \quad y \in E.$$

On vérifie que  $\pi P$  est bien une probabilité sur  $E$ , qui est reliée à la loi de la chaîne de la manière suivante.

**Proposition 2.2.5.** Si  $X_0$  suit la loi  $\pi$  alors pour tout  $n \in \mathbb{N}$  la loi de  $X_n$  est donnée par  $\pi P^n$ .

Preuve:  $\forall y \in E \quad \mathbb{P}(X_n=y) = \sum_{x \in E} \mathbb{P}(X_n=y | X_0=x) \mathbb{P}(X_0=x)$

$$= \sum_{x \in E} P^n(x,y) \pi(x) = (\pi P^n)(y)$$

$$\pi = [\pi(0) \quad \dots \quad \pi(k)] \quad \text{si } E = \{0, \dots, k\}$$

Ainsi, si  $\text{card}(E)$  est fini, alors pour déterminer la loi de  $X_n$  il suffit de calculer la matrice itérée  $P^n$ , donc la diagonaliser.

## 2.3 Classification des chaînes de Markov

**Définition 2.3.1.** On dit qu'un état  $x$  mène à un état  $y$ , et l'on note  $x \rightsquigarrow y$ , s'il existe  $n \in \mathbb{N}$  tel que  $P^n(x, y) > 0$ . Autrement dit, partant de l'état  $x$ , la chaîne peut atteindre  $y$  en  $n$  étapes.



**Proposition 2.3.2.** On définit une relation binaire sur  $E$  de la manière suivante: pour tous  $x, y \in E$ ,

$$x \sim y \iff x \rightsquigarrow y \text{ et } y \rightsquigarrow x. \quad \text{i.e. } \exists n_0, n_1 \in \mathbb{N} \text{ tq } \begin{cases} P^{n_0}(x, y) > 0 \\ P^{n_1}(y, x) > 0 \end{cases}$$

Cette relation de communication entre états est une relation d'équivalence sur  $E$ . Ainsi, l'ensemble des classes d'équivalence (mentionnées comme classes, ou classes de communication dans la suite) est une partition de l'espace des états  $E$ .

**Preuve :** " $\sim$ " Relation d'équivalence

- $\sim$  Réflexive car par définition  $x \rightsquigarrow x$  puisque  $P^0(x, x) = 1 > 0$
- $\sim$  Symétrique car  $x \sim y \iff \begin{cases} x \rightsquigarrow y \\ y \rightsquigarrow x \end{cases} \iff \begin{cases} y \rightsquigarrow x \\ x \rightsquigarrow y \end{cases} \iff y \sim x$
- $\sim$  Transitive si  $x \sim y$  et  $y \sim z$

•  $x \rightsquigarrow z$  ? :  
 Par définition  $\begin{cases} x \rightsquigarrow y \\ y \rightsquigarrow z \end{cases} \Rightarrow \begin{cases} \exists n_0 \in \mathbb{N} \text{ tq } P^{n_0}(x, y) > 0 \\ \exists n_1 \in \mathbb{N} \text{ tq } P^{n_1}(y, z) > 0 \end{cases}$   
 $\Rightarrow P^{n_0+n_1}(x, z) \geq P^{n_0}(x, y) P^{n_1}(y, z) > 0$   
 $\{ \text{chemin qui va de } x \text{ à } z \text{ en } n_0+n_1 \text{ étapes} \} \supset \{ \text{chemin qui va de } x \text{ à } z \text{ en } n_0+n_1 \text{ étapes} \text{ en passant par } y \text{ à l'étape } n_0 \}$

•  $z \rightsquigarrow x$  ? MÊME CHOSE  $\Rightarrow$  la relation est d'équivalence.

Notons qu'il est possible de sortir d'une classe. Par contre, une fois sorti, il est impossible d'y revenir, sinon cela signifie que l'on n'en serait pas sorti !! Le résultat suivant nous permet d'exhiber un critère simple pour montrer qu'un état mène à un autre état.

**Proposition 2.3.3.** Un état  $x$  mène à un autre état  $y$  si et seulement s'il existe  $n \in \mathbb{N}_*$  et  $x_1, x_2, \dots, x_{n-1} \in E$  tels que

$$P(x, x_1) P(x_1, x_2) \dots P(x_{n-1}, y) > 0.$$

En particulier, si  $A$  désigne une partie de  $E$ , alors, si pour tous  $x \in A$ , pour tous  $y \in A^c$ ,  $P(x, y) = 0$ ,  $A$  et  $A^c$  ne peuvent communiquer (au sens, il n'existe pas d'éléments  $x$  de  $A$  et  $y$  de  $A^c$  tels que  $x$  communique avec  $y$ ).

*Preuve.* Immédiate en remarquant que pour tout  $n \in \mathbb{N}$ ,

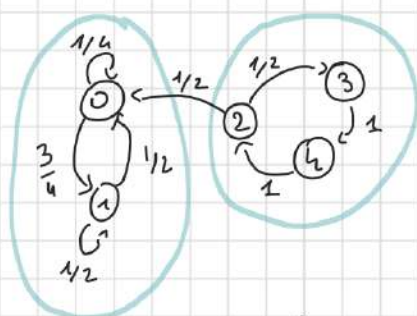
$$P^n(x, y) = \sum_{x_1, \dots, x_{n-1} \in E} P(x, x_1) P(x_1, x_2) \dots P(x_{n-1}, y).$$

■

Ainsi, pour montrer qu'un état  $x$  mène à  $y$ , il suffit de trouver un chemin d'arêtes sur le graphe associé, partant de  $x$  et arrivant en  $y$ .

**Définition 2.3.4.** Une classe de communication  $C$  est dite **fermée** si pour tout  $x \in C$  tel que  $x \rightsquigarrow y$ , alors  $y \in C$ . Dans le cas contraire, elle est dite **ouverte**. i.e.  $\exists x \in C$  et  $y \in C^c$  tq  $x \rightsquigarrow y$   
 En fait :  $C$  ouverte  $\iff \exists x \in C \exists y \in C^c$  tq  $P(x, y) > 0$

Exemple = Chaîne de graphe



$$0 \sim 1 \text{ car } \begin{cases} P(0,1) = 3/4 > 0 \\ P(1,0) = 1/4 > 0 \end{cases}$$

De plus  $A = \{0,1\}$  ne communique pas avec  $\{2,3,4\}$  car  $P(i,j) = 0 \forall i \in \{0,1\} \text{ et } j \in \{2,3,4\} = 0$

$$\Rightarrow C_0 = \{0,1\}$$

$$\begin{aligned} & \cdot 2 \sim 3 \text{ car } \begin{cases} P(2,3) = \frac{1}{2} > 0 \\ P^2(3,2) = P(3,4) \cdot P(4,2) = 1 > 0 \end{cases} \\ & \text{et } 3 \sim 4 \text{ et } 2 \sim 4 \end{aligned}$$

D'où  $C_2 = \{2,3,4\}$  On a ici 2 classes de communications

$C_0$  est fermée car  $\forall i,j \in \{0,1\} \times \{2,3,4\} \quad P(i,j) = 0$

$C_0$  est récurrente

$C_2$  est ouverte car  $P(2,0) > 0 \Rightarrow 2$  mène à 0 pourtant  $0 \notin C_2$

$C_2$  est transitoire

$$E = \mathbb{Z} \quad x \sim x+1 \quad \text{car} \quad P(x+1, x) = p > 0 \quad P(x, x+1) = q > 0$$

Par transitivité, toutes les états communiquent  $\Rightarrow$  chaîne irréductible



### 2.3. CLASSIFICATION DES CHAINES DE MARKOV

29

Un état  $x$  est dit **absorbant** si le singleton  $\{x\}$  est une classe fermée, i.e.  $P(x, x) = 1$ .

La chaîne de Markov est dite **irréductible** si tous les états communiquent, c'est-à-dire qu'il n'y a qu'une seule classe de communication : l'espace des états  $E$  tout entier.

Nous allons maintenant introduire les propriétés de classe des chaînes de Markov, à savoir la récurrence et la transience. Au préalable, nous devons définir deux objets très importants, le nombre de visites d'un état et le temps de retour en un état.

**Définition 2.3.5.** Soit  $y \in E$  un état. Le **nombre de visites de  $y$**  par la chaîne est défini par

$$V_y := \sum_{n \in \mathbb{N}} 1_{\{X_n = y\}} = \sum_{n \in \mathbb{N}} 1_{\{y\}}(X_n).$$

Le **temps de retour en  $y$**  de la chaîne est défini par

$$T_y := \inf\{n \geq 1 : X_n = y\},$$

avec la convention  $+\infty$  si  $\inf \emptyset$ .

Notons que l'on parle de temps de retour car  $n$  commence à 1 et non à 0. D'ailleurs on a  $T_y := \inf\{n \geq 0 : X_n = y\}$  dès que  $X_0 \neq y$ . Cette v.a. est un *temps d'arrêt* relativement à la suite croissante en  $n \in \mathbb{N}$  de tribus  $\mathcal{F}_n := \sigma(X_0, X_1, \dots, X_n)$ , c'est-à-dire que pour tout  $n \in \mathbb{N}_*$ ,

$$\{T_y \leq n\} = \bigcup_{k=1}^n \{X_k = y\} \in \mathcal{F}_n.$$

**Définition 2.3.6.** Un état  $y \in E$  est dit **récurent** si  $\mathbb{P}_y(V_y = \infty) = 1$  et **transitoire** si  $\mathbb{P}_y(V_y = \infty) = 0$ . *proba de passer une infinité de fois en  $y$  est égale à 1 (en partant de  $y$ )*

La récurrence traduit la propriété suivante: "partant d'un état, la chaîne  $y$  repasse à coup sûr une infinité de fois" tandis que la transience: "partant d'un état, la chaîne n'y repasse à coup sûr qu'un nombre fini de fois". Remarquons qu'*a priori*, il se peut que  $\mathbb{P}_y(V_y = \infty) \in ]0, 1[$ : un état peut être ni récurrent ni transitoire. On va voir dans la suite que **cette situation est impossible**.

**Définition 2.3.7.** Étant donné  $n \in \mathbb{N}_*$  et un état  $y \in E$ , on définit par récurrence le temps du  $n$ -ième retour en  $y$  de la chaîne par  $\mathbb{P}_y(V_y = \infty) = \mathbb{P}(\forall n, T_y^{(n)} < +\infty)$

$$T_y^{(n)} := \inf\{k \geq T_y^{(n-1)} + 1 : X_k = y\}, \quad \text{avec} \quad T_y^{(0)} := 0 \quad \text{et} \quad T_y^{(1)} := T_y.$$

*Premier temps de retour*

Le prochain résultat est quelque peu technique mais crucial dans ce qui va suivre. Il utilise fortement la propriété de Markov.

**Lemme 2.3.8.** Étant donné un état  $y \in E$ , on a pour tout  $n \in \mathbb{N}_*$ :

$$\mathbb{P}_y(T_y^{(n)} = T_y^{(n-1)} + m \mid T_y^{(n-1)} < \infty) = \mathbb{P}_y(T_y = m), \quad m \in \mathbb{N}_*.$$

$$\Leftrightarrow \mathbb{L}(T_y^{(n)} - T_y^{(n-1)} \mid T_y^{(n-1)} < +\infty) = \mathbb{L}(T_y)$$

Preuve :

$$\begin{aligned}
 \mathbb{P}(T_y^{(n)} - T_y^{(n-1)} = m \mid T_y^{(n-1)} = k) &= \frac{\mathbb{P}(T_y^{(n)} - T_y^{(n-1)} = m, T_y^{(n-1)} = k)}{\mathbb{P}(T_y^{(n-1)} = k)} \\
 &= \frac{\mathbb{P}(X_{m+k} = y, X_{m+k-1} \neq y, \dots, X_{k+1} \neq y, X_k = y, \dots, T_y^{(n-1)} = k)}{\mathbb{P}(T_y^{(n-1)} = k)} \\
 &= \mathbb{P}(A \mid B) \frac{\mathbb{P}(B)}{\mathbb{P}(T_y^{(n-1)} = k)} \\
 &= \underbrace{\mathbb{P}(X_{m+k} = y, X_{m+k-1} \neq y, \dots, X_{k+1} \neq y \mid X_k = y)}_{\mathbb{P}_y(T_y = m)} \times \underbrace{\frac{\mathbb{P}(X_k = y, T_y^{(n-1)} = k)}{\mathbb{P}(T_y^{(n-1)} = k)}}_{=1} = \mathbb{P}_y(T_y = m).
 \end{aligned}$$

Ainsi, lorsque l'on a  $\mathbb{P}_y(T_y^{(n-1)} < \infty) = 1$ , ce qui est vérifié dans le cas récurrent car l'on a pour tout  $k \in \mathbb{N}$ ,

$$1 = \mathbb{P}_y(V_y > k) = \mathbb{P}_y(T_y^{(k)} < \infty),$$

les v.a.  $T_y^{(n)} - T_y^{(n-1)}$ ,  $n \in \mathbb{N}_*$ , ont même loi que  $T_y$ . Avec un peu plus d'effort on peut montrer que les v.a.  $(T_y^{(n)} - T_y^{(n-1)})_{n \in \mathbb{N}_*}$  sont indépendantes, ce qui va nous servir dans la démonstration du théorème ergodique qui viendra plus tard.

Le lemme 2.3.8 va nous permettre de démontrer le résultat suivant, qui relie le nombre de visites d'un état  $y$  par la chaîne avec le temps de retour en ce même état. On note dans la suite

$$f_y := \mathbb{P}_y(T_y < \infty).$$

**Lemme 2.3.9.** *Étant donné un état  $y \in E$ , on a alors pour tout  $n \in \mathbb{N}$ :*

$$\mathbb{P}_y(V_y > n) = f_y^n.$$

Preuve par récurrence :

$$n=0 : P_y(V_y > 0) = 1 = f_y^0$$

$$\begin{aligned} n \Rightarrow n+1 : P_y(V_y > n+1) &= P(T_y^{(n+1)} < +\infty) \\ &= P(T_y^{(n+1)} < +\infty \mid T_y^{(n)} < +\infty) \times P(T_y^{(n)} < +\infty) \\ &= P_y(T_y < +\infty) \times f_y^n \quad (\text{H.R.}) \\ &= f_y \cdot f_y^n \quad \text{d'après le lemme précédent} \\ &= f_y^{n+1} \end{aligned}$$

À présent, nous allons pouvoir énoncer un des théorèmes principaux de cette partie, qui justifie le fait qu'un état ne peut être que récurrent ou transitoire.

**Théorème 2.3.10.** *Étant donné un état  $y \in E$ , on a la dichotomie suivante:*

- (i)  $P_y(T_y < \infty) = 1 \iff y$  est récurrent  $\iff \sum_{n \in \mathbb{N}} P^n(y, y) = \infty$ .
- (ii)  $P_y(T_y < \infty) < 1 \iff y$  est transitoire  $\iff \sum_{n \in \mathbb{N}} P^n(y, y) < \infty$ .

Preuve :

$= E_y[1_{\{V_y \leq n\}}]$  suite croissante

(i)  $f_y = 1 \iff P_y(V_y > n) = 1 \quad \forall n \iff P_y(V_y \leq n) = 0 \quad \forall n \Rightarrow P_y(V_y < +\infty) = 0$   
 $\Rightarrow P_y(V_y = +\infty) = 1 \Rightarrow$  Récurrent par convergence monotone (Beppo-Levi)

$P^n(x, y) = P(X_n = y, X_0 = x)$  et  $P^n(y, y) = P_y(X_n = y) \Rightarrow \sum_{n \geq 0} P^n(y, y) = \sum_{n \geq 0} P_y(X_n = y)$   
 $\sum_{n \geq 0} P^n(y, y) = E_y \left[ \sum_{n \geq 0} 1_{\{X_n = y\}} \right] = E_y[V_y]$   $E_y[1_{\{X_n = y\}}]$

②  $\Rightarrow$  ③ car si  $y$  récurrent alors  $V_y = +\infty \Rightarrow E_y[V_y] = +\infty \Rightarrow \sum_{n \geq 0} P^n(y, y) = +\infty$

③  $\Rightarrow$  ② : ③  $\Rightarrow E_y[V_y] = +\infty \Rightarrow f_y = 1 \Rightarrow y$  est récurrent

(ii) la même chose par symétrie

Dans la démonstration précédente nous avons utilisé le résultat suivant, connu sous le nom de lemme de Wald.

**Lemme 2.3.11.** *Si  $X$  est une v.a. à valeurs dans  $\mathbb{N}$ , alors de Wald*

$$E[X] = \sum_{k \in \mathbb{N}} P(X > k).$$



*Preuve.* Il suffit de remarquer que  $X$  peut s'écrire de manière artificielle comme

$$X = \sum_{k=0}^{X-1} 1 = \sum_{k \in \mathbb{N}} 1_{\{X > k\}},$$

puis de passer à l'espérance des deux côtés (qui sera éventuellement infinie si  $X$  n'est pas intégrable) et d'utiliser le théorème de convergence monotone pour inverser espérance et somme. ■

Comme nous allons le voir, la récurrence et la transience sont des propriétés de classe.

**Théorème 2.3.12.** *Tous les éléments d'une même classe sont de même nature: ils sont tous récurrents ou tous transitoires. On dit alors de cette classe qu'elle est récurrente ou transitoire. i.e.  $x, y \in \mathcal{C}$   $y$  est récurrent  $\Leftrightarrow x$  récurrent.*

Preuve:

$n_0 \in \mathbb{N}$  tq  $P^{n_0}(x, y) > 0$      $n_1 \in \mathbb{N}$  tq  $P^{n_1}(y, x) > 0$

$y \in \mathcal{C}_x$ ,  $\sum_{n \geq 0} P^{n_0+n+n_1}(y, y) \geq \underbrace{P^{n_1}(y, x)}_{>0} \underbrace{\sum_{n \geq 0} P^n(x, x)}_{=+\infty} \underbrace{P^{n_0}(x, y)}_{>0} = +\infty$

$\Rightarrow \sum_{k \geq 0} P^k(y, y) \geq \sum_{n \geq 0} P^{n_0+n+n_1}(y, y) = +\infty$  Donc  $y$  est récurrent

$x$  transitoire  $\Rightarrow y$  transitoire:

$P^{n_0+n+n_1}(x, x) \geq \underbrace{P^{n_0}(x, y)}_{\geq 0} P^n(y, y) \underbrace{P^{n_1}(y, x)}_{\geq 0}$

$\Rightarrow \sum_{n \geq 0} P^n(y, y) \leq \frac{\sum_{n \geq 0} P^{n_0+n+n_1}(x, x)}{P^{n_0}(x, y) P^{n_1}(y, x)} \leq \frac{1}{\underbrace{P^{n_0}(x, y) P^{n_1}(y, x)}_{<+\infty}} \times \sum_{n \geq 0} P^{n_0+n+n_1}(x, x) < +\infty$

$\Rightarrow y$  est transitoire

À présent, comment peut-on établir simplement qu'une classe est récurrente ou transitoire ? La proposition suivante répond à cette question.

**Proposition 2.3.13.** *Soit  $\mathcal{C}$  une classe de communication. Alors les propriétés suivantes sont vérifiées:*

- (i) Si  $\mathcal{C}$  est récurrente, alors elle est fermée. contraposée:  $\{\mathcal{C} \text{ est ouverte} \Rightarrow \mathcal{C} \text{ transitoire}\}$
- (ii) Si  $\mathcal{C}$  est fermée et finie, alors elle est récurrente.

En particulier, si l'espace d'état  $E$  est fini, alors la classe  $\mathcal{C}$  est récurrente si et seulement si elle est fermée.

*Preuve.* Nous n'allons démontrer que le point (i), le point (ii) utilisant la propriété de Markov forte, que nous n'énonçons pas dans ce cours. Pour ce faire, on va utiliser un raisonnement par l'absurde. Ainsi, supposons  $\mathcal{C}$  récurrente et ouverte. Alors il existe un état récurrent  $x \in \mathcal{C}$ , un état  $y \notin \mathcal{C}$  et  $n \in \mathbb{N}$  tels que  $P^n(x, y) > 0$ . Lorsque l'on sort d'une classe, on ne peut y revenir donc on a

$$\begin{aligned} \mathbb{P}_x(\{X_n = y\} \cup \{V_x = \infty\}) &= \mathbb{P}_x(X_n = y) + \mathbb{P}_x(V_x = \infty) - \mathbb{P}_x(X_n = y; V_x = \infty) \\ &= P^n(x, y) + 1 - 0 \\ &> 1, \end{aligned}$$

ce qui est bien évidemment impossible: la classe  $\mathcal{C}$  est donc fermée. ■

Ainsi, un état récurrent ne peut mener qu'à un état de sa classe, donc à un état récurrent. En revanche, un état transitoire peut mener à un autre état transitoire (de sa classe ou pas) comme à un état récurrent.

Pour résumer, on a les équivalences suivantes: si  $x \in E$ ,

- (i)  $x$  est un état récurrent.
- (ii)  $\mathbb{P}_x(V_x = \infty) = 1$ .
- (iii)  $\mathbb{P}_x(T_x < \infty) = 1$ .
- (iv)  $\sum_{n \in \mathbb{N}} P^n(x, x) = \infty$ .

De même, les assertions suivantes sont équivalentes: si  $x \in E$ ,

- (i)  $x$  est un état transitoire.
- (ii)  $\mathbb{P}_x(V_x = \infty) = 0$ .
- (iii)  $\mathbb{P}_x(T_x < \infty) < 1$ .
- (iv)  $\sum_{n \in \mathbb{N}} P^n(x, x) < \infty$ .

On peut aussi démontrer les résultats suivants:

- (i) si  $x$  et  $y$  sont dans la même classe récurrente, alors on a

$$\begin{cases} \mathbb{P}_x(V_y = \infty) &= 1; \\ \mathbb{P}_x(T_y < \infty) &= 1; \\ \sum_{n \in \mathbb{N}} P^n(x, y) &= \infty. \end{cases}$$

- (ii) si  $x$  et  $y$  sont dans la même classe transitoire,

$$\begin{cases} \mathbb{P}_x(V_y = \infty) &= 0; \\ \mathbb{P}_x(T_y < \infty) &< 1; \\ \sum_{n \in \mathbb{N}} P^n(x, y) &< \infty. \end{cases}$$

Enfin, si la chaîne est irréductible alors il n'est pas difficile de montrer que l'on peut enlever le conditionnement relatif à  $X_0$  dans ce qui précède, c'est-à-dire que ces critères sont valables pour n'importe quelle loi initiale.

## 2.4 Probabilité invariante

Étant donnée une chaîne de Markov, on peut se demander si elle a vocation à se stabiliser statistiquement, c'est-à-dire existe-il une probabilité  $\pi$  telle que:

- si  $X_0$  suit la loi  $\pi$  alors  $X_n$  la suit aussi pour tout  $n \in \mathbb{N}$  ?
- la loi de  $X_n$  converge vers  $\pi$  lorsque  $n \rightarrow \infty$  ?

Il s'avère que sous certaines hypothèses, la réponse est oui. Pour étudier ces questions, introduisons tout d'abord la notion de probabilité invariante.

**Définition 2.4.1.** Soit  $\pi$  une mesure sur  $E$ , de masse éventuellement infinie. On dit que  $\pi$  est une mesure invariante pour la chaîne si l'équation suivante est vérifiée:

$$\pi(y) = \pi P(y), \quad y \in E,$$

où l'on rappelle que la définition de  $\pi P$  est donnée par

$$\pi P(y) = \sum_{x \in E} \pi(x) P(x, y), \quad y \in E.$$

De surcroît, si  $\pi$  est une probabilité, on parle de probabilité invariante pour la chaîne.

Dire que  $\pi$  est une probabilité invariante signifie que si  $X_0$  a pour loi  $\pi$ , alors  $X_1$  suit aussi cette loi. On en déduit de proche en proche que pour tout  $n \in \mathbb{N}$ , la v.a.  $X_n$  suit la loi  $\pi$ , i.e. que  $\pi = \pi P^n$ .

**Remarque 2.4.1.** Une mesure invariante est donc un vecteur propre à gauche de la matrice (éventuellement infinie)  $P$  associé à la valeur propre 1. Autrement dit, c'est un vecteur propre à droite de la matrice  $P^T$ . Or, d'après la définition 2.2.2, on sait que  $\underline{1}$  est valeur propre de  $P$ . Ainsi,  $\underline{1}$  est aussi valeur propre de  $P^T$  (lorsque  $E$  est fini, on peut montrer ce résultat en utilisant que  $\det(A) = \det(A^T)$ ) et donc  $P$  admet bien un vecteur propre à gauche. Néanmoins, pour que ça soit une mesure invariante, il faut que ce dernier soit positif. Pour que ça soit une probabilité invariante, il faut en plus que sa norme soit égale à 1...

On va voir dans la suite que sous de bonnes hypothèses, il existe une unique probabilité invariante pour la chaîne. Pour ce faire, on va considérer le nombre de visites d'un état  $y$  avant le retour en un autre état  $x$ :

nombre de visite en  $y$  avant de revenir en  $x$

$$V_y^x := \sum_{n=0}^{T_x-1} 1_{\{X_n=y\}} = \sum_{n=0}^{\infty} 1_{\{X_n=y; T_x > n\}}.$$

On remarque que  $V_x^x = 1_{\{X_0=x\}}$  et donc que  $\mathbb{E}_x[V_x^x] = \mathbb{P}_x(X_0 = x) = 1$ . Étant donné un état  $x$ , notons dans la suite  $\gamma_x$  la mesure sur  $E$  donnée par

nombre moyen de visite de  $y$  avant de revenir en  $x$

$$\gamma_x(y) := \mathbb{E}_x[V_y^x] = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y; T_x > n),$$

quantité qui peut éventuellement être infinie. Commençons par plusieurs lemmes techniques.

$\gamma_x$  : mesure positive sur  $E$  et par définition  $\gamma_x(x) = 1$  tjs !

**Lemme 2.4.2.** On suppose la chaîne de Markov irréductible (une seule classe,  $E$ ) et récurrente (tous les états le sont). Alors pour tout  $x \in E$ , la mesure  $\gamma_x$  est une mesure invariante pour la chaîne.

$$\gamma_x P = \gamma_x$$

**Remarque 2.4.2.** Sur un exemple à deux points  $E = \{0, 1\}$  avec  $P(0, 1) = \alpha$  et  $P(1, 0) = \beta$ , on peut facilement calculer  $\gamma_x$  et constater qu'il s'agit d'une mesure invariante (Faire l'exercice...)



Preuve:  $(X_n)$  I.R. Vérifions que  $\gamma_o = [\gamma_o(0), \gamma_o(1)]$  est invariante

$$\gamma_o(0) = \mathbb{E}_o[V_o^0] = 1 \text{ par définition}$$

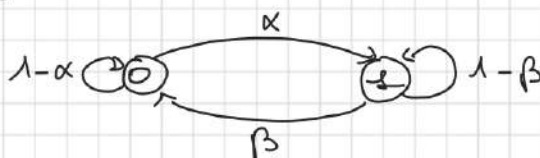
$$\gamma_o(1) = \mathbb{E}_o[V_o^1] = \text{nombre moyen de visites des 1 avant de revenir en 0}$$

$$= 0 \times \underbrace{\mathbb{P}_o(X_1 = 0)}_{1-\alpha} + \underbrace{\text{temps moyen de sortie de 1}}_{\frac{1}{\beta} \text{ car } \tilde{g}(\beta)} \times \underbrace{\mathbb{P}_o(X_1 = 1)}_{\alpha}$$

$$\text{Finalement } \gamma_o = \begin{bmatrix} 1 & \frac{\alpha}{\beta} \end{bmatrix} \quad P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

$$\gamma_o P = \begin{pmatrix} 1-\alpha + \beta \times \frac{\alpha}{\beta} & \alpha + 1-\beta \times \frac{\alpha}{\beta} \end{pmatrix} = \begin{pmatrix} 1-\alpha-\alpha & \alpha + \frac{\alpha}{\beta} - \beta \times \frac{\alpha}{\beta} \end{pmatrix} = \begin{pmatrix} 1 & \frac{\alpha}{\beta} \end{pmatrix}$$

$$\pi = \frac{\gamma_o}{1 + \frac{\alpha}{\beta}} = \left[ \frac{\beta}{\beta + \alpha} \quad \frac{\alpha}{\alpha + \beta} \right] \text{ est la probabilité invariante pour } P.$$



La mesure  $\gamma_x$  étant une mesure invariante pour la chaîne, on a immédiatement que  $\gamma_x = \gamma_x P^n$  pour tout  $n \in \mathbb{N}$ . Ceci va nous servir dans la démonstration du prochain résultat.

I.R

**Lemme 2.4.3.** On suppose la chaîne irréductible et récurrente. Alors pour tout état  $x \in E$ , on a  $0 < \gamma_x(y) < \infty$  pour tout  $y \in E$ .

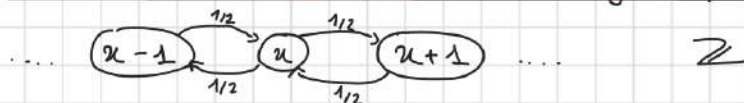
Preuve: Faire en exercice

$$\text{Par l'absurde } \mathbb{E}_x[V_y^n] \Rightarrow \dots \quad \mathbb{P}_x(T_y < T_x) = 0 \dots$$

Impossible car  $x \sim y$

Terminons par un dernier lemme important mais dont la preuve trop technique est admise.

## Exemple de la marche aléatoire symétrique (M.A.S) (I.R).



D'après le lemme 2.4.2, il y a au moins une mesure invariante

Notons  $\gamma$  cette mesure invariante  $\gamma = (\gamma(j))_{j \in \mathbb{Z}}$

$\gamma$ : mesure invariante  $\Leftrightarrow \gamma = \gamma \cdot P$  et  $\gamma(j) \geq 0 \quad \forall j \in \mathbb{Z}$

$$\gamma = \gamma P \Leftrightarrow [\dots \gamma(j) \dots] \begin{bmatrix} 1/2 & 1/2 \end{bmatrix} = [\dots \gamma(j) \dots]$$

$$\Leftrightarrow \forall j \in \mathbb{Z} \quad \gamma(j) = (\gamma P)(j)$$

$$\Leftrightarrow \forall j \in \mathbb{Z} \quad \gamma(j) = \frac{1}{2} \gamma(j-1) + \frac{1}{2} \gamma(j+1)$$

$$\Leftrightarrow \forall j \in \mathbb{Z} \quad \frac{1}{2} (\gamma(j+1) - \gamma(j)) = \frac{1}{2} (\gamma(j) - \gamma(j-1))$$

$$\Leftrightarrow \forall j \in \mathbb{Z} \quad \gamma(j+1) - \gamma(j) = \gamma(j) - \gamma(j-1) = c \in \mathbb{R}$$

• Si  $c = 0$ , on obtient  $\gamma(j) = \gamma(0) \quad \forall j \in \mathbb{Z}$ .

Quitte à renormaliser, il s'agit de la mesure de comptage sur  $\mathbb{Z}$ .

$$\gamma = [\dots 1, \dots 1, \dots 1] = \sum_{k \in \mathbb{Z}} \delta_k \rightarrow \text{symbole de Dirac}$$

• Si  $c \neq 0$   $\gamma(j) = \gamma(0) + cj \quad \forall j \in \mathbb{Z}$

$$\begin{array}{ll} \text{Si } c > 0 & \gamma(j) < 0 \text{ pour } j \in (\mathbb{Z}^+)^* \\ \text{Si } c < 0 & \gamma(j) < 0 \text{ pour } j \in (\mathbb{Z}^-)^* \end{array}$$

Ce ne sont donc pas des mesures (positives).

CCL: à constante multiplicative près, la M.A.S. admet une ! mesure invariante, la mesure de comptage.

(Ca n'est pas transformable en proba!).

Question: Parmi les chaînes irréductibles récurrentes, peut-on caractériser celle qu'ont:

- une mesure invariante de masse infinie  $\xleftrightarrow{\text{suite de l'histoire}} \begin{matrix} \text{récence} \\ \text{nulle} \end{matrix}$
- une proba invariante  $\longleftrightarrow$  récence positive.

**Lemme 2.4.4.** On suppose la chaîne irréductible et récurrente, et qu'il existe  $\lambda$  une mesure invariante pour la chaîne telle que  $\lambda(u) = 1$  pour un certain état  $u \in E$ . Alors les mesures  $\lambda$  et  $\gamma_u$  coïncident.  $\lambda = \gamma_u \iff \forall x \in E \quad \lambda(x) = \gamma_u(x)$

En particulier, ce résultat signifie qu'il existe une mesure invariante à une constante près : soit  $\lambda_1$  une mesure invariante : alors, pour toute mesure invariante  $\lambda$ , il existe  $C > 0$  tel que  $\lambda = C\lambda_1$ . Toute mesure invariante "charge donc tous les points" et si  $\sum_{x \in E} \lambda_1(x) < +\infty$ , il existe une unique probabilité invariante. Revenons à  $T_x$ , le temps de retour en un état  $x$ . On remarque qu'il peut s'écrire en fonction des  $V_y^x$ :

$$T_x = \sum_{n=0}^{T_x-1} 1 = \sum_{n=0}^{T_x-1} \sum_{y \in E} 1_{\{X_n=y\}} = \sum_{y \in E} V_y^x,$$

d'où en passant à l'espérance, on obtient en utilisant le théorème de convergence monotone que

$$\mathbb{E}_x[T_x] < \infty \iff \sum_{y \in E} \mathbb{E}_x[V_y^x] = \sum_{y \in E} \gamma_x(y) < \infty.$$

En particulier, d'après le lemme 2.4.3, une condition suffisante dans le cas irréductible et récurrent pour que  $\mathbb{E}_x[T_x]$  soit finie pour tout état  $x$  est de supposer  $E$  de cardinal fini. Ceci nous amène à la définition suivante.

Les mesures invariantes sont :  $\left\{ \begin{array}{l} - \text{ de masse infinie si } \mathbb{E}_x[T_x] = +\infty \rightarrow \text{récurrent nul} \\ - \text{ de masse finie si } \mathbb{E}_x[T_x] < +\infty \rightarrow \text{récurrent } \oplus \end{array} \right.$

**Définition 2.4.5.** Un état récurrent  $x$  est dit récurrent positif si  $\mathbb{E}_x[T_x]$  est fini, et récurrent nul sinon. Si tous les états sont récurrents positifs (resp. récurrents nuls), on dit que la chaîne est récurrente positive (resp. récurrente nulle).

Notons que dans le cas transitoire on a  $\mathbb{P}_x(T_x < \infty) < 1$ , ce qui entraîne immédiatement que  $\mathbb{E}_x[T_x] = \infty$ .

Si l'on suppose maintenant qu'un état  $x$  est récurrent positif, on définit alors la probabilité  $\pi_x$  sur  $E$  par:

$$\pi_x(y) = \frac{\mathbb{E}_x[V_y^x]}{\mathbb{E}_x[T_x]} = \frac{\gamma_x(y)}{\sum_{y \in E} \gamma_x(y)}, \quad y \in E.$$

Par le lemme 2.4.2, la probabilité  $\pi$  est invariante pour la chaîne sous l'hypothèse supplémentaire de l'irréductibilité. Cependant, en existe-il d'autres ? À présent, nous sommes en mesure de répondre à cette question en énonçant le théorème suivant.

**Théorème 2.4.6.** On suppose la chaîne irréductible et récurrente. Alors les assertions suivantes sont équivalentes:

- (i) tout état de  $E$  est récurrent positif. si  $\mathbb{E}_x[T_x] < +\infty$
- (ii) il existe un état récurrent positif.
- (iii) il existe une unique probabilité invariante  $\pi$ .

Dans ce cas, on a  $\pi = \pi_x$  pour tout  $x \in E$ : on prendra alors

$$\pi(y) = \pi_y(y) = \frac{1}{\mathbb{E}_y[T_y]}, \quad y \in E.$$

*Preuve.*

Notons que le théorème précédent admet une version “réurrence nulle”, dont la démonstration est quelque peu similaire à celle que l’on vient de faire.

**Théorème 2.4.7.** *On suppose la chaîne irréductible et récurrente. Alors les assertions suivantes sont équivalentes:*

- (i) *tout état de  $E$  est récurrent nul.*
- (ii) *il existe un état récurrent nul.*
- (iii) *il existe une mesure invariante  $\lambda$  pour la chaîne, unique à constante multiplicative près, et l’on a pour tout  $x \in E$  fixé,*

$$\lambda(y) = \lambda(x) \gamma_x(y) > 0, \quad y \in E.$$

*De plus,  $\lambda$  est forcément de masse infinie.*

correspond à la M.A.S par exemple : la M.A.S est récurrent nul .  
Et  $\forall x \in \mathcal{X} \quad \mathbb{E}_x[T_x] = +\infty$   
Car sa mesure invariante (mesure de comptage) est de masse infinie

Pour terminer ce paragraphe, résumons dans le théorème qui suit les résultats importants que nous venons de voir jusqu’à présent.

**Théorème 2.4.8.** *On suppose la chaîne de Markov irréductible. Alors les états sont tous transitoires, tous récurrents nuls ou tous récurrents positifs. Et dans ce dernier cas seulement, la chaîne admet une unique probabilité invariante  $\pi$  donnée par*

$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x]}, \quad x \in E.$$

*Enfin, si la chaîne est irréductible sur  $E$  supposé de cardinal fini, alors tous les états sont récurrents positifs et l’existence et l’unicité de  $\pi$  sont assurées.*

## 2.5 Convergence vers la probabilité invariante

Dans le cas irréductible et récurrent positif, on a vu qu'il existait une unique probabilité invariante  $\pi$ , c'est-à-dire une probabilité satisfaisant la propriété suivante: si  $X_0$  suit la loi  $\pi$  alors pour tout  $n \in \mathbb{N}$ , la v.a.  $X_n$  la suit aussi. Cependant, si  $X_0$  est distribuée selon une autre loi, peut-on espérer que la loi de  $X_n$  converge vers  $\pi$ ? Oui, à condition d'éviter un problème de périodicité.

Considérons la matrice de transition triviale  $P$  sur  $E = \{0, 1\}$  donnée par  $P(0, 1) = P(1, 0) = 1$ . Autrement dit, il n'y a pas d'aléa. En calculant les puissances successives de  $P$ , on montre que  $P^{2n} = I_2$  et  $P^{2n+1} = P$ . Ainsi, on remarque que  $P^n(x, y)$  ne peut pas converger lorsque  $n \rightarrow \infty$  à cause de cette périodicité. Donnons à présent la définition de la période d'un état.

**Définition 2.5.1.** On appelle **période** de l'état  $x$  le nombre

$$d_x := \text{pgcd}\{n \in \mathbb{N}_* : P^n(x, x) > 0\}.$$

L'état  $x$  est dit **apériodique** si  $d_x = 1$ .

**Proposition 2.5.2.** Le concept de période est une propriété de classe, c'est-à-dire que si  $x$  et  $y$  sont deux états appartenant à une même classe  $\mathcal{C}$ , alors ils ont même période. On parlera alors de période d'une classe.

**Preuve:** Soit  $x$  et  $y \in \mathcal{C}$ . Montrons que  $d_x$  divise  $d_y$ .  
 Soit  $A_y = \{n \in \mathbb{N}^*, P^n(y, y) > 0\}$  (Par définition  $\forall n \in A_y, d_y | n$ )  
 $\exists n_0, n_1$  tq  $P^{n_0}(x, y) > 0$  et  $P^{n_1}(y, x) > 0$ . Soit  $n \in A_y$   
 $P^{n_0+n+n_1}(x, x) \geq P^{n_0}(x, y) \underbrace{P^n(y, y)}_{>0} P^{n_1}(y, x) > 0 \quad \forall n \in A_y$   
 $d_x | n_0+n+n_1$  or  $d_x | n_0+n_1$  car  $P^{n_0+n_1}(x, x) = P^{n_0}(x, y) P^{n_1}(y, x) > 0$   
 d'où  $d_x | (n_0+n+n_1) - (n_0+n_1)$  c-à-d  $d_x | n \quad \forall n \in A_y$   
 $\Rightarrow d_x$  divise le pgcd de  $A_y$  i.e  $d_x | d_y$   
 De même  $d_y | d_x \Rightarrow d_y = d_x$

On peut maintenant énoncer l'un des derniers résultats fondamentaux de ce cours.

**IRPA**

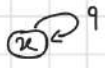
**Théorème 2.5.3.** On suppose la chaîne **irréductible et récurrente positive et apériodique**. Alors la loi de  $X_n$  converge lorsque  $n \rightarrow \infty$  vers l'unique probabilité invariante  $\pi$ , en oubliant sa condition initiale. Autrement dit, pour tout  $x \in E$ ,

$$P^n(x, y) = \mathbb{P}_x(X_n = y) \xrightarrow[n \rightarrow \infty]{} \pi(y), \quad y \in E. \quad \text{i.e.} \quad X_n \xrightarrow[n \rightarrow \infty]{\text{L}} \pi$$

convergence  $\perp$  du point de départ

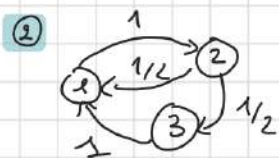


### Exemple :

①  où  $q > 0$

Ici  $P(x, x) = P^1(x, x) = q > 0 \Rightarrow \exists n \in \mathbb{N}^* : P^n(x, x) > 0$

$\Rightarrow d_x \mid 1 \Rightarrow d_x = 1 \Rightarrow x$  apériodique.



On a  $P^2(1, 1) \geq 1 \times \frac{1}{2} > 0$

$P^3(1, 1) \geq P(1, 2) P(2, 3) P(3, 1) = \frac{1}{2} > 0$

$\Rightarrow 2 \text{ et } 3 \in \{n \in \mathbb{N}^* : P^n(1, 1) > 0\}$

$\Rightarrow d_1 \mid 2 \text{ et } d_1 \mid 3 \Rightarrow d_1 = 1 \Rightarrow 1$  est apériodique.

...  $\Rightarrow 2$  et  $3$  sont apériodiques aussi, d'où la propriété suivante.

③ M.A.S  $p \in \{0, 1\}$

On a déjà vu que  $P^{2n+1}(x, x) = 0 \quad \forall n$

$\Rightarrow d_x$  est pair  $\Rightarrow d_x = 2$  car  $P^2(x, x) \geq p(1-p) > 0$

④ Un cas où  $(X_n)$  est récurrent positif mais non apériodique

Cette chaîne est IRP (car irréductible sur  $E$  fini)

Mais de période 2

En effet  $P^k(0, 0) = \begin{cases} 1 & \text{si } k \text{ pair} \\ 0 & \text{si } k \text{ impair} \end{cases}$

Une manière de prouver ce résultat consiste dans un premier temps à montrer qu'une matrice irréductible et apériodique est "fortement irréductible" (réciproque vraie). Ceci signifie qu'il existe un  $n_0$  tel que

$$\min_{x,y \in E} P^{n_0}(x,y) > 0.$$

Nous allons terminer ce cours par l'énoncé de la Loi des Grands Nombres pour les chaînes de Markov, connue sous le nom de théorème ergodique. Dans ce qui suit, la convergence a lieu au sens de la convergence presque sûre, c'est-à-dire pour tout  $\omega$  en dehors d'un certain ensemble de probabilité 0. Notons

nombre de passage en  $x$  entre 0 et  $n-1$

$$V_x^n := \sum_{k=0}^{n-1} 1_{\{X_k=x\}} = \sum_{k=0}^{n-1} 1_{\{x\}}(X_k), \quad x \in E, \quad n \in \mathbb{N}_*,$$

le nombre de visites de l'état  $x$  avant l'instant  $n$ . Si la chaîne est irréductible et transitoire alors la v.a.  $V_x^n$  est majorée par  $V_x$ , qui est une quantité finie, l'état  $x$  étant transitoire. On en déduit alors que

$$\frac{V_x^n}{n} \xrightarrow{n \rightarrow \infty} 0 = \frac{1}{\mathbb{E}_x[T_x]},$$

c'est-à-dire que la proportion de temps passé en l'état  $x$  tend vers 0.

**Théorème 2.5.4.** *On suppose la chaîne irréductible et récurrente, de loi initiale quelconque. Alors pour toute mesure invariante  $\lambda$ , on a*

$$\frac{\sum_{k=0}^{n-1} 1_{\{X_k=y\}}}{\sum_{k=0}^{n-1} 1_{\{X_k=x\}}} = \frac{V_y^n}{V_x^n} \xrightarrow{n \rightarrow \infty} \frac{\lambda(y)}{\lambda(x)} = \gamma_x(y), \quad x, y \in E,$$

mais aussi

$$\frac{V_x^n}{n} \xrightarrow{n \rightarrow \infty} \frac{1}{\mathbb{E}_x[T_x]}, \quad x \in E.$$

0 si récurrent nul car  $\mathbb{E}_x[T_x] = +\infty$

$\frac{T(x)}{\pi(x)}$  si récurrent positif car dans ce cas  $\pi(x) = \frac{1}{\mathbb{E}_x[T_x]}$

Exemple d'application dans le cas récurrent nul.

Si  $(X_n)$  M.A.S alors  $\forall x, y \in \mathbb{Z}$

$$\frac{V_z^n}{V_y^n} \xrightarrow{n \rightarrow \infty} \frac{\lambda(y)}{\lambda(x)} = \frac{1}{1} = 1$$

En travaillant un peu, on peut remplacer dans le résultat précédent les indicatrices  $1_{\{x\}}$  et  $1_{\{y\}}$  par des fonctions  $f$  et  $g$  à valeurs réelles. On obtient ainsi un résultat rappelant celui connu pour les v.a.r. i.i.d., la mesure invariante de la chaîne remplaçant la loi commune du cas i.i.d.

**Théorème 2.5.5.** *On suppose la chaîne irréductible et récurrente, de loi initiale quelconque. Alors pour toute mesure invariante  $\lambda$  et toutes fonctions  $f, g : E \rightarrow \mathbb{R}$  intégrables par rapport à  $\lambda$ , avec de surcroît  $g > 0$ , on a*

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} g(X_k)} = \frac{\sum_{z \in E} f(z) \lambda(z)}{\sum_{z \in E} g(z) \lambda(z)}.$$

*Si la chaîne est récurrente nulle, alors pour toute fonction  $f : E \rightarrow \mathbb{R}$  qui est  $\lambda$ -intégrable,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = 0.$$

*Si la chaîne est récurrente positive, alors pour toute fonction  $f : E \rightarrow \mathbb{R}$  qui est  $\pi$ -intégrable, où  $\pi$  désigne l'unique probabilité invariante,  $\pi(x) = 1/\mathbb{E}_x[T_x]$ , on a*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \sum_{z \in E} f(z) \pi(z).$$



## Chapitre 3

# Simulation et Méthodes de Monte-Carlo

### 3.1 Simulation de variables aléatoires réelles

L'objectif de ce premier chapitre est de présenter les principales méthodes permettant de simuler une variable aléatoire sur un ordinateur. On commencera par rappeler comment "créer de l'aléa raisonnable" sur un ordinateur, ou plus précisément comment (pseudo)-simuler la loi uniforme sur  $[0, 1]$  puis dans un second temps, comment, à partir de ce point de départ, construire des méthodes pour simuler d'autres lois. On parlera d'abord de méthodes générales (inversion, rejet) puis on se concentrera sur quelques méthodes spécifiques.

#### 3.1.1 Génération de nombres pseudo-aléatoires

##### Problématique

Question : Comment générer de l'aléa raisonnable sur un ordinateur ?

D'abord, par aléa raisonnable, on entend l'idée de savoir simuler une loi (*i.e.* tirer un nombre suivant cette loi à l'aide d'un ordinateur) suffisamment riche pour qu'elle soit ensuite utilisable comme point de départ pour en simuler d'autres. Par exemple, simuler une variable de Bernoulli consiste bien à "créer" du hasard mais celui-ci n'est pas raisonnable. Le hasard n'est pas assez riche (seulement deux possibilités).

La variable suffisamment riche la plus simple est la loi uniforme sur  $[0, 1]$ . On a par exemple la propriété suivante :

**Proposition 3.1.1.** *Soit  $X$  une variable aléatoire à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  (ou même dans un espace plus général) de loi notée  $\mathbb{P}_X$ . Notons  $\lambda_{[0,1]}$  la loi uniforme sur  $[0, 1]$  (*i.e.* la mesure de Lebesgue sur  $[0, 1]$ ). Alors, il existe une application mesurable  $\varphi : ([0, 1], \mathcal{B}([0, 1])) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  telle que*

$$\mathbb{P}_X = \lambda \circ \varphi^{-1},$$

*i.e.* pour tout  $A \in \mathcal{B}(\mathbb{R}^d)$ ,

$$\mathbb{P}_X(A) = \mathbb{P}(\varphi(U) \in A) \quad \text{où } U \sim \mathcal{U}_{[0,1]}.$$

En terme de "richesse aléatoire", il "suffit" de générer une loi  $\mathcal{U}_{[0,1]}$

Ce théorème dont nous ne ferons pas la démonstration dans un cadre général (seulement dans le cas réel, voir plus loin) signifie qu’une loi uniforme contient suffisamment d’aléa pour simuler n’importe quelle loi. (Notons cependant qu’il s’agit évidemment d’un résultat théorique et que déterminer  $\varphi$  en pratique est un problème compliqué).

On prend donc la loi uniforme comme point de départ et on se donne comme objectif de simuler une suite de réalisations i.i.d de cette loi :

**Définition 3.1.2.** Soit  $(x_n)_{n \geq 1}$  une suite de nombres à valeurs dans  $[0, 1]$ . Alors, on dit que  $(x_n)$  est une suite de nombres aléatoires s’il existe un espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$  et une suite  $(U_n)_{n \geq 1}$  de variables i.i.d sur cet espace de loi uniforme sur  $[0, 1]$  telle que  $x_n = U_n(\omega)$  pour tout  $n \geq 1$  où  $\omega \in \Omega$ .

Cette définition n’est en fait pas consistante car tout évènement est défini  $\mathbb{P}$  – p.s.. Néanmoins, cela signifie que l’on souhaite que la suite soit une réalisation de cette suite.

### Nombres pseudo-aléatoires

Dans la réalité, on n’aura pas possibilité de générer une approximation d’une suite de nombres aléatoires qu’on appellera suite de nombres pseudo-aléatoires. L’idée est la suivante : on construit une suite déterministe mais relativement *chaotique* en cherchant à minimiser le coût de calcul de cette suite. Le procédé standard consiste à considérer un nombre  $N$  “grand” et à définir  $(x_n)$  par :

$$x_n = \frac{y_n}{N}, n \geq 0$$

où  $(y_n)_{n \geq 0}$  est défini récursivement

$$y_{n+1} = (ay_n + b) \bmod N$$

avec  $a$  et  $b$  entiers compris entre 0 et  $N$  (voir plus bas pour le choix de  $a$ ). La suite  $(y_n)$  est définie de manière déterministe excepté la valeur initiale  $y_0$  que l’on appelle la *graine aléatoire* et qui est souvent construite à partir de l’heure à laquelle est lancé le programme. Nous ne détaillerons pas plus ici la construction de la graine aléatoire. On gardera simplement en mémoire l’idée suivante : on peut créer de l’aléa en lançant le programme.

Revenons un instant sur le choix de  $a$  et  $b$ . Supposons pour simplifier que  $b = 0$ . Dans ce cas, on s’appuie sur la proposition suivante :

**Proposition 3.1.3.** Si  $N$  est premier, alors  $(\frac{\mathbb{Z}}{N\mathbb{Z}^*}, \times)$  est un groupe monogène, i.e. engendré par un seul élément:

$$\frac{\mathbb{Z}}{N\mathbb{Z}^*} = \{\bar{a}^k, k = 1, \dots, N-1\}.$$

On rappelle que  $\bar{a}$  est alors appelé un générateur du groupe.

En particulier, la propriété ci-dessus implique que tous les  $\bar{a}^k$ ,  $k = 0, \dots, N-1$  sont distincts. On constate donc que la suite  $(x_n)_{n \geq 1}$  ainsi définie chargera de manière uniforme tous les points de l’ensemble  $\{1/N, \dots, \frac{N-1}{N}\}$ . La question de l’indépendance semble plus compliquée. Sans rentrer dans les détails, remarquons simplement qu’il semble naturel de choisir  $a/N$  grand pour rendre “non prévisible” le résultat de la  $(N+1)$ -ième simulation connaissant les précédents.

Ce sujet n'étant pas l'objectif principal fixé par le cours, nous nous limiterons ici à ces considérations imprécises sur les nombres pseudo-aléatoires et supposerons à présent que la simulation d'une suite de variables aléatoires de loi uniforme sur  $[0, 1]$  est possible. On s'intéresse alors ci-dessous à plusieurs méthodes générales permettant de simuler d'autres lois à partir de la loi uniforme. Pour commencer, considérons le cas relativement simple des variables aléatoires discrètes :

### 3.1.2 Simulation de variables aléatoires discrètes

On a la proposition suivante :

**Proposition 3.1.4.** Soit  $X$  à valeurs dans  $\mathbb{R}^d$  de loi  $\mu = \sum_{i=1}^N p_i \delta_{x_i}$  (i.e.,  $\mathbb{P}(X = x_i) = p_i$ ). Alors,  $X$  peut être représentée de la manière suivante :

*Donc on sait simuler une v.a. discrète de manière exacte.*

$$X = \sum_{i=1}^N x_i 1_{P_{i-1} \leq U < P_i}$$

où  $U \sim \mathcal{U}_{[0,1]}$ ,  $P_0 = 0$  et pour tout  $i \in \{1, \dots, N\}$ ,  $P_i = \sum_{k=1}^i p_k$ .

Cette représentation fournit en particulier un algorithme simple pour simuler la loi  $\mu$ .

**Exemple 3.1.1.** Si  $\mu = p\delta_a + (1-p)\delta_b$ , alors, on simule  $\mu$  de la manière suivante. On simule une loi uniforme  $U$  sur  $[0, 1]$  puis on pose

$$X = \begin{cases} a & \text{si } U < p \\ b & \text{si } U > p. \end{cases}$$

La simulation de variables aléatoires discrètes peut être vue comme un cas particulier de la méthode par inversion que l'on présente maintenant.

### 3.1.3 Méthode par inversion (de la fonction de répartition).

On suppose dans cette partie que l'on cherche à simuler une variable aléatoire réelle  $X$  de fonction de répartition  $F$ . On rappelle que  $F$  est croissante, continue à droite, admettant des limites à gauche.

On définit alors la pseudo-inverse de  $F$  par :

**Définition 3.1.5.** La pseudo-inverse (à gauche) de  $F$  est définie par :

$$\forall t \in ]0, 1[, \quad F^{\leftarrow}(t) = \inf\{z \in \mathbb{R}, F(z) \geq t\}.$$

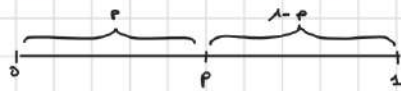
**Exemple 3.1.2.** Calculer la fonction pseudo-inverse relative à la loi de Bernoulli de paramètre  $p$ .

Notons que  $F^{\leftarrow}$  est toujours croissante car  $F$  l'est. Le cas le plus agréable est celui où  $F$  est continue strictement croissante. En effet, on a alors la propriété suivante :

**Proposition 3.1.6.** Supposons que  $F$  soit strictement croissante continue sur  $\mathbb{R}$ . Alors,  $F^{\leftarrow} = F^{-1}$  où  $F^{-1}$  est la fonction réciproque de  $F$ . Plus généralement, s'il existe  $a, b \in \mathbb{R}$  tels que  $a < b$ , et  $F$  réalise une bijection de  $]a, b[$  sur  $]0, 1[$ , alors  $F^{\leftarrow} = F^{-1}$  où  $F^{-1}$  désigne la réciproque de  $F$  restreinte à l'intervalle  $]a, b[$ .

### 3.1.2. Simulation de variables aléatoires discrètes

$X$  à valeurs dans  $\{a, b\}$



$$U \sim U_{[0,1]}$$

$$P(X=a) = p \quad \text{et} \quad P(X=b) = 1-p$$

$$X = \begin{cases} a & \text{si } u \leq p \\ b & \text{si } u > p \end{cases}$$

Lemme :  $X$  a pour loi  $\begin{cases} P(X=a) = p \\ P(X=b) = 1-p \end{cases}$

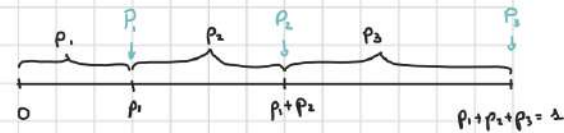
Preuve :  $P(X=a) = P(U \leq p) = p$ . car  $P(x_1 \leq U \leq x_2) = x_2 - x_1$  si  $0 \leq x_1 < x_2 \leq 1$ .  
 $\left( = \int_{x_1}^{x_2} 1 \cdot dy \right)$

Avec 3 "valeurs" :

$$P(X=x_1) = p_1$$

$$P(X=x_2) = p_2$$

$$P(X=x_3) = p_3$$



$$X = \begin{cases} x_1 & \text{si } u \leq p_1 \\ x_2 & \text{si } p_1 < u \leq p_1 + p_2 \\ x_3 & \text{si } p_1 + p_2 < u \leq 1 \end{cases}$$

Ici  $\phi(u) = \begin{cases} x_1 & \text{si } u \leq p_1 \\ x_2 & \text{si } p_1 < u \leq p_1 + p_2 \\ x_3 & \text{si } p_1 + p_2 < u \leq 1 \end{cases}$

Pseudo-programme :

```
def Discrete ([p1, pn], [x1, xn]) :
    p = p1
    u = rand() (tirage entre 0 et 1)
    i = 1
    while u > p :
        p = p + pi
        i = i + 1
    return xi
```

Temps de calcul :

$$E[T] \text{ où } T_n := \inf \{ n \geq 1, u \leq p_n \}$$

$$P(T=1) = P(0 \leq u \leq p_1) \quad \dots \quad P(T=k) = P(p_{k-1} \leq u \leq p_k) = p_k - p_{k-1} = p_k$$

$$\Rightarrow E[T] = \sum_{k=1}^N k P(T=k) = \sum_{k=1}^N k p_k$$

Remarque :

Pour optimiser le temps de calcul, le mieux est de réordonner les  $p_k$  dans l'ordre décroissant :

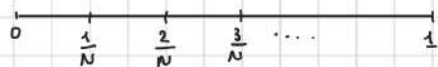
$$p_1 = 0,2 \quad p_2 = 0,2 \quad p_3 = 0,6$$

Avec l'algo standard :  $E[T] = 1 \times 0,2 + 2 \times 0,2 + 3 \times 0,6 = 2,4$ .

Ordre décroissant :  $\tilde{p}_1 = p_3 \quad \tilde{p}_2 = p_2 \quad \tilde{p}_3 = p_1$

$$E[\tilde{T}] = 1 \times 0,6 + 2 \times 0,2 + 3 \times 0,2 = 1,6$$

Exemple : Cas de la loi uniforme sur  $\{x_1, \dots, x_N\}$



$$X = \begin{cases} x_1 & \text{si } u \leq \frac{1}{N} \\ x_2 & \text{si } \frac{1}{N} < u \leq \frac{2}{N} \\ \vdots & \vdots \\ x_N & \text{si } \frac{N-1}{N} < u. \end{cases}$$

$$p_1 = \dots = p_N = \frac{1}{N}$$

Avec la méthode "standard" :  $E[T] = \sum_{k=1}^N k \cdot \frac{1}{N} = \frac{1}{N} \cdot \frac{N(N+1)}{2} = \frac{N+1}{2}$

Si on pose  $X = x_{[Nu]+1}$  alors  $X \sim \mathcal{U}(\{x_1, \dots, x_N\})$ .

$$P(X = x_k) = P([Nu] + 1 = k) = P(k-1 \leq Nu < k) = P\left(\frac{k-1}{N} \leq u < \frac{k}{N}\right) = \frac{1}{N}$$

Par cette méthode, le coût du calcul est "de l'ordre de 1"

### 3.1.3. Méthode par inversion

"MORALEMENT",  $X = F^{-1}(u)$  suit une loi de fonction de répartition  $F$  <sup>loi uniforme</sup>

Problèmes =  $F^{-1}$  n'est pas bjs bien définie  
 $F^{-1}$  n'est pas tjrs connue

Résultat simple (et utile en pratique) :

→  $X$  a une fonction de répartition :  $F_X : F_X(t) = P(X \leq t)$  telle que  $F_X$  réalise une bijection de  $I \subset \mathbb{R}$  sur  $J = ]0, 1[$

Dans ce cas  $Z = F_X^{-1}(u)$  a pour fonction de répartition  $F_X$  (a même loi que  $X$ )

Preuve :  $t \in \mathbb{R}$

$$F_Z(t) = P(Z \leq t) = P(F_X^{-1}(u) \leq t) = P((F_X \circ F_X^{-1})(u) \leq F_X(t)) = P(u \leq F_X(t)) = F_X(t)$$

$$\Rightarrow F_Z = F_X \Rightarrow X \stackrel{L}{=} Z$$

$$P(u \leq k) = k \text{ si } u \sim \mathcal{U}(a, b)$$

Exemple :  $\mathcal{M}_\lambda \quad -\frac{1}{\lambda} \log u \sim \mathcal{E}(\lambda)$ .

Prop : Si  $F$  réalise une bijection de  $I \subset \mathbb{R}$  dans  $J = ]0, 1[$ , alors  $F^{-1}(u)$  a pour fonction de répartition  $F$  où  $u \sim \mathcal{U}_{[0,1]}$

$$F_\lambda(t) = 1 - e^{-\lambda t} \text{ où } X \sim \mathcal{E}(\lambda) \text{ avec } f_\lambda(t) = \lambda e^{-\lambda t} \mathbb{1}_{t \geq 0}$$

$$I = ]0, +\infty[ \quad J = ]0, 1[$$

$F_\lambda$  est strictement croissante (car  $F_\lambda'(t) = e^{-\lambda t} > 0$  sur  $I$ ) et continue sur  $I$  et  $F_\lambda(]0, +\infty[) = ]0, 1[$ . Donc  $F_\lambda$  réalise une bijection de  $I$  sur  $J$ , donc admet une réciproque  $F_\lambda^{-1} : J \rightarrow I$  tq :

$$\forall x, y \in I \times J \quad y = F_\lambda(x) \Leftrightarrow x = F_\lambda^{-1}(y).$$

Déterminons  $F_{\lambda}^{-1}(y) =$  Soit  $(x, y) \in I \times ]$

$$y = F_{\lambda}(x) \Leftrightarrow y = 1 - e^{-\lambda x} \Leftrightarrow e^{-\lambda x} = 1 - y \Leftrightarrow x = -\frac{1}{\lambda} \ln(1 - y).$$

$$\text{d'où } \forall y \in I \quad F_{\lambda}^{-1}(y) = -\frac{1}{\lambda} \ln(1 - y).$$

D'après le résultat du cours  $-\frac{1}{\lambda} \ln(1 - U) \sim \mathcal{E}(\lambda) \quad (U \sim \mathcal{U}_{[0,1]})$ .

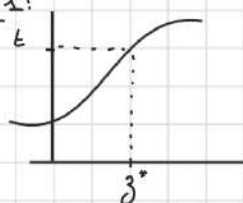
Remarque : En fait,  $1 - U \sim U$  donc  $\frac{1}{\lambda} \ln(U) \sim \mathcal{E}(\lambda)$ .

$$\begin{aligned} F_{1-U}(t) &= \mathbb{P}(1 - U \leq t) = \mathbb{P}(U \geq 1 - t) = 1 - \mathbb{P}(U \leq 1 - t) = 1 - (1 - t) = t \\ &= \mathbb{P}(U \leq t) \quad \text{pour } U \sim \mathcal{U}_{[0,1]} \\ &= F_U(t). \end{aligned}$$

Définition 3.1.5 :

$$F^{\leftarrow}(t) = \inf \{ z \in \mathbb{R}, F(z) \leq t \} \rightarrow \text{c'est un quantile}$$

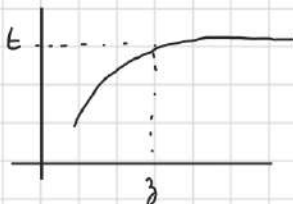
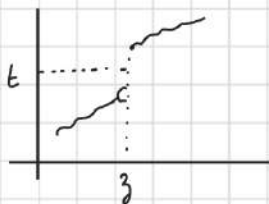
Ex 1 :



$$A = \{ z \in \mathbb{R}, F(z) \geq t \} = [z^*, +\infty[ \Rightarrow F^{\leftarrow}(t) = z^*$$

Si  $F$  strict.  $\mathcal{C}$  et continue sur  $\mathbb{R}$  alors  $F^{\leftarrow} = F^{-1}$

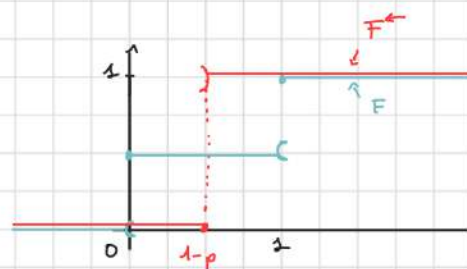
Ex 2 :



$$F^{\leftarrow}(t) = z.$$

Ex 3 :

$$X \sim \mathcal{B}(p) \quad F_X(t) = \begin{cases} 0 & t < 0 \\ \mathbb{P}(X=0) = 1-p & 0 \leq t < 1 \\ (1-p) + \mathbb{P}(X=1) = 1 & t \geq 1 \end{cases}$$



$$F^{\leftarrow}(t) = \inf \{ z, F(z) \geq t \}$$

$$t \in ]p, 1[ \quad \forall t \in ]0, 1-p[ \quad F^{\leftarrow}(t) = [0, +\infty[ = 0$$

$$\forall t \in ]1-p, 1[ \quad F^{\leftarrow}(t) = 1.$$

$$\text{Conclusion : } F^{\leftarrow}(t) = 1 \text{ if } t > 1-p \quad \forall t \in ]0, 1[$$



La généralisation est notamment pratique quand la variable aléatoire n'est pas à support dans  $\mathbb{R}$  tout entier. C'est le cas par exemple de la loi exponentielle ou de l'exemple suivant :

**Exemple 3.1.3.** Calculer la pseudo-inverse relative à la loi de densité  $f$  définie par :

$$f(x) = \begin{cases} 1 - |x| & \text{si } x \in [-1, 1] \\ 0 & \text{sinon} \end{cases}$$

La méthode par inversion est alors basée sur le résultat suivant :

**Proposition 3.1.7.** Pour tout  $x \in \mathbb{R}$ ,

$$\{u \in ]0, 1[, F^{\leftarrow}(u) \leq x\} = \{u \in ]0, 1[, F(x) \geq u\}$$

.

La propriété est quasi-évidente sous les hypothèses de la proposition précédente. Démontrons-la dans le cas général :

*Preuve.* Posons  $A = \{u \in ]0, 1[, F^{\leftarrow}(u) \leq x\}$ ,  $B = \{u \in ]0, 1[, F(x) \geq u\}$  et procédons par double inclusion. D'abord, si  $u \in A$ , alors  $F^{\leftarrow}(u) \leq x$ . Si  $F(x) < u$ , alors comme  $F$  est continue à droite, il existe un intervalle  $[x, x+h]$  tel que pour tout  $z \in [x, x+h]$ ,  $F(z) < u$  et dans ce cas, comme  $F$  est croissante, on en déduit que  $F^{\leftarrow}(u) \geq x+h$ . Ceci fournit une contradiction et nous permet de déduire que  $F(x) \geq u$  puis que  $A \subset B$ .

Si maintenant  $u \in B$ , alors  $F(x) \geq u$  et donc  $x \in \{z, F(z) \geq u\}$ . Ainsi, par définition de  $F^{\leftarrow}$ , il est alors clair que  $F^{\leftarrow}(u) \leq x$ . On en déduit que  $B \subset A$ . ■

La méthode par inversion est alors un corollaire de cette proposition :

**Corollaire 3.1.4.** Soit  $\mu$  une loi de probabilité sur  $\mathbb{R}$  de fonction de répartition  $F$ . Si  $U \sim \mathcal{U}_{[0,1]}$ , alors  $F^{\leftarrow}(U)$  suit la loi  $\mu$ .

*Preuve.* Par la proposition précédente,  $\mathbb{P}(F^{\leftarrow}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$ . Ainsi,  $F^{\leftarrow}(U)$  a pour fonction de répartition  $F$ . ■

**Exemple 3.1.5.** Pour une loi exponentielle de paramètre  $\lambda > 0$ ,  $F(t) = (1 - e^{-\lambda t})1_{t>0}$ . Alors, pour tout  $u \in ]0, 1[$ ,

$$F^{\leftarrow}(u) = -\frac{1}{\lambda} \log(1 - u)$$

de sorte que si  $U \sim \mathcal{U}_{[0,1]}$   $X = -\lambda^{-1} \log(1 - U)$  suit la loi exponentielle de paramètre  $\lambda > 0$ . Comme  $U$  a même loi que  $1 - U$ , il en est de même pour  $\tilde{X} = -\lambda^{-1} \log(U)$  (on utilisera la seconde alternative en pratique).

En conclusion de cette section, on peut insister sur le fait que la méthode par inversion est a priori facile à mettre en oeuvre et très peu coûteuse lorsque  $F^{\leftarrow}$  est explicite. Malheureusement, cette propriété nécessite en général une expression explicite de  $F$  ce qui n'est bien sûr pas toujours le cas. On doit alors recourir à d'autres méthodes.

Proposition 3.1.7: 1.  $A \subset B$  ? 2.  $B \subset A$  ?

1. Si  $u \in A$  alors  $F^{\leftarrow}(u) \leq x$   
 $\forall z > F^{\leftarrow}(u) \quad F(z) \geq u$  (car  $F$  croissante).

Comme  $F$  continue à droite, on a forcément,  $F(F^{\leftarrow}(u)) \geq u \Rightarrow F(x) \geq u$  car  $x \geq F^{\leftarrow}(u)$  et  $F \nearrow$

$\Rightarrow A \subset B$   
2.  $u \in B \Rightarrow F(x) \geq u \Rightarrow x \in \{z, F(z) \geq u\} \Rightarrow x \geq \inf \{z, F(z) \geq u\} = F^{\leftarrow}(u) \Rightarrow u \in A$ .

Corollaire 3.1.4 :  $G$  : la fonction de répartition de  $F^{\leftarrow}(U)$ .

$$\forall x \in \mathbb{R}, \quad G(x) = \mathbb{P}(F^{\leftarrow}(U) \leq x) \stackrel{\substack{\uparrow \\ \text{prop. 3.1.7}}}{=} \mathbb{P}(F(x) \geq U) = \mathbb{P}(0 \leq U \leq \underbrace{F(x)}_{\in [0,1]}) = F(x) - 0 = F(x).$$

Ex de la loi Bernoulli  $\text{Ber}(p)$ .

$$X = F^{\leftarrow}(U) = \mathbb{1}_{\{U > 1-p\}} \stackrel{\substack{\sim \\ \text{cor. 3.1.4}}}{\sim} B(p)$$

$$\begin{aligned} \text{En effet, } \mathbb{P}(X=1) &= \mathbb{P}(U > 1-p) = p. \\ \mathbb{P}(X=0) &= 1-p. \end{aligned}$$

### 3.1.4 Méthode du rejet

#### Le cas de la loi uniforme

Supposons que l'on veuille simuler la loi uniforme sur un ensemble  $D$  et que l'on sache la simuler sur un ensemble  $P$  tel que  $D \subset P \subset E$  (Dans la suite, on supposera que  $E = \mathbb{R}^d$  pour simplifier). Alors une manière de simuler  $\mathcal{U}_D$  est de simuler  $\mathcal{U}_P$  loi uniforme sur  $P$  autant de fois que l'on ne tombe pas dans  $D$ . On "jette" donc toutes les simulations de  $\mathcal{U}_P$  antérieures à la première fois où l'on entre dans  $D$ . Plus précisément, on a le résultat suivant :

**Théorème 3.1.8.** *Supposons  $D \subset P \subset \mathbb{R}^d$  et  $\lambda(D) > 0$  (où  $\lambda$  est la mesure de Lebesgue sur  $\mathbb{R}^d$ ). Notons  $(Y_k)_{k \geq 1}$  une suite de variables aléatoires i.i.d. telle que  $Y_1 \sim \mathcal{U}_P$ . Notons  $T$  la variable aléatoire à valeurs dans  $\mathbb{N}^*$  définie par*

$$T := \inf\{k \geq 1, Y_k \in D\}.$$

*Alors,  $T$  suit une loi géométrique de paramètre*

$$p = \mathbb{P}(Y_1 \in D) = \frac{\lambda(D)}{\lambda(P)} \in ]0, 1].$$

*En particulier,  $T$  est finie p.s.. La variable  $X := Y_{T_{\text{fin}}}$  est alors bien définie et satisfait :  $X \sim \mathcal{U}_D$ . De plus,  $X$  et  $T$  sont indépendants.*

**Remarque 3.1.6.** *Cette méthode fournit donc bien un algorithme de simulation de la loi uniforme sur  $D$ . Néanmoins, son coût d'exécution dépend fortement du paramètre  $p$ , i.e. de la taille de  $D$  relativement à celle de  $P$ . Le temps de calcul moyen est en  $1/p = \lambda(P)/\lambda(D)$ .*

*Preuve.* Le fait que  $T$  suive une loi géométrique est laissé en exercice. Pour montrer que  $X$  suit la loi uniforme sur  $D$ , remarquons d'abord que  $\mathcal{L}(Y_1 | Y_1 \in D) = \mathcal{U}_D$ . En effet, pour tout  $A \subset D$ ,

$$\mathbb{P}(Y_1 \in A | Y_1 \in D) = \frac{\mathbb{P}(Y_1 \in A)}{\mathbb{P}(Y_1 \in D)} = \frac{\lambda(A)}{\lambda(P)} \times \frac{\lambda(P)}{\lambda(D)} = \frac{\lambda(A)}{\lambda(D)}.$$

Ainsi, en utilisant que la suite  $(Y_k)_{k \geq 1}$  est i.i.d.,

$$\mathbb{P}(Y_T \in A, T = k) = \mathbb{P}(Y_k \in A, Y_1 \notin D, \dots, Y_{k-1} \notin D) = \frac{\lambda(A)}{\lambda(D)} p (1-p)^{k-1},$$

ce qui implique d'une part que

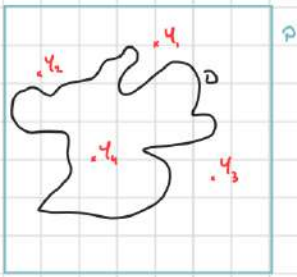
$$\mathbb{P}(Y_T \in A) = \sum_{k \geq 1} \mathbb{P}(Y_k \in A, T = k) = \frac{\lambda(A)}{\lambda(D)} = \mathbb{P}(U_D \in A)$$

où  $U_D$  suit la loi uniforme sur  $D$  puis que pour tous  $A$  et  $k$ ,

$$\mathbb{P}(Y_T \in A, T = k) = \mathbb{P}(Y_T \in A) \mathbb{P}(T = k).$$

Les deux propriétés précédentes prouvent à la fois que  $Y_T$  a la loi voulue et que  $Y_T$  et  $T$  sont indépendantes. ■

### 3.1.4 - Méthode du Rejet



on ne sait pas simuler une loi uniforme sur  $D$ , mais on sait la faire un pavé droit  $[(X, Y) \sim \mathcal{U}^2 \text{ chaque coordonnée.}]$ .

$Y_n \sim \mathcal{U}_P$  on rejette  $y_1, y_2, y_3$ .

$T=4$  (théorème 3.1.8).

Lemme : (Rq 3.1.8).

Si  $P = [a_1, b_1] \times \dots \times [a_d, b_d]$ . Si  $U_1, \dots, U_d$  sont  $\mathcal{U}_{[0,1]}$ , alors :

$Y = (a_1 + (b_1 - a_1)U_1, \dots, a_d + (b_d - a_d)U_d)$  suit la loi  $\mathcal{U}_P$

Preuve : Soit  $h$  muette. (essayez avec  $d=2$ )  $E[h(Y)] = \dots = \int h(y) \frac{1_P(y)}{\lambda(P)} \lambda(dy)$ .

$X = Y_T \sim \mathcal{U}_D$ . De plus  $T \sim \mathcal{G}\left(\frac{\lambda(D)}{\lambda(P)}\right)$  et  $T \perp\!\!\!\perp Y_T$

Algo :

$Y \sim \mathcal{U}_P$

compteur = 1

Tant que  $Y \notin D$  :

$Y \sim \mathcal{U}_P$

compteur += 1

return  $Y$ .

Temps de calcul moyen :  $E[T] = \frac{1}{p}$   $p$  : paramètre de la loi géométrique

Remarque : On a intérêt à avoir le  $\oplus$  petit  $p \supset D$ .

Théorème 3.1.8 : preuve de  $T \sim \mathcal{G}(p)$ . Soit  $k \in \mathbb{N}^*$ .

$$\begin{aligned} P(T=k) &= P(\{Y_1 \notin D\} \cap \dots \cap \{Y_{k-1} \notin D\} \cap \{Y_k \in D\}) \stackrel{!}{=} \prod_{i=1}^{k-1} P(\{Y_i \notin D\}) \times P(\{Y_k \in D\}) \\ &= \left( \prod_{i=1}^{k-1} \frac{\lambda(P \setminus D)}{\lambda(P)} \right) \times \frac{\lambda(D)}{\lambda(P)} = p \cdot \left( \prod_{i=1}^{k-1} \frac{\lambda(P) - \lambda(D)}{\lambda(P)} \right) = p(1-p)^{k-1} \quad \text{Donc } T \sim \mathcal{G}(p) \end{aligned}$$

Etape 2 :  $L(Y|Y \in D) = \mathcal{U}_D$  ?

Soit  $h$  une fonction muette :

$$\begin{aligned} E[h(Y) | Y \in D] &= \frac{1}{P(Y \in D)} E[h(Y) \cdot 1_{\{Y \in D\}}] = \frac{\lambda(P)}{\lambda(D)} \int_{\mathbb{R}^d} h(y) 1_{\{y \in D\}} \frac{1_P(y)}{\lambda(P)} \lambda(dy) \\ &= \int_{\mathbb{R}^d} h(y) \frac{1_{\{y \in D\}}}{\lambda(D)} \lambda(dy) \end{aligned}$$

Etape 3 :  $A \subset D$ .

$$\begin{aligned} P(Y_T \in A) &= \sum_{k=1}^{+\infty} P(Y_T \in A, T=k) = \sum_{k=1}^{+\infty} P(Y_k \in A, T=k) = \sum_{k=1}^{+\infty} P(Y_k \in A, Y_k \in D, Y_{k-1} \notin D, \dots, Y_1 \notin D) \\ &= \sum_{k=1}^{+\infty} P(Y_k \in A, Y_{k-1} \notin D, \dots, Y_1 \notin D) \stackrel{!}{=} \sum_{k=1}^{+\infty} \frac{P(Y_k \in A)}{\frac{\lambda(D)}{\lambda(P)}} \cdot (1-p)^{k-1} = \sum_{k=1}^{+\infty} \frac{\lambda(A)}{\lambda(D)} \cdot \frac{\lambda(D)}{\lambda(P)} \cdot (1-p)^{k-1} \\ &= \frac{\lambda(A)}{\lambda(D)} \cdot \underbrace{\sum_{k=1}^{+\infty} p(1-p)^{k-1}}_1 = \frac{\lambda(A)}{\lambda(D)} = P(X \in A) \quad \text{lorsque } X \sim \mathcal{U}_D. \end{aligned}$$

Remarque : On a aussi montré dans l'étape 3 que  $P(Y_T \in A, T=k) = P(Y \in A | Y \in D) P(T=k)$   
Ce qui prouve l'indépendance entre  $Y_T$  et  $T$ .

**Remarque 3.1.7.** L'indépendance entre  $Y_T$  et  $T$  est assez naturelle. En gros, là où tombe  $Y_T$  (à l'intérieur de  $D$ ) ne dépend pas de l'instant auquel la suite  $(Y_k)_{k \geq 1}$  entre dans  $D$ .

**Remarque 3.1.8.** En pratique,  $P$  peut être un hypercube. En effet, si  $P = \prod_{i=1}^d [a_i, b_i]$ , alors  $Z = (Z_1, \dots, Z_d)$ , où  $Z_i \sim \mathcal{U}_{[a_i, b_i]}$ , et les  $Z_i$  sont indépendants suit la loi uniforme sur  $P$ .

**Remarque 3.1.9.** On verra en TD que la méthode du rejet peut aussi avoir un intérêt dans le cas de variables discrètes, par exemple, pour la simulation du tirage sans remise de  $k$  boules dans une urne en contenant  $N$  ( $N > k$ ).

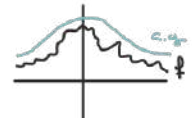
### Cadre général

On cherche à simuler une variable aléatoire à densité, relativement à une mesure donnée (mesure de Lebesgue dans la suite), proportionnelle à une certaine fonction  $f$ . Une généralisation (non triviale) de la méthode précédente va en fait permettre de fabriquer un algorithme de simulation exacte de cette loi.

Dans ce cadre, l'idée de la méthode du rejet est de dominer la loi que l'on cherche à simuler par une autre que l'on sait simuler par un moyen simple (typiquement, par inversion). Plus précisément, on se donne  $f$  et  $g$  deux fonctions positives et mesurables sur  $\mathbb{R}^d$  et on fait les hypothèses suivantes :

(a)  $g$  est une densité de probabilité, i.e.  $\int g(x)\lambda(dx) = 1$  où  $\lambda$  est la mesure de Lebesgue. On note alors  $\mu$  la loi de densité  $g$ .

(b) Il existe  $c > 0$  tel que  $\forall x \in \mathbb{R}^d$ ,  $f(x) \leq cg(x)$ .  $g$  doit être simulable facilement



L'objectif est alors de simuler la loi  $\nu$  de densité  $\frac{f}{\int f(y)\lambda(dy)}$ . On fait donc naturellement l'hypothèse que  $\int f(y)\lambda(dy) > 0$ .

**Remarque 3.1.10.** On n'a pas besoin de connaître  $\int f(y)\lambda(dy)$  pour utiliser cette méthode. En pratique, cela peut avoir un intérêt lorsque l'on ne connaît la densité à simuler qu'à une constante près et que le calcul de  $\int f(y)\lambda(dy)$  est coûteux.

On a alors le théorème suivant :

**Théorème 3.1.9.** Notons  $(U_k)_{k \geq 1}$  et  $(Y_k)_{k \geq 1}$  deux suites indépendantes de variables aléatoires i.i.d. telles que  $U_1 \sim \mathcal{U}_{[0,1]}$  et  $Y_1 \sim \mu$ . Notons  $T$  la variable aléatoire à valeurs dans  $\mathbb{N}^*$  définie par

$$T := \inf\{k \geq 1, cU_k g(Y_k) \leq f(Y_k)\}.$$

Alors,  $T$  suit une loi géométrique de paramètre  $p \in ]0, 1]$  satisfaisant

$$p = \mathbb{P}(cU_1 g(Y_1) \leq f(Y_1)) \stackrel{(*)}{=} \frac{\int f(y)\lambda(dy)}{c}.$$

En particulier,  $T$  est finie p.s.. La variable  $X := Y_T$  est alors bien définie et satisfait :  $X \sim \nu$  de densité  $\frac{f}{\int f}$ . De plus  $X$  et  $T$  sont indépendants.

Exemple:  $f(x) = e^{-x^2/2} \leq c \underbrace{\left( \frac{1}{2} e^{-|x|} \right)}_{g(x)}$  avec  $c$ ?

"g est simulable" On pose  $Y = (2Z - 1)E$  où  $Z \sim \beta\left(\frac{1}{2}\right)$  et  $E \sim \mathcal{E}(\lambda)$ .  
 $Y$  a pour densité  $g$ .

Calcul de  $c$ ,  $\forall x \geq 0$

$$e^{x^2/2 - x} \cdot \frac{c}{2} \geq 1 \Leftrightarrow \ln\left(\frac{c}{2}\right) + \frac{x^2}{2} - x \geq 0 \Leftrightarrow \ln\left(\frac{c}{2}\right) + \frac{1}{2}(x-1)^2 - \frac{1}{2} \geq 0 \Leftrightarrow \ln\left(\frac{c}{2}\right) - \frac{1}{2} \geq 0.$$

$$\Leftrightarrow c \geq 2e^{1/2}$$

Théorème 3.1.9: Application à la simulation de la loi  $\mathcal{N}(0,1)$ .

- $(Y_n)$  suite de v.a de densité  $g(x) = \frac{1}{2}e^{-|x|}$
- $(U_n)$  suite de v.a  $U_{[0,1]}$

Alors si  $T = \inf \left\{ k \geq 1, 2\sqrt{e} U_k \frac{1}{2} e^{-|Y_k|} \leq e^{-\frac{|Y_k|^2}{2}} \right\}$   $Y_T \sim \mathcal{N}(0,1)$ .

Simplification de  $T$ :  $u = U_k$   $y = Y_k$

$$2\sqrt{e} u \frac{1}{2} e^{-|y|} \leq e^{-\frac{y^2}{2}} \Leftrightarrow e^{-y^2/2 + |y|} \geq \sqrt{e} u \Leftrightarrow -\frac{y^2}{2} + |y| \geq \ln(\sqrt{e} u).$$

$$T = \inf \left\{ k \geq 1 \mid |Y_k| + \frac{|Y_k|^2}{2} \geq \ln(U_k) + \frac{1}{2} \right\}$$

Car  $U_0$ :  $f(x) = 1_D(x)$   $g(x) = 1_P(x) \cdot \frac{1}{\text{Vol}(P)}$

$$1_D(x) \leq 1_P(x) \text{ car } D \subset P. \text{ d'où } f(x) \leq \frac{\text{Vol}(P)}{c} \cdot \underbrace{g(x)}_{\text{simulable}}.$$

$$c \text{ et } g(x) \leq f(x) \Leftrightarrow \underbrace{u}_{\in ]0,1[} 1_P(x) < 1_D(x) \Leftrightarrow x \in D \Rightarrow T = \inf \{ n, Y_n \in D \}$$



**Exemple 3.1.11.** Considérons  $f(x) = e^{-\frac{x^2}{2}}$ . On peut vérifier que  $f(x) \leq cg(x)$  avec  $c = 2e^{\frac{1}{2}}$  et  $g(x) = \frac{1}{2}e^{-|x|}$ .  $g$  est une densité de probabilité et il est évident que la condition (c) est satisfaite. Ainsi, on peut utiliser la méthode du rejet pour simuler la loi  $f / \int f(y)dy$  qui n'est autre que la loi normale centrée réduite. Il reste à vérifier que la loi de densité  $g$  est bien simulable facilement. C'est bien le cas car si  $S$  et  $Z$  sont deux variables indépendantes telles que  $S \sim \frac{1}{2}(\delta_1 + \delta_{-1})$  ( $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$ ) et  $Z \sim \mathcal{E}(1)$  alors  $SZ$  a pour densité  $g$  (et  $S$  et  $Z$  sont facilement simulables).

**Remarque 3.1.12.** Le temps de calcul moyen est égal à  $\kappa \mathbb{E}[T]$  où  $\kappa$  est le nombre de calculs par itération de la procédure. Comme

$$\mathbb{E}[T] = \frac{c}{\int f(x)\lambda(dx)},$$

on en déduit donc qu'il est important de minimiser  $c$  pour minimiser le temps de calcul. On remarque également que  $c \geq \int f(x)\lambda(dx)$  (car  $\mathbb{E}[T] \geq 1$ )

**Remarque 3.1.13.** On peut vérifier qu'il s'agit bien d'une généralisation de la simulation de la loi uniforme en posant  $f(x) = \mathbf{1}_D(x)$  et  $g(x) = \frac{1}{\lambda(P)}\mathbf{1}_P(x)$ . On remarque que dans ce cas, la suite de variables  $(U_k)_{k \geq 1}$  ne joue aucun rôle (à vérifier).

*Preuve.*

Étape 1 : Soit  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction mesurable bornée. Alors, si  $U$  et  $Y$  sont deux variables aléatoires indépendantes telles que  $U \sim \mathcal{U}_{[0,1]}$  et  $Y \sim \mu$ , on a :

$$\begin{aligned} \mathbb{E}[\varphi(Y)\mathbf{1}_{cUg(Y) \leq f(Y)}] &= \int_{\mathbb{R}^d} \left( \int_0^1 \varphi(y)\mathbf{1}_{\{cug(y) \leq f(y)\}} du \right) g(y)dy \\ &= \int_{\mathbb{R}^d} \varphi(y) \left( \int_0^{\frac{f(y)}{cg(y)}} du \right) g(y)dy \\ &= \int_{\mathbb{R}^d} \varphi(y) \frac{f(y)}{cg(y)} g(y)dy \\ &= \frac{1}{c} \int_{\mathbb{R}^d} \varphi(y) f(y)dy. \end{aligned}$$

En appliquant avec  $\varphi = 1$ , on en déduit (\*). On a alors :

$$\mathbb{E}[\varphi(Y)/cUg(Y) \leq f(Y)] = \mathbb{E}[\varphi(X)]$$

où  $X \sim \nu$  de sorte que  $\mathcal{L}(Y|cUg(Y) \leq f(Y)) = \nu$ .

Étape 2 : On montre d'abord que  $T$  suit une loi géométrique de paramètre  $p$ . Comme  $\int f d\mu > 0$ , on remarque d'abord que  $p > 0$ . Ensuite, pour tout  $n \geq 0$ ,

$$\mathbb{P}(T > n) = \mathbb{P}(cU_1g(Y_1) > f(Y_1), \dots, cU_ng(Y_n) > f(Y_n)) = \prod_{k=1}^n \mathbb{P}(cU_kg(Y_k) > f(Y_k)) = (1-p)^n$$

d'où le résultat. En particulier,  $T$  est finie p.s.. Montrons maintenant que  $Y_T$  suit la loi  $\nu$  de densité  $f$ . Pour toute fonction  $\varphi$  mesurable et bornée,

$$\mathbb{E}[\varphi(Y_T)] = \sum_{k=1}^{+\infty} \mathbb{E}[\varphi(Y_k)\mathbf{1}_{T=k}]$$

$$\begin{aligned}
&= \sum_{k=1}^{+\infty} \mathbb{E}[\varphi(Y_k) 1_{\{cU_k g(Y_k) \geq f(Y_k)\}} 1_{\{cU_1 g(Y_1) > f(Y_1), \dots, cU_{k-1} g(Y_{k-1}) > f(Y_{k-1})\}}] \\
&= \sum_{k=1}^{+\infty} \mathbb{E}[\varphi(Y_k) 1_{\{cU_k g(Y_k) \leq f(Y_k)\}}] \mathbb{P}(cU_1 g(Y_1) > f(Y_1), \dots, cU_{k-1} g(Y_{k-1}) > f(Y_{k-1})) \\
&= \sum_{k=1}^{+\infty} \mathbb{E}[\varphi(Y_k) | \{cU_k g(Y_k) \leq f(Y_k)\}] p(1-p)^{k-1} \\
&= \sum_{k=1}^{+\infty} \mathbb{E}[\varphi(X)] p(1-p)^{k-1} \quad \text{où } X \sim \nu \\
&= \mathbb{E}[\varphi(X)]
\end{aligned}$$

Dans les équations précédentes, on a d'abord utilisé l'indépendance puis le début de l'étape 2 et enfin l'étape 1. On a donc prouvé que  $Y_T$  a la bonne loi. Pour prouver l'indépendance de  $Y_T$  et  $T$ , il suffit de vérifier que pour tout  $k$  et pour tout  $\varphi$ ,

$$\mathbb{E}[\varphi(Y_T) 1_{T=k}] = \mathbb{E}[\varphi(Y_T)] \mathbb{P}(T = k).$$

Cette propriété a en fait déjà été prouvée plus haut. ■

### 3.1.5 Simulation de variables aléatoires gaussiennes

On a vu précédemment que les variables gaussiennes pouvaient être simulées par rejet. Néanmoins, on préférera dans la pratique une autre méthode : l'algorithme de Box-Muller. Cette méthode est basée sur le résultat suivant :

**Proposition 3.1.10.** *Soit  $R$  et  $\Theta$  deux variables aléatoires indépendantes telles que  $R \sim \mathcal{E}(1/2)$  et  $\Theta \sim \mathcal{U}_{[0, 2\pi]}$ . Alors,*

$$(\sqrt{R} \cos(\Theta), \sqrt{R} \sin(\Theta)) \sim \mathcal{N}(0, I_2).$$

*Preuve.* On rappelle dans un premier temps que la densité de  $\mathcal{N}(0, I_2)$  est donnée par :

$$\left( \int e^{-\frac{x^2}{2}} dx \right)^2 = g(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right). \rightarrow \text{produit des deux densités car } \mathbb{1}$$

Soit  $(X_1, X_2)$  une variable aléatoire de loi  $\mathcal{N}(0, I_2)$ . On a alors pour toute  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  positive mesurable :

$$\mathbb{E}[f(X_1, X_2)] = \int \int \underbrace{f(x_1, x_2)}_{\text{intégrale de } f \text{ contre la densité}} \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) dx_1 dx_2$$

En passant en coordonnées polaires, on obtient :

$$\mathbb{E}[f(X_1, X_2)] = \int_0^{2\pi} \int_0^{+\infty} f(r \cos \theta, r \sin \theta) \frac{1}{2\pi} r \exp\left(-\frac{r^2}{2}\right) dr d\theta. \quad \begin{array}{l} x_1 = r \cos \theta \\ x_2 = r \sin \theta \\ dx_1 dx_2 = r dr d\theta \end{array}$$

En posant  $\rho = r^2$ , on obtient

$$\mathbb{E}[f(X_1, X_2)] = \int_0^{2\pi} \int_0^{+\infty} f(\sqrt{\rho} \cos \theta, \sqrt{\rho} \sin \theta) \frac{1}{2\pi} \frac{1}{2} \exp\left(-\frac{\rho}{2}\right) d\rho d\theta.$$

$\mathcal{N}(m, \Sigma)$  avec  $m \in \mathbb{R}^d$ ,  $\Sigma$  matrice  $d \times d$  SDP.

Rappel:  $(X_1, \dots, X_d)$  vecteur Gaussien de moyenne  $m$  et de matrice de covariance  $\Sigma$  si:  
 $\forall \lambda \in \mathbb{R}^d$   $\sum \lambda_i X_i$  est une v.a. gaussienne et si  $m = \mathbb{E}[X_i]$  et  $\Sigma_{i,j} = \text{Cov}(X_i, X_j)$ .

$$\text{Si } (X_1, X_2) \sim \mathcal{N}(0, I_2) \Rightarrow \begin{cases} \mathbb{E}[X_1] = \mathbb{E}[X_2] = 0 \\ \text{Var}(X_1) = \text{Var}(X_2) = 1 \\ \text{Cov}(X_1, X_2) = 0 \end{cases} \Rightarrow X_1 \perp\!\!\!\perp X_2$$

$\uparrow$   $\Delta$  spécifiques aux vecteurs gaussiens

Proposition 3.1.10:

$$\mathbb{E} = \iint_{\mathbb{R}^2} \frac{1}{r} (r \cos \theta, r \sin \theta) \underbrace{\frac{1}{2\pi} \mathbb{1}_{[0, 2\pi]}(\theta)}_{\text{densité de: } \theta \sim \mathcal{U}_{[0, 2\pi]}} \underbrace{\frac{1}{2} e^{-r/2} \mathbb{1}_{\{r > 0\}}}_{\text{densité de: } R \sim \mathcal{E}\left(\frac{1}{2}\right)} dr d\theta$$

$\perp\!\!\!\perp$

Ainsi,

$$\mathbb{E}[f(X_1, X_2)] = \mathbb{E}[f(\sqrt{R} \cos \theta, \sqrt{R} \sin \Theta)]$$

où  $R$  et  $\Theta$  sont deux variables indépendantes telles que  $R$  suit une loi exponentielle de paramètre  $1/2$  et  $\Theta$  suit la loi uniforme sur  $[0, 2\pi]$ . ■

On peut alors simuler la loi  $\mathcal{N}(0, I_2)$  de la manière suivante :

*Corollaire 3.1.14.* Soit  $(U_1, U_2)$  un couple de variables aléatoires indépendantes telles que  $U_1 \sim \mathcal{U}_{[0,1]}$  et  $U_2 \sim \mathcal{U}_{[0,1]}$ . Alors,

↑  
méthode de  
Box Muller

$$X := \left( \underbrace{\sqrt{-2 \log(U_1)}}_{R \sim \mathcal{E}(\frac{1}{2})} \cos(\underbrace{2\pi U_2}_{\sim [0, 2\pi]}), \sqrt{-2 \log(U_1)} \sin(2\pi U_2) \right)$$

suit la loi  $\mathcal{N}(0, I_2)$ . En particulier, les deux marginales suivent la loi  $\mathcal{N}(0, 1)$ .

*Preuve.* Exercice. ■

Considérons maintenant le cas des vecteurs gaussiens. Plus précisément, considérons une variable aléatoire  $X$  de loi  $\mathcal{N}(0, \Sigma)$  où  $\Sigma$  est une matrice symétrique positive  $d \times d$ . D'après ce qui précède, on sait simuler  $X$  lorsque  $\Sigma = I_d$ . Dans le cas général, on peut s'appuyer sur la propriété suivante (voir section 3.1.6) :

**Proposition 3.1.11.** (i) Soit  $\Gamma$  une matrice symétrique positive. Alors, il existe une unique matrice symétrique positive  $A$  telle que  $A^2 = \Gamma$ . On la note donc  $\sqrt{\Gamma}$ . Alors, si  $Z \sim \mathcal{N}(0, I_d)$ ,  $\sqrt{\Gamma}Z$  suit la loi  $\mathcal{N}(0, \Gamma)$ .

(ii) Plus généralement, si  $T$  est une matrice carrée telle que  $TT^* = \Gamma$ , alors  $TZ$  suit la loi  $\mathcal{N}(0, \Gamma)$ .

**Remarque 3.1.15.** La seconde propriété est reliée à la méthode Cholesky. Plus précisément, lorsque  $\Sigma$  est inversible, alors, on peut décomposer  $\Sigma$  sous la forme  $TT^*$  où  $T$  est une matrice triangulaire inférieure. Cette seconde méthode est en général moins coûteuse que le calcul de la racine carrée (mais n'est utilisable que si l'on sait que  $\Sigma$  est inversible).

*Preuve.* (i) est un cas particulier de (ii). On prouve donc simplement (ii). Il suffit de vérifier que si  $Z$  est une variable aléatoire centrée à valeurs dans  $\mathbb{R}^d$  de matrice de covariance  $C_Z$  (pas nécessairement gaussienne), alors  $TZ$  a pour matrice de covariance  $TC_ZT^*$ . C'est bien le cas car pour tous  $i$  et  $j$  appartenant à  $\{1, \dots, d\}$ ,

$$\text{Cov}((TZ)_i, (TZ)_j) = \mathbb{E}[(TZ)_i(TZ)_j] = \sum_{k,l} T_{i,k} \mathbb{E}[Z_k Z_l] T_{j,l} = \sum_{k,l} T_{i,k} (C_Z)_{k,l} T_{l,j}^* = (TC_ZT^*)_{i,j}.$$

■

### 3.1.6 Annexe : quelques rappels sur les vecteurs gaussiens

Si  $X$  est un vecteur aléatoire en dimension  $d$  (i.e. une v.a. à valeurs dans  $\mathbb{R}^d$  vue comme un vecteur colonne) tel que chacune de ses coordonnées  $X_i$ ,  $i \in \{1, \dots, d\}$ , soit dans  $L^2(\mathcal{A})$ , alors

on définit dans la suite son vecteur espérance et sa matrice de covariance par

$$\mathbb{E}[X] = m := \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_d] \end{pmatrix} \quad \text{et} \quad \text{Cov}(X) = \Gamma := \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,d} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d,1} & \sigma_{d,2} & \cdots & \sigma_{d,d} \end{pmatrix},$$

où  $\sigma_{i,j}$  désigne la covariance entre les variables  $X_i$  et  $X_j$ :

$$\sigma_{i,j} = \text{Cov}(X_i, X_j) := \mathbb{E}[(X_i - m_i)(X_j - m_j)] = \mathbb{E}[X_i X_j] - m_i m_j,$$

où  $m_i := \mathbb{E}[X_i]$  pour tout  $i \in \{1, \dots, d\}$ . En particulier, cette covariance est bien définie par l'inégalité de Cauchy-Schwarz. De plus, on remarque que la matrice réelle  $\Gamma$  est symétrique et qu'elle peut s'écrire comme

$$\Gamma = \mathbb{E}[(X - m)(X - m)^T] = \mathbb{E}[X X^T] - m m^T,$$

où le symbole  $T$  désigne la transposition et l'espérance d'une matrice est définie comme la matrice des espérances de chacun de ses éléments. On en déduit qu'elle est semi-définie positive au sens où pour tout  $x \in \mathbb{R}^d$ ,

$$x^T \Gamma x = \mathbb{E} \left[ \underbrace{x^T (X - m)}_{\in \mathbb{R}} \underbrace{(X - m)^T x}_{\in \mathbb{R}} \right] = \mathbb{E} \left[ (x^T (X - m))^2 \right] \geq 0.$$

Dans la suite de ce chapitre, on supposera que les vecteurs gaussiens considérés sont non dégénérés, c'est-à-dire que  $\det \Gamma \neq 0$  (elle est donc définie positive, i.e.  $x^T \Gamma x > 0$  pour tout  $x \neq 0$  dans  $\mathbb{R}^d$ ).

**Définition 3.1.12.** Soit  $X$  un vecteur aléatoire en dimension  $d$  de vecteur espérance  $m$  et de matrice de covariance  $\Gamma$ . Il est dit gaussien si la densité jointe est donnée par

$$f_X(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Gamma}} \exp \left( -\frac{1}{2} (x - m)^T \Gamma^{-1} (x - m) \right), \quad x \in \mathbb{R}^d.$$

On note alors  $X \sim \mathcal{N}_d(m, \Gamma)$  (et  $\mathcal{N}(m, \Gamma)$  si  $d = 1$ ).

Ainsi, comme dans le cas unidimensionnel, la donnée du vecteur espérance et de la matrice de covariance caractérise la loi d'un vecteur gaussien. La fonction caractéristique étant vue comme la transformée de Fourier de la densité, que l'on sait injective dans  $L^1(\mathcal{A})$ , on a la caractérisation suivante de la loi d'un vecteur gaussien.

**Proposition 3.1.13.** Un vecteur aléatoire  $X$  est gaussien de vecteur espérance  $m$  et de matrice de covariance  $\Gamma$  si et seulement si sa fonction caractéristique est donnée par

$$\phi_X(\theta) := \mathbb{E}[e^{i\theta^T X}] = \exp \left( i\theta^T m - \frac{1}{2} \theta^T \Gamma \theta \right), \quad \theta \in \mathbb{R}^d.$$

À présent, posons-nous la question suivante: un vecteur gaussien a-t-il toutes ses composantes unidimensionnelles gaussiennes ? Et réciproquement, suffit-il d'avoir toutes les coordonnées gaussiennes pour que le vecteur associé soit gaussien ? La proposition suivante permet de répondre à la question.

**Proposition 3.1.14.** *Un vecteur aléatoire  $X$  est gaussien si et seulement si  $\theta^T X$  est une v.a. gaussienne unidimensionnelle pour tout  $\theta \in \mathbb{R}^d$  différent du vecteur nul, i.e. toute combinaison linéaire non nulle de ses coordonnées est gaussienne.*

*Preuve.* Notons respectivement  $m$  et  $\Gamma$  le vecteur espérance et la matrice de covariance du vecteur  $X$ . Supposons le gaussien. Alors pour tout  $\theta \in \mathbb{R}^d$  différent du vecteur nul, la fonction caractéristique de la v.a.r.  $\theta^T X$  s'écrit pour tout  $u \in \mathbb{R}$ :

$$\begin{aligned}\phi_{\theta^T X}(u) &= \mathbb{E} \left[ e^{iu\theta^T X} \right] \\ &= \phi_X(u\theta) \\ &= \exp \left( iu\theta^T m - \frac{u^2}{2} \theta^T \Gamma \theta \right).\end{aligned}$$

Ainsi, on en déduit que pour tout  $\theta \in \mathbb{R}^d$  non nul, la v.a.r.  $\theta^T X$  suit une loi gaussienne d'espérance  $\theta^T m$  et de variance  $\theta^T \Gamma \theta$ , qui est strictement positive car  $\Gamma$  est définie positive. Réciproquement, si  $\theta^T X$  suit une loi gaussienne pour tout  $\theta \in \mathbb{R}^d$  non nul, alors

$$\begin{cases} \mathbb{E} [\theta^T X] &= \sum_{i=1}^d \theta_i \mathbb{E}[X_i] &= \theta^T m; \\ \text{Var} (\theta^T X) &= \text{Var} \left( \sum_{i=1}^d \theta_i X_i \right) &= \sum_{i,j=1}^d \theta_i \theta_j \text{Cov}(X_i, X_j) = \theta^T \Gamma \theta. \end{cases}$$

Alors on a pour tout  $u \in \mathbb{R}$

$$\phi_{\theta^T X}(u) = \exp \left( iu\theta^T m - \frac{u^2}{2} \theta^T \Gamma \theta \right),$$

et en choisissant  $u = 1$  on obtient que pour tout  $\theta \in \mathbb{R}^d$  non nul,

$$\phi_X(\theta) = \exp \left( i\theta^T m - \frac{1}{2} \theta^T \Gamma \theta \right),$$

c'est-à-dire que  $X \sim \mathcal{N}_d(m, \Gamma)$ . ■

Ainsi, chacune des coordonnées d'un vecteur gaussien suit forcément une loi gaussienne. En revanche, la réciproque est fautive: il se peut que 2 variables aléatoires réelles  $Y$  et  $Z$  soient gaussiennes sans que le vecteur  $(Y, Z)^T$  soit un vecteur gaussien. En effet, considérons une v.a.r.  $Y \sim \mathcal{N}(0, 1)$ , indépendante d'une variable  $\varepsilon$  dite de Rademacher, de loi donnée par

$$\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}.$$

On peut montrer que  $Z = \varepsilon Y$  est une v.a.r. gaussienne alors que le vecteur  $(Y, Z)^T$  ne l'est pas. Ainsi, demander à ce qu'un vecteur aléatoire soit gaussien requiert plus que le caractère gaussien des coordonnées.

Le résultat qui suit est une propriété phare du cadre gaussien : il suffit de démontrer la nullité des différentes covariances afin d'obtenir l'indépendance des coordonnées.



**Proposition 3.1.15.** *Soit  $X$  un vecteur gaussien. Alors ses coordonnées sont indépendantes si et seulement si sa matrice de covariance est diagonale.*

*Preuve.* Les coordonnées du vecteur gaussien  $X$  sont indépendantes si et seulement si sa densité  $f_X$  est le produit des densités des coordonnées. Ceci équivaut à dire que la forme quadratique  $(x - m)^T \Gamma^{-1} (x - m)$  apparaissant dans  $f_X$  s'écrit de la forme  $\sum_{i=1}^d \alpha_i (x_i - m_i)^2$ . En d'autres termes, il n'y a aucun terme croisé du type  $(x_i - m_i)(x_j - m_j)$  pour  $i \neq j$ , c'est-à-dire que la matrice  $\Gamma^{-1}$  (donc  $\Gamma$ ) est diagonale. ■

**Proposition 3.1.16** (Transformation linéaire). *Soit  $A$  une matrice inversible  $d \times d$  et  $b$  un vecteur dans  $\mathbb{R}^d$ . Considérons le vecteur gaussien  $X \sim \mathcal{N}_d(m, \Gamma)$ . Alors le vecteur  $AX + b$  est gaussien de vecteur espérance  $Am + b$  et de matrice de covariance  $A\Gamma A^T$ . En particulier, si  $X \sim \mathcal{N}_d(0, \sigma^2 I_d)$  et que la matrice  $A$  est orthogonale, i.e.  $AA^T = I_d$ , alors les vecteurs  $X$  et  $AX$  ont même loi.*

Cette propriété s'étend au cadre non inversible à condition de définir la loi à l'aide de la fonction de caractéristique (pour laquelle on n'a pas besoin d'inverser la matrice de covariance).

## 3.2 Méthodes de Monte-Carlo

### 3.2.1 Principe

Les méthodes de Monte-Carlo désignent dans leur ensemble les méthodes d'approximation numérique basées sur la Loi Forte des Grands Nombres pour approcher l'espérance (et donc pour approcher aussi des quantités définies par des intégrales). On rappelle donc pour commencer la loi forte des grands nombres :

**Théorème 3.2.1.** *Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires i.i.d. de même loi que  $X$  telle que  $\mathbb{E}[|X|] < +\infty$ . Alors,*

$$\bar{X}_N(\omega) := \frac{X_1(\omega) + \dots + X_N(\omega)}{N} \xrightarrow{N \rightarrow +\infty} m_X := \mathbb{E}[X].$$

Ainsi, si l'on sait simuler la suite de variables aléatoires ci-dessus, on sait *a priori* approcher l'espérance de cette même variable aléatoire. Bien entendu, bien que basées sur ce principe général, les méthodes dites de Monte-Carlo ne s'y limitent pas. Elles contiennent à la fois la maîtrise de l'erreur associée et de manière générale, la recherche de moyens appropriés pour minimiser l'erreur (et donc le temps de calcul).

**Remarque 3.2.1.** *En pratique, on cherchera souvent à simuler  $\mathbb{E}[Y]$  où  $Y = f(X)$ . On rappelle au passage que si  $f$  est mesurable et  $(X_n)_{n \geq 1}$  est une suite de variables aléatoires i.i.d.,  $(f(X_n))_{n \geq 1}$  est aussi une suite de variables aléatoires i.i.d. Par conséquent, si  $\mathbb{E}[|f(X)|] < +\infty$ ,*

$$\frac{f(X_1) + \dots + f(X_N)}{N} \xrightarrow{N \rightarrow +\infty} m_f := \mathbb{E}[f(X)].$$

*En pratique, c'est souvent sous cette forme que l'on utilisera la loi des grands nombres.*

**Exemples :** Voici deux exemples où l'on peut avoir la nécessité de faire appel à ce type de méthode :

- On veut calculer un volume ou plus généralement  $\int_{[0,1]^d} f(x_1, \dots, x_d) dx_1 \dots dx_d$ . Les méthodes de Monte-Carlo peuvent constituer une alternative aux méthodes déterministes d'approximation d'intégrales. Elles peuvent en particulier s'avérer efficaces lorsque la dimension est grande.
- On veut calculer le prix d'un produit dérivé en finance :  $\mathbb{E}[\varphi(S_T^1, S_T^2, \dots, S_T^d)]$  où pour tout  $i$ ,  $(S_t^i)_{t \geq 0}$  est la dynamique du prix d'un actif donné et  $\varphi$  est une fonction de  $\mathbb{R}^d$  dans  $\mathbb{R}$ . Comme l'option ne dépend que de la valeur terminale, on parle alors d'option européenne. Dans le cas  $d = 1$  et  $\varphi(x) = (x - K)_+$ , on rappelle que l'on parle de call européen.

### 3.2.2 Contrôle de l'erreur

On rappelle ici des principes standards de l'estimation.

#### Erreur $L^2$ (non asymptotique)

On pose  $\tilde{X}_k = X_k - m_X$ . On a (en dimension 1),

$$\begin{aligned} \mathbb{E} \left[ \left( \frac{1}{N} \sum_{k=1}^N X_k - m_X \right)^2 \right] &= \mathbb{E} \left[ \left( \frac{1}{N} \sum_{k=1}^N \tilde{X}_k \right)^2 \right] \stackrel{\text{car } \tilde{X}_i \perp}{=} \frac{1}{N^2} \sum_{k=1}^N \mathbb{E}[\tilde{X}_k^2] + 0 \\ &= \frac{\text{Var}(X)}{N}. \end{aligned}$$

On obtient donc que

$$\text{dist}_{L^2} \left( \frac{1}{N} \sum_{k=1}^N X_k, m_X \right) = \frac{\sigma_X}{\sqrt{N}}$$

où  $\sigma_{X_1} = \sqrt{\text{Var}(X_1)}$ . On constate donc les points suivants :

- La vitesse de convergence a priori est assez lente de l'ordre de  $1/\sqrt{N}$ . Néanmoins, le coût de calcul de chaque itération est a priori faible et l'erreur n'augmente que linéairement avec la dimension.
- La variance joue un rôle fondamental dans la taille de l'erreur. On développera dans la section suivante des moyens pour réduire sa taille. Pour le moment, remarquons que l'on a a priori pas connaissance de la variance explicite. Ainsi, en pratique, afin d'avoir une idée raisonnable de l'erreur, on pourra calculer la variance empirique de l'estimateur :

$$\begin{aligned} \bar{V}_N &= \frac{1}{N-1} \sum_{k=1}^N (X_k - \bar{X}_N)^2 \\ &= \frac{1}{N-1} \sum_{k=1}^N X_k^2 - \frac{N}{N-1} (\bar{X}_N)^2 \xrightarrow{N \rightarrow +\infty} \text{Var}(X) = \sigma_X^2. \end{aligned}$$

**Remarque 3.2.2.** On rappelle que la renormalisation par  $N-1$  (et non par  $N$ ) permet à cet estimateur de la variance d'être sans biais :  $\mathbb{E}[\bar{V}_N] = \sigma_X^2$ .

### Erreur asymptotique

On s'appuie ici sur le TCL dont on rappelle l'énoncé :

**Théorème 3.2.2.** *Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires réelles i.i.d. de même loi que  $X$  telle que  $\mathbb{E}[X^2] < +\infty$ . Alors,*

$$\sqrt{N} \left( \frac{\bar{X}_N - m_X}{\sigma} \right) \xrightarrow[N \rightarrow +\infty]{} \mathcal{N}(0, 1).$$

Naturellement, on retrouve que l'erreur est de l'ordre de  $\sigma/\sqrt{N}$ . Comme en statistiques, le TCL a ici l'intérêt de fournir des intervalles de confiance asymptotiques : pour un  $\alpha$  donné appartenant à  $[0, 1]$ , on note  $a_\alpha$  l'unique nombre satisfaisant :

$$\mathbb{P}(|\mathcal{N}(0, 1)| \leq a_\alpha) = \alpha, \text{ ou encore tel que } 2F_Z(a_\alpha) - 1 = \alpha.$$

Alors, si l'on note

$$I_N^\alpha = [\bar{X}_N - a_\alpha \frac{\sigma_X}{\sqrt{N}}, \bar{X}_N + a_\alpha \frac{\sigma_X}{\sqrt{N}}],$$

on a

$$\mathbb{P}(m_X \in I_N^\alpha) \xrightarrow[N \rightarrow +\infty]{} \alpha.$$

**Remarque 3.2.3.** *Comme en statistiques, un résultat d'estimation a peu de valeur s'il n'est accompagné d'une estimation de l'erreur ou d'intervalles de confiance. Comme dans la section précédente, on n'a pas accès en général à  $\sigma_X$  de sorte que l'on remplace  $I_N^\alpha$  par  $J_\alpha^N$  défini par*

$$J_\alpha^N = [\bar{X}_N - a_\alpha \sqrt{\frac{\bar{V}_N}{N}}, \bar{X}_N + a_\alpha \sqrt{\frac{\bar{V}_N}{N}}].$$

En utilisant que  $\bar{V}_N \xrightarrow[N \rightarrow +\infty]{} \sigma_X^2$ , on peut à nouveau vérifier que

$$\mathbb{P}(m_X \in J_\alpha^N) \xrightarrow[N \rightarrow +\infty]{} \alpha.$$

### Effort

Dans ce qui précède, on n'a pas tenu compte du coût de simulation. Si l'on veut estimer de manière précise la qualité d'une méthode de Monte-Carlo, il est cependant important de le prendre en compte. On définit alors l'effort de la variable  $X$  de la manière suivante :

**Définition 3.2.3.** *Soit  $X$  une variable aléatoire telle que  $\mathbb{E}[X^2] < +\infty$ . Alors, l'effort de  $X$  noté  $\mathcal{E}(X)$  est défini par*

$$\mathcal{E}(X) = \kappa_X \text{Var}(X)$$

où  $\kappa_X$  désigne le nombre d'opérations nécessaires à la simulation de la variable  $X$ .

Ainsi, après  $M$  opérations, on a effectué  $M/\kappa_X$  itérations (pour simplifier, on suppose que ce nombre est entier). L'erreur quadratique après  $M$  opérations est égale à

$$\frac{\text{Var}(X)}{M/\kappa_X} = \frac{\mathcal{E}(X)}{M}.$$

**Remarque 3.2.4.** *Dans la littérature, on mesure pour simplifier la qualité d'une méthode de Monte-Carlo à la taille de sa variance. Le plus souvent, on fera aussi cet abus ici excepté dans quelques remarques (cf variables antithétiques par exemple).*

### Et si l'on a du biais

Dans la majeure partie du cours, on présentera les résultats en supposant que l'on sait simuler de manière exacte la variable aléatoire dont on cherche à calculer l'espérance. En termes statistiques, cela revient en fait à supposer que l'estimateur est sans biais. En pratique, on ne sait parfois simuler que des approximations  $X^h$  de la variable aléatoire (discrétisation de processus par exemple). Dans ce cas, on doit ajouter également l'erreur due au biais. Par exemple, dans un sens  $L^2$ , on retrouve en notant  $m_{X^h}$  l'espérance de  $X^h$ ,

$$\begin{aligned}\mathbb{E} \left[ \left( \frac{1}{N} \sum_{k=1}^N X_k^h - m_X \right)^2 \right] &= \mathbb{E} \left[ \left( \frac{1}{N} \sum_{k=1}^N \tilde{X}_k^h - m_{X^h} \right)^2 \right] + (m_{X^h} - m_X)^2 \\ &= \frac{\text{Var}(X)}{N} + (m_{X^h} - m_X)^2.\end{aligned}$$

### 3.2.3 Réduction de variance

Jusqu'à présent, les résultats présentés ne diffèrent pas de ceux que l'on en utilise en statistique classique. La différence majeure entre statistiques et méthodes Monte-Carlo réside en fait dans la remarque suivante : "En Statistiques, on ne choisit pas les observations", ou inversement, dans les méthodes de MC, on peut choisir le moyen le mieux adapté pour approcher une quantité déterministe. Prenons un exemple : supposons que l'on veuille calculer  $\mathbb{E}[f(U)]$  où  $U$  est la loi uniforme sur  $[0, 1]$ . Alors, on peut utiliser la loi des grands nombres pour la variable aléatoire  $U$  mais aussi pour  $1 - U$ . En d'autres termes, on n'a pas un seul moyen de calculer cette quantité et l'objectif est maintenant de déterminer un moyen permettant de minimiser la variance (ou l'effort).

On va présenter ci-dessous une famille de méthodes permettant de réduire la variance.

#### Variable de contrôle

Le principe est très simple : on cherche à approcher  $\mathbb{E}[X]$  et on sait qu'il existe une variable aléatoire  $X'$  telle que :

- $\mathbb{E}[X']$  peut être calculée de manière exacte à un coût très faible.
- La simulation de  $X - X'$  a un coût comparable à celle de  $X$  seule.
- $\text{Var}(X - X') < \text{Var}(X)$ .

Alors, pour approcher  $\mathbb{E}[X]$ , on peut appliquer la méthode de Monte-Carlo à

$$\tilde{X} = X - X' + \mathbb{E}[X']$$

puisque par définition,

$$\mathbb{E}[\tilde{X}] = \mathbb{E}[X - X'] + \mathbb{E}[X'] = \mathbb{E}[X].$$

De plus, par hypothèse,

$$\text{Var}(\tilde{X}) = \text{Var}(X - X') < \text{Var}(X).$$

**Définition 3.2.4.** La variable  $X'$  est alors appelée une variable de contrôle pour  $X$ .

**Remarque 3.2.5.** (i) L'utilisation d'une variable de contrôle n'a un sens que si  $X$  et  $X'$  ne sont pas indépendantes. En effet, dans le cas contraire,

$$\text{Var}(X - X') = \text{Var}(X) + \text{Var}(X')$$

et l'on ne peut donc espérer réduire la variance.

(ii) Pour que la méthode soit réellement intéressante, il faut plus précisément que

$$\mathcal{E}(X - X') < \mathcal{E}(X)$$

mais encore une fois, on se limitera en pratique à l'aspect réduction de variance.

(iii) En pratique, une variable  $X'$  est un bon candidat pour être une variable de contrôle si  $X - X'$  est "petit". Par exemple, si  $|X - X'| \leq \varepsilon$  p.s., alors

$$\text{Var}(X - X') \leq \varepsilon^2.$$

Sans être certain de réduire la variance (car cela dépend de celle de  $X$ ), cette technique a alors pour intérêt d'avoir une variance contrôlée (et donc une erreur contrôlée). Un autre cas intéressant est celui où  $0 \leq X' < X$  de sorte que

$$\mathbb{E}[(X - X')^2] < \mathbb{E}[X].$$

Donnons quelques exemples d'application :

**Exemple 3.2.6.** Voici un exemple sans intérêt pratique mais qui permet de donner une idée du principe : approximation de  $I = \int e^x dx$  ( $I = e - 1$ ). Si l'on note  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ , alors  $I = \mathbb{E}[X]$  où  $X = e^U$ . L'idée la plus basique pour approcher cette espérance consiste alors à simuler

$$\bar{X}_N = \frac{1}{N} \sum_{k=1}^N e^{U_k}.$$

La variance associée est égale à :

$$\text{Var}(X) = \mathbb{E}[e^{2U}] - (\mathbb{E}[e^U])^2 = \frac{e^2 - 1}{2} - (e - 1)^2 \approx 0,24.$$

En utilisant le développement au voisinage de 0 de  $e^x$ , on peut envisager de considérer :

$$X' = 1 + U.$$

On constate que  $\mathbb{E}[X'] = 3/2$  de sorte que  $X'$  est a priori utilisable comme variable de contrôle. On peut vérifier que

$$\text{Var}(X - X') \approx 0,04.$$

L'utilisation de la variable de contrôle réduit donc considérablement la variance dans cet exemple.

On verra en TD un exemple d'utilisation de variable de contrôle pour le calcul d'options européennes sur un panier d'actifs.

## Variables antithétiques

L'idée est d'essayer de construire des variables aléatoires de même loi négativement corrélées en se basant sur le lemme suivant :

**Lemme 3.2.7.** *Considérons  $X$  et  $X'$ , deux variables aléatoires telles que  $\mathbb{E}[X] = \mathbb{E}[X']$  et  $\mathbb{E}[X^2] = \mathbb{E}[(X')^2]$ . Alors,*

$$\text{Var}\left(\frac{X + X'}{2}\right) \leq \frac{1}{2}\text{Var}(X) \iff \text{Cov}(X, X') \leq 0.$$

*Preuve.* Le résultat est évident en utilisant que  $\text{Var}(X) = \text{Var}(X')$ . ■

**Remarque 3.2.8.** *Si  $X$  et  $X'$  ont des coûts de simulation égaux (ou quasi-égaux), alors*

$$\mathcal{E}\left(\frac{X + X'}{2}\right) \approx 2\kappa_X \text{Var}\left(\frac{X + X'}{2}\right).$$

*On déduit donc de ce qui précède que l'effort*

$$\mathcal{E}\left(\frac{X + X'}{2}\right) \leq \mathcal{E}(X) \iff \text{Cov}(X, X') \leq 0.$$

*Ceci explique le rôle du 1/2 dans le lemme précédent.*

**Définition 3.2.5.** *Deux variables aléatoires négativement corrélées telles que leur espérance et leur variance sont égales, sont dites antithétiques.*

La question est maintenant la suivante : comment construire de telles variables ? On a la proposition suivante sur  $\mathbb{R}$  :

**Proposition 3.2.6.** *Supposons que  $Z$  soit une variable aléatoire réelle de loi invariante par une transformation  $T$  décroissante. Alors, pour toute fonction  $\varphi$  monotone de  $\mathbb{R}$  dans  $\mathbb{R}$ ,  $X = \varphi(Z)$  et  $X' = \varphi(T(Z))$  sont antithétiques.*

*Preuve.* Comme  $Z \stackrel{\mathcal{L}}{=} T(Z)$ ,  $X$  et  $X'$  ont mêmes espérance et variance. Il suffit maintenant de montrer que  $\text{Cov}(X, X') \leq 0$ . Soient  $Z$  et  $Z'$  deux copies indépendantes. Alors, comme  $\varphi$  est monotone et que  $T$  est décroissante,

$$(\varphi(Z) - \varphi(Z'))(\varphi(T(Z)) - \varphi(T(Z'))) \leq 0 \quad \text{p.s.}$$

Ainsi,

$$\mathbb{E}[(\varphi(Z) - \varphi(Z'))(\varphi(T(Z)) - \varphi(T(Z')))] \leq 0.$$

En développant et en utilisant l'indépendance, il vient

$$2\mathbb{E}[\varphi(Z)\varphi(T(Z))] - 2\mathbb{E}[\varphi(Z)]\mathbb{E}[\varphi(T(Z))] \leq 0$$

puis le résultat. ■

Conséquence :

$$\textcircled{1} \text{ Var}\left(\frac{X+X'}{2}\right) \leq \frac{1}{2} \text{Var}(X)$$

$\textcircled{2}$   $\mathcal{M}$  est plus efficace d'approcher

$$\mathbb{E}[X] = \mathbb{E}[\varphi(Z)] \quad \text{par :}$$

$$\frac{1}{N} \sum_{k=1}^N \varphi(Z_k) + \varphi(T(Z_k))$$

que par  $\frac{1}{N} \sum_{k=1}^N \varphi(Z_k)$  (Moyenne Naïve).  
où  $(Z_k)$  i.i.d de la loi que  $Z$ .

$$\begin{array}{ll} \text{Ex 1: } T: u \mapsto 1-u & \text{Ex 2: } T: z \mapsto -z \\ T(u) = 1-u \sim u & \text{et } T(z) = -z \sim z \end{array}$$



Ci-dessous, deux exemples importants d'applications :

- Lorsque  $Z$  est symétrique,  $T(x) = -x$  convient. Ceci s'applique en particulier à la loi normale centrée réduite. On peut par exemple utiliser cette méthode pour le calcul du call Européen dans le modèle de Black-Scholes. On rappelle que celui-ci est donné par :

$$C_{eur} = e^{-rT} \mathbb{E}[\varphi(Z)]$$

où  $Z \sim \mathcal{N}(0, 1)$  et

$$\varphi(z) = (S_0 e^{(r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}z} - K)_+.$$

L'application  $\varphi$  est croissante donc la méthode fonctionne.

- Lorsque  $U$  suit la loi uniforme sur  $[0, L]$ , l'application  $T(x) = L - x$  convient.

### Combinaison convexe quelconque

**Principe :** On suppose données deux variables  $X$  et  $X'$  de carré intégrable telles que  $\mathbb{E}[X] = \mathbb{E}[X']$  et  $X \neq X'$ . Alors, pour tout  $\theta \in \mathbb{R}$ ,

$$X^\theta = \theta X + (1 - \theta)X'$$

satisfait  $\mathbb{E}[X^\theta] = \mathbb{E}[X]$ . On souhaite déterminer le choix de  $\theta$  qui minimise la variance (*i.e.* tel que  $\mathbb{E}[(X^\theta)^2]$  est minimal puisque  $\mathbb{E}[X^\theta]$  est constant). Cette valeur de  $\theta$  n'est en pratique pas calculable mais peut s'approcher "en ligne".

**Proposition 3.2.7.** *Sous les hypothèses ci-dessus, la fonction  $\theta \mapsto \text{Var}(X^\theta)$  admet un unique minimum en  $\theta^*$  défini par :*

*pb: on ne sait pas le calculer!*

$$\theta^* = \frac{\mathbb{E}[X'(X' - X)]}{\mathbb{E}[(X' - X)^2]} = \frac{\text{Cov}(X', X' - X)}{\text{Var}(X' - X)}.$$

*Approximation en ligne.*

*Preuve.* Posons  $f(\theta) = \text{Var}(X^\theta) = \mathbb{E}[(X^\theta)^2] - \mathbb{E}[X]^2$ . On peut appliquer le théorème de dérivation de Lebesgue car l'application  $(\theta, x, x') \mapsto (\theta x + (1 - \theta)x')^2$  est  $\mathcal{C}^1$  et que pour tout  $R > 0$ ,

$$\sup_{\theta \in [-R, R]} |\partial_\theta ((\theta X + (1 - \theta)X')^2)| \leq C_R(|X|^2 + |X'|^2).$$

Ainsi,  $f$  est dérivable sur  $\mathbb{R}$  et

$$f'(\theta) = 2\mathbb{E}[(X - X')(\theta(X - X') + X')].$$

On en déduit facilement le résultat. ■

**Exemple 3.2.9.** *Un exemple classique est donné par la relation de parité Call/Put. Celui sera vu en TD.*

Exemple: Le calcul de  $I$  où  $I = \int_0^1 e^u du = \mathbb{E}[e^u]$

• Méthode Naïve  $\frac{1}{N} \sum_{k=1}^N e^{u_k}$

• MSE: (Mean-Squared Error)

$$\mathbb{E}\left[\left(\frac{1}{N} \sum_{k=1}^N e^{u_k} - I\right)^2\right] = \frac{\text{Var}(e^u)}{N}$$

$$\begin{aligned} \text{or } \text{Var}(e^u) &= \mathbb{E}[e^{2u}] - \mathbb{E}[e^u]^2 = \int_0^1 e^{2u} du - I^2 \\ &= \left[\frac{1}{2} e^{2u}\right]_0^1 - \left([e^u]_0^1\right)^2 = \frac{1}{2}(e^2 - 1) - (e - 1)^2 \\ &= \frac{1}{2}e^2 - \frac{1}{2} - e^2 + 2e - 1 = -\frac{1}{2}e^2 - \frac{3}{2} + 2e \approx 0,24 \end{aligned}$$

Méthode antithétique:

$$\frac{1}{N} \sum_{k=1}^N \left( \frac{e^{u_k} + e^{1-u_k}}{2} \right) \quad \text{MSE} \leq \frac{\text{Var}\left(\frac{e^u + e^{1-u}}{2}\right)}{N}$$

$$\text{Var}\left(\frac{e^u + e^{1-u}}{2}\right) = \frac{1}{4} [\text{Var}(e^u) + \text{Var}(e^{1-u})] = \frac{1}{4} e^2 - \frac{1}{4} + \frac{1}{4} e - (e-1)^2 = -\frac{3}{4}e^2 + \frac{5}{2}e - \frac{5}{4} \approx 0,004$$

On a divisé la variance par 60.

**Complément : Echantillonnage préférentiel**

**Principe :** On cherche à calculer  $\mathbb{E}[h(X)]$  où  $f$  est une fonction donnée. L'idée de l'échantillonnage préférentiel est de modifier la loi sous laquelle est a priori définie l'espérance pour tirer des points mieux choisis par rapport à la fonction. Par exemple, supposons que  $Y$  suit la loi normale centrée réduite et que le support de  $h$  est contenu dans l'intervalle  $[2, 3]$ . Dans ce cas, la suite  $(h(X_k))_{k=1}^n$  contiendra très peu de valeurs différentes de 0. Ainsi, on peut considérer que l'on "perd" beaucoup de simulations via la méthode de Monte-Carlo standard et la variance est alors énorme relativement à la taille de  $\mathbb{E}[h(X)]$ . Il serait donc plus approprié de modifier la loi pour tirer plus de points dans l'intervalle  $[2, 3]$ . Comment faire ?

Supposons que  $X$  ait une densité  $f$  par rapport à une mesure de référence donnée notée  $\mu$ . Pour simplifier, on supposera dans la suite que  $\mu(dx) = \lambda(dx)$  (mesure de Lebesgue). Cependant, on pourra avoir à l'esprit que ce type de méthode a aussi un sens dans le cadre discret.  $\mu$  peut alors être la mesure de comptage sur  $\mathbb{Z}$ .

Pour modifier l'échantillonnage, on introduit une densité de probabilité  $g$  que l'on suppose strictement positive sur le support de  $f$ . On a :

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f(x)\lambda(dx) = \int_{\mathbb{R}} \frac{h(x)f(x)}{g(x)}g(x)\lambda(dx) = \mathbb{E}\left[\frac{h(Y)f(Y)}{g(Y)}\right]$$

où  $Y$  a pour densité  $g$ . On a donc une autre représentation de  $\mathbb{E}[h(X)]$ . Pour que cette nouvelle représentation génère une méthode de Monte-Carlo plus efficace, il faut que l'effort soit diminué. Pour simplifier, on ne considèrera que la réduction de variance (en faisant l'hypothèse que l'effort pour simuler  $(h(Y_k)f(Y_k)/g(Y_k))_{k \geq 1}$  est comparable à celui de la suite  $(h(X_k))_{k \geq 1}$ ).

Calculons donc la variance associée à la nouvelle variable.

$$\begin{aligned} \text{Var}\left(\frac{h(Y)f(Y)}{g(Y)}\right) &= \mathbb{E}\left[\left(\frac{h(Y)f(Y)}{g(Y)}\right)^2\right] - \mathbb{E}[h(X)]^2 \\ &= \int_{\mathbb{R}} \left(\frac{h(y)f(y)}{g(y)}\right)^2 g(y)\lambda(dy) - \mathbb{E}[h(X)]^2 \\ &= \int_{\mathbb{R}} \left(\frac{h^2(x)f(x)}{g(x)}\right) f(x)\lambda(dx) - \mathbb{E}[h(X)]^2 \end{aligned}$$

On a donc réduction de variance si

$$\mathbb{E}\left(\frac{h^2(X)f(X)}{g(X)}\right) < \mathbb{E}[h^2(X)].$$

**Question :** Comment construire une densité  $g$  qui a la propriété ci-dessus ?

Sans donner une réponse très précise à cette question, on peut commencer par la remarque suivante basée sur l'inégalité de Cauchy-Schwarz.

$$\begin{aligned} \mathbb{E}[h(X)]^2 &= \left(\int h(x)f(x)\lambda(dx)\right)^2 \\ &= \left(\int \frac{h(x)f(x)}{\sqrt{g(x)}}\sqrt{g(x)}\lambda(dx)\right)^2 \end{aligned}$$

$$\begin{aligned}
&\leq \int \frac{h^2(x)f^2(x)}{g(x)} \lambda(dx) \int g(x) \lambda(dx) \\
&\leq \mathbb{E} \left( \frac{h^2(X)f(X)}{g(X)} \right).
\end{aligned}$$

On retrouve bien que la variance est positive. Comme on sait qu'on a égalité dans Cauchy-Schwarz si et seulement si les vecteurs considérés sont colinéaires, on remarque donc que théoriquement,

$$\text{Var} \left( \frac{h(Y)f(Y)}{g(Y)} \right) = 0$$

si et seulement si

$$\frac{h(x)f(x)}{\sqrt{g(x)}} = c\sqrt{g(x)} \quad \lambda(dx) - pp.$$

On peut donc **en théorie** annuler la variance si

$$g(x) = \frac{1}{c} h(x)f(x) \quad \lambda(dx) - pp.$$

mais comme  $g$  doit être une densité de probabilité, cela impose que

$$c = \int h(x)f(x)\lambda(dx) = \mathbb{E}[h(X)].$$

Ainsi, **en pratique**, construire une telle fonction  $g$  n'est possible que si l'on sait calculer  $\mathbb{E}[h(X)]$ , ce qui est justement l'objectif de départ.

En revanche, ce qui précède permet de comprendre comment doit être choisie la densité  $g$ . Il faut donc choisir  $g$  aussi proche que possible de  $fh$  tout en assurant que vérifier que  $\int g(x)\lambda(x) = 1$  est possible et simuler sous  $g$  n'est pas trop coûteux.

**Exemple 3.2.10.** *Approximation de  $\mathbb{P}(X > 2)$  où  $X$  suit une loi de Cauchy, i.e. a pour densité  $f(x) = \frac{1}{\pi(1+x^2)}$ .*

*Il s'agit du calcul de  $\mathbb{E}[h(X)]$  avec  $h(x) = 1_{x>2}$ . Via ce qui précède, il est naturel de poser  $g(x) = 0$  si  $x < 2$ . Ensuite, en faisant l'hypothèse que  $\int h(x)f(x)dx$  n'est pas calculable, on choisit une approximation  $h(x)f(x)$  dont on sait calculer explicitement l'intégrale. Lorsque  $x > 2$ ,  $h(x)f(x)$  est proche de  $\frac{2}{\pi x^2}$ . On pose donc*

$$g(x) = \frac{1}{\int_2^{+\infty} \frac{2}{\pi x^2} dx} \frac{2}{\pi x^2} 1_{x>2} = \frac{2}{x^2} 1_{x>2}.$$

### 3.3 Ouverture : Méthodes MCMC

Les méthodes MCMC (littéralement, “Monte-Carlo Markov Chains”) sont des méthodes de simulation ou d'approximation Monte-Carlo<sup>1</sup> basées sur la convergence des chaînes de Markov à l'équilibre. Plus exactement, à minima dans un cadre discret, le principe de cette méthode est, pour une probabilité  $\pi$  fixée, de construire une chaîne de Markov (ou plutôt une famille de

<sup>1</sup>Ces deux problèmes sont finalement évidemment reliés puisque l'on cherche généralement à simuler une variable aléatoire sous une loi donnée  $\pi$  pour être capable de calculer  $\int f(x)\pi(dx)$ .

chaînes de Markov) dont la probabilité invariante est justement la probabilité  $\pi$  puis de simuler la chaîne de Markov jusqu'à un temps suffisamment important pour que cette dernière soit "proche de son équilibre".

Ce thème, devenu très populaire, dans le cadre de la *grande dimension* notamment (ex: calcul sur des grands graphes), est hors programme mais peut faire l'objet d'un projet TER pour les étudiant(e)s intéressé(e)s par les méthodes de simulation.