

# Cours de probabilités Math-Eco-S6

Yoann Dabrowski

4 mai 2017

# Chapitre 1

## Rappels sur les fonctions caractéristiques, la convergence en loi et le Théorème central limite

On suppose connu le modèle des probabilités en terme de théorie de la mesure.

L'espace des réalisations  $\Omega$  d'une expérience aléatoire est l'ensemble des résultats possibles de cette expérience.

**Définition 1.** La famille  $\mathcal{T}$  des événements constitue une partie de  $P(\Omega)$  ayant la structure d'une tribu qui, par définition, possède les propriétés suivantes :

1.  $\emptyset \in \mathcal{T}, \Omega \in \mathcal{T}$ .
2. Si  $A \in \mathcal{T}$  alors  $A^c \in \mathcal{T}$ .
3. Pour toute suite finie ou infinie dénombrable  $(A_n)_{n \geq 1}$  de parties de  $\mathcal{T}$ , alors  $\cup_{n \geq 1} A_n \in \mathcal{T}$ .

Pour  $\Omega = \mathbb{R}^n$ , on utilisera en général la tribu des boréliens  $\mathcal{B}(\mathbb{R}^n)$ , plus petite tribu engendrée par les ensembles ouverts.

**Définition 2** (Définition d'une probabilité). On appelle *probabilité* sur une tribu  $\mathcal{T}$  toute application  $P : \mathcal{T} \rightarrow [0, 1]$  ayant les propriétés suivantes :

1.  $P(\emptyset) = 0, P(\Omega) = 1$ .
2. Pour toute suite finie ou infinie (dénombrable)  $(A_n)_{n \geq 1}$  d'événements deux à deux incompatibles,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Autrement dit,  $P$  est une mesure positive avec  $P(\Omega) = 1$ .

## 1 Lois, Invariant des lois, Fonction Caractéristique (Transformée de Fourier)

### 1.1 Rappels sur les Lois

On utilisera souvent le lemme suivant vu en théorie de la mesure.

**Lemme 1** (de classe monotone). *Si deux probabilités  $P_1, P_2$  sur  $(\Omega, \mathcal{T})$  sont égales sur une classe  $\mathcal{C}$  stable par intersection finie (c'est-à-dire  $\forall C \in \mathcal{C}, P_1(C) = P_2(C)$ ) et telle que  $\sigma(\mathcal{C}) = \mathcal{T}$  alors  $P_1 = P_2$ .*

**Définition 3.** Soit  $(E, \mathcal{E})$  un espace mesurable. Une variable aléatoire (v.a.) définie sur  $\Omega$  est une application mesurable  $X : \Omega \rightarrow E$ , c'est-à-dire telle que, pour tout  $B \in \mathcal{E}$

$$\{X \in B\} \equiv X^{-1}(B) \equiv \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{T}.$$

Le plus souvent l'espace d'arrivée sera  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , cas des variables aléatoires réelles (v.a.r.) ou  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ , cas des variables aléatoires vectorielles (v.a.v.) ou vecteurs aléatoires.

Comme cette tribu est engendrée par les produits d'intervalles, une application  $X : \Omega \rightarrow \mathbb{R}^n$  est un v.a.v si et seulement si  $X^{-1}(\prod_{i=1}^n [a_i, b_i]) \in \mathcal{T}$ , pour tout  $a_i < b_i$ .

**Définition 4.** Soit  $X : \Omega \rightarrow (E, \mathcal{E})$  une v.a. On appelle **loi de  $X$**  (ou distribution, ou mesure image de  $X$ ) la probabilité sur  $(E, \mathcal{E})$ , notée  $P_X$ , définie par

$$P_X(B) = P(\{X \in B\}) = P(X^{-1}(B)), B \in \mathcal{E}$$

On vérifie que  $P_X$  ainsi définie est effectivement une probabilité (car l'image inverse respecte unions et complémentaires). Cela ramène l'étude des probabilités associées aux vecteurs aléatoires à l'étude de probabilités sur  $\mathbb{R}^n$  puisqu'on veut des résultats ne dépendant pas de  $\Omega$  mais seulement de la loi  $P_X$ .

**Définition 5** (Egalité en loi). Deux v.a.  $X$  et  $Y$  ayant même loi  $P_X = P_Y$  sont dites équidistribuées ou égales en loi ou *identiquement distribuées*, ce que l'on peut noter  $X =_{\mathcal{L}} Y$ . Insistons sur le fait que l'égalité en loi est distincte de l'égalité ponctuelle.

Pour les vecteurs aléatoires, on a la conséquence importante du lemme de classe monotone (cf. Th de la mesure, utilisé dans l'unicité de la mesure de Lebesgue)

**Proposition 2.** *Soient  $X, Y : \Omega \rightarrow \mathbb{R}^n$ . Les propriétés suivantes sont équivalentes*

1.  $X, Y$  sont égales en loi :  $P_X = P_Y$ .
2. Pour tout  $h : \mathbb{R}^n \rightarrow \mathbb{R}$ , continue bornée,  $\int h(X)dP = \int h(Y)dP$
3. Pour tout ouvert  $O$  de  $\mathbb{R}^n$ ,  $P_X(O) = P_Y(O)$ .
4. pour tout  $(x_1, \dots, x_n) \in \mathbb{R}^n$  :

$$P_X(]-\infty, x_1] \times \dots \times ]-\infty, x_n]) = P_Y(]-\infty, x_1] \times \dots \times ]-\infty, x_n]).$$

La fonction  $F_X(x_1, \dots, x_n) = P_X(]-\infty, x_1] \times \dots \times ]-\infty, x_n])$  appelée **fonction de répartition** caractérise donc la loi.

*Démonstration.* Les produits d'intervalles  $]-\infty, x_1] \times \dots \times ]-\infty, x_n]$  et les ouverts sont des familles stables par intersection finie et engendrent la tribu des boréliens de  $\mathbb{R}^n$  (car par intersection et complémentaire on obtient les boules carrées de la norme infini et que tout ouvert de  $\mathbb{R}^n$  est union dénombrable de telles boules, de centre un point de  $\mathbb{Q}^n$  par densité de  $\mathbb{Q}^n$ .) On applique donc le lemme de classe monotone 1 pour obtenir les 2 dernières équivalences. 1 implique 2 vient du th de transfert plus bas comme l'équivalence de 2 avec : Pour tout  $h : \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $\int_{\mathbb{R}^d} h(x)dP_X(x) = \int_{\mathbb{R}^d} h(x)dP_Y(x)$ .

Pour montrer 3 à partir de 2 et conclure, il suffit de remarquer que  $h_n(x) = \max(1, nd(., O^c))$  sont des fonctions continues bornées par 1 (car la distance à un fermé  $x \mapsto d(x, O^c) = \inf\{d(x, y), y \in O^c\}$

est continue, cf. MASS 31). Si  $x \in O^c$ ,  $h_n(x) = 0$  et sinon,  $h_n$  est une suite croissante qui tend vers  $h_n(x) \rightarrow 1_{O^c}(x)$  (car si  $x \in O$ ,  $nd(\cdot, O^c) \rightarrow \infty$  donc  $\geq 1$  pour  $n$  assez grand donc  $h_n(x) = 1$  pour  $n$  assez grand). Donc par convergence monotone,  $\int_{\mathbb{R}^d} h_n(x) dP_X(x) \rightarrow P_X(O)$  d'où l'égalité du 3. par celle du 2. □

## 1.2 Théorème de transfert

Le premier théorème fondamental montrant que seule la loi  $P_X$  d'une v.a. est importante et pas la réalisation concrète sur  $\Omega$  est la suivante :

**Théorème 3** (Théorème de transfert). *Soit  $X : (\Omega, \mathcal{T}, P) \rightarrow (E, \mathcal{E})$  une v.a. et  $h : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  une fonction mesurable. Alors, si  $h$  est à valeur positive :*

$$\mathbf{E}(h(X)) = \int_E h(x) dP_X(x).$$

*De plus, si  $h$  n'est pas à valeur positive  $h(X) \in L^1(\Omega, P)$  si et seulement si  $h \in L^1(E, P_X)$  et on a encore  $\mathbf{E}(h(X)) = \int h(x) dP_X(x)$ .*

Autrement dit on ramène une intégrale sur  $\Omega$  à une intégrale sur  $\mathbb{R}$  :

$$\int_{\Omega} h(X(\omega)) dP(\omega) = \int_{\mathbb{R}} h(x) dP_X(x).$$

PREUVE : On procède comme pour la construction de l'intégrale. Si  $h = 1_B$  avec  $B \in \mathcal{E}$ ,  $h \circ X = 1_{X^{-1}(B)}$  et donc  $\mathbf{E}(h(X)) = P(X^{-1}(B)) = P_X(B) = \int h(x) dP_X(x)$ . Par linéarité on obtient le cas de  $h$  étagé. Si  $h$  positive,  $h$  est la limite croissante d'une suite de fonctions étagées, par exemple :

$$h_n(x) = \sum_{k=0}^{4^n} \frac{k}{2^n} 1_{[\frac{k}{2^n}, \frac{k+1}{2^n}[}(h(x)) = \sum_{k=0}^{4^n} \frac{k}{2^n} 1_{h^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}[)}(x)$$

qui est étagée mesurable (car  $h$  est mesurable). En effet, voyons que  $h_n(x) \leq h_{n+1}(x)$ , si  $h(x) \in [\frac{k}{2^n}, \frac{k+1}{2^n}[$ , soit  $h(x) \in [\frac{2k}{2 \cdot 2^n}, \frac{2k+1}{2 \cdot 2^n}[$  et alors  $h_n(x) = h_{n+1}(x)$ , soit  $h(x) \in [\frac{2k+1}{2 \cdot 2^n}, \frac{2k+2}{2 \cdot 2^n}[$  et alors  $h_n(x) = k/2^n \leq h_{n+1}(x) = (2k+1)/2^{n+1}$ . Dans les cas en dehors de ces intervalles,  $h_n(x) = 0$  qui est forcément inférieur aux valeurs positives prises par  $h_{n+1}$ . Comme  $h_n(x) \rightarrow h(x)$  par construction, on applique le théorème de convergence monotone aux deux mesures :

$$\mathbf{E}(h(X)) = \lim_{n \rightarrow \infty} \mathbf{E}(h_n(X)) = \lim_{n \rightarrow \infty} \int h_n(x) dP_X(x) = \int h(x) dP_X(x).$$

Le dernier résultat du cas intégrable est évident par le cas positif pour l'équivalence et par linéarité pour l'égalité. ■

**Exemple 1. Loi de  $Y = \frac{1}{1+U}$  si  $U \sim \mathcal{U}([0, 1])$ .**

Remarquons que lorsque  $X$  et  $Y$  sont égales en loi, on voit que, pour toute fonction  $h$ ,  $E[h(X)] = E[h(Y)]$ , pourvu que l'un des deux membres existe. En particulier,  $X$  et  $Y$  ont mêmes paramètres (moyenne, variance, fonction de répartition, fonctions caractéristiques etc).

**Exemple 2. Cas discret, cas continu.**

**Proposition 4** (Lemme de Doob-Dynkin). *Soit  $X$  une variable aléatoire,  $X : (\Omega, \mathcal{T}, P) \rightarrow (E, \mathcal{E})$ ,  $\sigma(X) = \{A = X^{-1}(B), B \in \mathcal{E}\}$  est une tribu (appelé **tribu engendrée par  $X$** ) et si  $Y : \Omega \rightarrow \mathbb{R}^n$  v.a. est  $\sigma(X)$ -mesurable si et seulement si il existe  $h : (E, \mathcal{E}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  tel que  $Y = h(X)$ .*

De même, si  $X = (X_i)_{i \in I}$  une famille de variables aléatoires,  $X_i : (\Omega, \mathcal{T}, P) \rightarrow (E_i, \mathcal{E}_i)$ , on note  $\sigma(X) = \sigma(X_i^{-1}(B_i), B_i \in \mathcal{E}_i)$  somme tribu engendrée par les tribus  $\sigma(X_i)$ .

PREUVE : La condition suffisante est évidente. Réciproquement, on raisonne comme pour le transfert par le cas étagé  $Y = \sum_i \lambda_i 1_{A_i}$  et  $A_i = X^{-1}(B_i)$  et alors  $h = \sum \lambda_i 1_{B_i}$  convient. Sinon, si  $Y$  positive, on la prend pour limite simple de  $Y_n$  étagée de la forme  $h_n(X)$  par le cas étagé, et  $h(x) = \liminf_{n \rightarrow \infty} h_n(x)$  convient car mesurable et car  $Y(\omega) = \lim_n h_n(X(\omega)) = h(X(\omega))$  en  $X(\omega)$   $h_n$  converge d'après le choix de  $Y_n$ . ■

### 1.3 Théorème d'injectivité de la transformée de Fourier

**Définition 6.** La *fonction caractéristique* (f.c. ou transformée de Fourier) du v.a.  $(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$  est définie par

$$\Phi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \mathbf{E}[e^{i\langle t, X \rangle}],$$

pour tout  $t = (t_1, \dots, t_n) \in \mathbb{R}^n$  et en notant le produit scalaire  $\langle t, X \rangle := \sum_{i=1}^n t_i X_i$ .

La fonction  $\varphi_X$  caractérise la loi de  $X$  par le théorème d'injectivité de la transformée de Fourier/ théorème d'inversion de la transformée de Fourier ci-dessous. On utilisera aussi plus tard au chapitre 2 la fonction caractéristique pour caractériser une notion de convergence, au chapitre 3 pour l'introduction des vecteurs gaussiens qui seront la base du chapitre 5 sur le mouvement brownien. C'est une notion FONDAMENTALE...

**Lemme 5.** *Soit  $X \sim \mathcal{N}(m, \sigma^2)$  de loi normale alors  $\Phi_X(t) = \exp(-\frac{t^2 \sigma^2}{2} + imt)$ .*

PREUVE : On a vu une preuve à l'exercice 8 du TD 3 de MASS 31 utilisant que la partie imaginaire est nulle par parité et le calcul de la partie réelle en établissant une équation différentielle par intégration dépendant d'un paramètre.

On donne ici une autre preuve par prolongement analytique. Par transfert, on doit montrer  $\int \frac{1}{\sigma\sqrt{2\pi}} e^{ixt - \frac{(x-m)^2}{2\sigma^2}} = \exp(-\frac{t^2 \sigma^2}{2} + imt)$  en faisant le changement de variables  $u = (x - m)/\sigma$  on se ramène au cas  $\sigma = 1, m = 0$ .

En prenant  $m = z$  dans le calcul de la densité, on a pour  $z \in \mathbb{R}$

$$\int_{-\infty}^{\infty} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2 + z^2 - 2xz}{2}} = \int_{-\infty}^{\infty} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-z)^2}{2}} = 1.$$

Pour  $z \in \mathbb{C}$ , en appliquant le résultat précédent

$$\sum_{n=0}^{\infty} \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} \frac{|zx|^n}{n!} e^{-\frac{x^2}{2}} = \lim_{N \rightarrow \infty} \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} \sum_{n=0}^N \frac{|zx|^n}{n!} e^{-\frac{x^2}{2}} \leq \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + |zx|} \leq \exp\left(\frac{|z|^2}{2}\right) < \infty$$

La première bornitude permet d'appliquer le TCD pour les séries (ou Fubini pour la mesure discrète) et intervertir somme et série :

$$\sum_{n=0}^{\infty} z^n \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} \frac{x^n}{n!} e^{-\frac{x^2}{2}} = \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + zx}$$

la fonction de droite est donc la somme d'une série entière  $\exp(\frac{z^2}{2})$  pour  $z \in \mathbb{R}$ , donc par identification des coefficients, elle vaut cette valeur pour tout  $z \in \mathbb{C}$ , en particulier pour  $z = it$  et on trouve le résultat. ■

On démontrera le théorème suivant dans la prochaine section puisque la preuve utilise des propriétés générales de l'indépendance importante à noter pour elles-mêmes :

**Théorème 6** (Théorème d'injectivité de la transformation de Fourier). *Deux v.a.  $(X_1, \dots, X_n), (Y_1, \dots, Y_n)$  tels que*

$$\Phi_{(X_1, \dots, X_n)}(t) = \Phi_{(Y_1, \dots, Y_n)}(t) \forall t \in \mathbb{R}^n$$

*sont égales en loi, c'est à dire :*

$$P_{(X_1, \dots, X_n)} = P_{(Y_1, \dots, Y_n)}.$$

*De plus, si  $\Phi_X \in L^1(\mathbb{R}^n, \text{Leb})$  alors  $P_{(X_1, \dots, X_n)}$  a une densité par rapport à la mesure de Lebesgue donnée par (la transformée de Fourier inverse) qui est une fonction continue :*

$$f_{(X_1, \dots, X_n)}(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \Phi_{(X_1, \dots, X_n)}(t) \exp(-i\langle x, t \rangle) dt.$$

**Proposition 7** (Caractérisations de l'indépendance). *Soient  $(X_1, \dots, X_n)$  des v.a. réelles. On a équivalence des assertions :*

1.  $X_1, \dots, X_n$  sont indépendantes.
2.  $\forall a_1 \in \mathbb{R}, \dots, \forall a_n \in \mathbb{R}$

$$P(X_1 \leq a_1, \dots, X_n \leq a_n) = \prod_{i=1}^n P(X_i \leq a_i)$$

3.  $\forall (t_1, \dots, t_n) \in \mathbb{R}^n$

$$\Phi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \prod_{i=1}^n \Phi_{X_i}(t_i).$$

PREUVE : (i) implique (ii) ou (iii) est évident par le théorème de la loi produit et sa conséquence. (ii) implique (i) vient de la caractérisation de la loi par les fonctions de répartition (Prop 2). (iii) implique (i) vient de même en utilisant l'injectivité de la transformée de Fourier. ■

## 1.4 Sommes de variables aléatoires indépendantes (Rappels)

Vous avez probablement vu en TD de théorie de la mesure la définition de la convolution que l'on rappelle ici et relie aux sommes de variables aléatoires indépendantes.

**Définition 7** (Convolution). Soit  $\mu$  une mesure de Proba sur  $S \subset \mathbb{R}^d$  et  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction mesurable telle que pour tout  $x \in S, y \mapsto f(x - y)$  est dans  $L^1(\mathbb{R}^d, \mu)$ , la convolution de  $f$  et  $\mu$  est la fonction  $f * \mu$  définie par :

$$(f * \mu)(x) = \int_{\mathbb{R}^d} f(x - y) d\mu(y).$$

Si  $\mu$  est absolument continue par rapport à la mesure de Lebesgue de densité  $g$ , on note aussi  $f * g$ .

**Proposition 8.** *Soient  $X, Y : \Omega \rightarrow \mathbb{R}^d$  des v.a. indépendantes :*

1.  $\forall t \in \mathbb{R}^d, \Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t)$
2. Si  $X_i, Y_i$  sont dans  $L^2(\Omega)$ ,  $Cov(X_i + Y_i, X_j + Y_j) = Cov(X_i, X_j) + Cov(Y_i, Y_j)$ .
3. Si  $P_X(dx) = f(x)dx, P_Y(dx) = g(y)dy$  alors  $P_{X+Y}$  est absolument continue par rapport à Lebesgue (sur  $\mathbb{R}^d$ ) de densité  $f * g$  définie Lebesgue p.p. :

$$P_{X+Y}(dz) = (f * g)(z)dz.$$

4. Si seulement  $X$  est de loi absolument continue mais de densité continue bornée  $f$ , alors quel que soit  $Y$ ,  $P_{X+Y}$  est absolument continue par rapport à Lebesgue (sur  $\mathbb{R}^d$ ) de densité  $f * P_Y$  (définie partout). De plus, pour tout  $h$  continue bornée :

$$E((h * f)(Y)) = E(h(X + Y)).$$

PREUVE : 1. On a  $\Phi_{X+Y}(t) = \mathbf{E}[e^{it(X+Y)}] = \mathbf{E}[e^{itX}e^{itY}] = \mathbf{E}[e^{itX}]\mathbf{E}[e^{itY}] = \Phi_X(t)\Phi_Y(t)$  l'avant dernière égalité par indépendance car  $f(x) = e^{itx}$  est bornée donc intégrable (par rapport à une probabilité).

2. En général par bilinéarité  $Cov(X_i + Y_i, X_j + Y_j) = Cov(X_i, X_j) + Cov(Y_i, Y_j) + Cov(Y_i, X_j) + Cov(Y_i, X_j)$ , mais ici par indépendance les deux derniers termes sont nuls.

3. Il faut d'abord vérifier que  $f * g$  est bien définie. Par Fubini-Tonelli vu le caractère positif :

$$\int_{\mathbb{R}^n} dx \int_{\mathbb{R}^n} dy f(x-y)g(y) = \int_{\mathbb{R}^n} dy \left( \int_{\mathbb{R}^n} dx f(x-y) \right) g(y) = \int_{\mathbb{R}^n} dy g(y) = 1$$

donc  $\int_{\mathbb{R}^n} dy f(x-y)g(y)$  existe et est fini p.p.

En prenant  $h$  mesurable positive et en appliquant le transfert, on obtient par changement de variables  $z = x + y$  dans l'intégrale sur  $y$  obtenue par Fubini :

$$E(h(X + Y)) = \int_{\mathbb{R}^{2d}} h(x+y)f(x)dxP_Y(dy) = \int_{\mathbb{R}^{2d}} h(z)f(z-y)dzP_Y(dy) = \int_{\mathbb{R}^d} h(z)(f * P_Y)(z)dz$$

ce qui donne le calcul de densité (égalité de la loi avec seulement le cas  $h = 1_B$ ). Dans le cas de 4. on raisonne pareil sauf que  $f$  continue bornée donne  $x \mapsto f(x-y)$  intégrable par rapport à la proba  $P_Y$  directement. L'application de Fubini vient de  $\int_{\mathbb{R}^{2d}} |h(z)f(z-y)|dzP_Y(dy) \leq \|h\|_\infty$ . L'égalité intermédiaire donne aussi  $E(h(X + Y)) = \int_{\mathbb{R}^d} (h * f)(y)P_Y(dy) = E((h * f)(Y))$  par transfert. ■

## 1.5 Preuve [Facultative] du Thm d'injectivité de la transformée de Fourier

On va utiliser les lois gaussiennes pour se ramener au cas avec densité tout en exploitant leurs propriétés de stabilité par cette transformée.

**Lemme 9.** Soit  $g_\sigma$  la densité sur  $\mathbb{R}^n$  d'un  $n$ -uplet de variable gaussienne i.i.d.  $\mathcal{N}(0, \sigma^2)$ . Pour tout  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  continue bornée,  $(h * g_\sigma)(x) \rightarrow_{\sigma \rightarrow 0} h(x)$ . On a même convergence uniforme sur tout compact.

En terme de convergence en loi, cela signifiera au chapitre 2 que si  $(X_1(\sigma), \dots, X_n(\sigma))$  sont les variables de densités  $g_\sigma$ , alors  $x + (X_1(\sigma), \dots, X_n(\sigma)) \rightarrow_{\sigma \rightarrow 0} x$  en loi en utilisant la proposition 8.(4) au cas  $Y = x$ .

PREUVE : Par transfert et changement de variables

$$(h * g_\sigma)(x) - h(x) = \int_{\mathbb{R}^d} (h(x - \sigma z) - h(x))g_1(z)dz.$$

En prenant, en prenant le supremum sur un compact  $K$  :

$$\sup_{x \in K} |(h * g_\sigma)(x) - h(x)| \leq \int_{\mathbb{R}^d} \sup_{x \in K} |(h(x - \sigma z) - h(x))| g_1(z) dz$$

la limite vient de la convergence dominée par une constante  $2\|h\|_\infty$  puisque une constante est intégrable par rapport à une probabilité comme  $g_1(z)dz$ , et la limite ponctuelle en  $z$  vient de la continuité de  $h$  qui est donc uniformément continue sur  $K + B(0, |z|)$  et donc pour  $|\sigma| < 1$ ,  $x - \sigma z, x$  sont dans ce compact de distance  $\sigma|z|$  tendant vers 0. Si  $h$  est uniformément continue sur  $\mathbb{R}^d$  on a même convergence uniforme sur  $\mathbb{R}^d$ . ■

PREUVE DU THM 6 : Pour montrer l'injectivité, par le lemme 2, il suffit de montrer que l'égalité des transformée de Fourier implique égalité de  $\mathbf{E}(h(X))$  pour tout  $h$  continue bornée.

Or par le lemme précédent,  $(h * g_\sigma)(x) \rightarrow h(x)$  tout en étant borné par  $\|h\|_\infty$  donc par TCD :

$$\mathbf{E}(h(X)) = \lim_{\sigma \rightarrow 0} \mathbf{E}((h * g_\sigma)(X)) = \lim_{\sigma \rightarrow 0} \mathbf{E}(h(X + Y_\sigma))$$

la dernière égalité avec  $Y_\sigma$  de densité  $g_\sigma$  et indépendant de  $X$  par la proposition 8 (4) puisque la densité  $g_\sigma$  est continue bornée. Or la transformée de Fourier de  $X + Y_\sigma$  est  $\Phi_{X+Y_\sigma}(t) = \Phi_X(t)\Phi_{Y_\sigma}(t)$  par la proposition 8 (2) et donc

$$\Phi_{X+Y_\sigma}(t) = \Phi_X(t) \exp\left(-\frac{\|t\|_2^2 \sigma^2}{2}\right)$$

par le calcul du lemme 5. Comme ceci est intégrable, on s'attend à avoir la formule d'inversion de Fourier de la deuxième partie qui va donner  $\mathbf{E}(h(X + Y_\sigma))$  en fonction de  $\Phi_{X+Y_\sigma}(t)$ , nous allons donc la montrer à la main dans ce cas pour conclure la preuve.

Or en interprétant la densité comme une variante de la transformée de Fourier dans le cas gaussien :

$$(g_\sigma * P_X)(x) = \int_{\mathbb{R}^d} \frac{1}{\sigma^d (2\pi)^{d/2}} \exp\left(-\frac{\|x - y\|_2^2}{2\sigma^2}\right) P_X(dy) = \int_{\mathbb{R}^{2d}} P_X(dy) dv \frac{1}{\sigma^d (2\pi)^d} \exp\left(-\frac{\|v\|^2}{2} + i\langle \frac{y - x}{\sigma}, v \rangle\right)$$

soit par le changement de variables  $u = v/\sigma$  de jacobien  $\sigma^{-d}$  on obtient

$$\mathbf{E}(h(X + Y_\sigma)) = \int_{\mathbb{R}^d} dx h(x) (g_\sigma * P_X)(x) = \int_{\mathbb{R}^{3d}} dx P_X(dy) dv h(x) \frac{1}{(2\pi)^d} \exp\left(-\frac{\sigma^2 \|v\|^2}{2} + i\langle y - x, v \rangle\right)$$

soit en appliquant Fubini sur les intégrales en  $y, v$

$$\mathbf{E}(h(X + Y_\sigma)) = \int_{\mathbb{R}^{2d}} dx dv \frac{h(x)}{(2\pi)^d} \exp\left(-\frac{\sigma^2 \|v\|^2}{2} - i\langle x, v \rangle\right) \Phi_X(v) = \int_{\mathbb{R}^{2d}} dx dv \frac{h(x)}{(2\pi)^d} \exp(-i\langle x, v \rangle) \Phi_{X+Y_\sigma}(v)$$

qui est la formule souhaitée qui ne dépend bien que de la transformée de Fourier  $\Phi_X$  et conclut l'injectivité.

Maintenant si  $\Phi_X$  est intégrable  $|h(x)\Phi_{X+Y_\sigma}(v)| \leq h(x)|\Phi_X(v)|$  est une domination (si  $h$  est à support compacte) et puisque  $\Phi_{X+Y_\sigma}(v) \rightarrow_{\sigma \rightarrow 0} \Phi_X(v)$  par les formules précédentes, on obtient par le TCD la formule souhaitée pour la densité à la limite. La continuité de la densité vient du Théorème de continuité des intégrales à paramètres. On remarque qu'en utilisant  $\mathbf{E}(h(X)) = \int_{\mathbb{R}^d} dx h(x) f_X(x)$  pour tout  $h$  positive continue à support compact, on déduit  $f_X$  positive (sinon par continuité elle est négative sur un ouvert dans lequel on peut prendre le support de  $h$  pour contredire positivité de l'intégrale) et par convergence monotone et faisant tendre  $h \rightarrow 1$ , on déduit  $f_X$  intégrable et densité de proba. D'où on peut utiliser  $\mathbf{E}(h(X)) = \int_{\mathbb{R}^d} dx h(x) f_X(x)$  (maintenant valable pour  $h$  continue bornée car  $f_X$  peut servir de domination) pour identifier  $P_X(dx) = f_X(x)dx$  en utilisant le lemme 2. ■



## 2 Convergence en loi, Théorème de Paul Lévy

On note  $C_b(\mathbb{R}^d)$  l'espace de Banach (e.v.n complet) des fonctions continues bornées sur  $\mathbb{R}^d$  muni de la norme de la convergence uniforme :  $\|\phi\| = \sup_{x \in \mathbb{R}^d} |\phi(x)|$ .

**Définition 8.** Soit  $Y_n : \Omega_n \rightarrow \mathbb{R}^d$  une suite de variables aléatoires et  $\mu$  une probabilité sur  $\mathbb{R}^d$ . On dit que  $Y_n$  **converge en loi vers**  $\mu$ , notée  $Y_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \mu$ , si

$$\forall \varphi \in C_b(\mathbb{R}^d), \quad \mathbf{E}[\varphi(Y_n)] \rightarrow_{n \rightarrow \infty} \int_{\mathbb{R}^d} \varphi(x) d\mu(x).$$

Si de plus  $Y : \Omega \rightarrow \mathbb{R}^d$  une autre v.a.,  $Y_n$  converge en loi vers  $Y$ , notée  $Y_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} Y$ , si  $Y_n$  converge en loi vers  $P_Y$ .

La convergence en loi ne dépend que des lois de  $Y_n$  et c'est pour cela que les espaces probabilisés  $\Omega_n$  n'ont pas besoin d'être égaux. On relie d'abord la convergence en loi avec la convergence la plus faible déjà rencontrée, la convergence en probabilité : On cherche ensuite des définitions équivalentes. On note  $C_c(\mathbb{R}^d)$  les fonctions continues à support compact (c'est à dire nulles en dehors d'un compact).

**Proposition 10.** 1. Soit  $X_n : \Omega_n \rightarrow \mathbb{R}^d$  une suite de variables aléatoires et  $\mu$  une probabilité sur  $\mathbb{R}^d$ . Alors  $X_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \mu$ , si et seulement si

$$\forall \varphi \in C_c(\mathbb{R}^d), \quad \mathbf{E}[\varphi(X_n)] \rightarrow_{n \rightarrow \infty} \int_{\mathbb{R}^d} \varphi(x) d\mu(x).$$

2. (cas des v.a. entières) Si  $Y_n : \Omega_n \rightarrow \mathbb{Z}^d = S$  et  $\mu$  une probabilité sur  $\mathbb{Z}^d$  alors  $Y_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \mu$ , si et seulement si

$$\forall x \in \mathbb{Z}^d, \quad P(Y_n = x) \rightarrow_{n \rightarrow \infty} \mu(\{x\}).$$

**Exemple 3.** Limite en loi de  $X_n$  de loi binomiale  $\mathcal{B}(n, p = \frac{\lambda}{n})$ .

PREUVE : 1. L'implication directe est évidente puisque  $C_c(\mathbb{R}^d) \subset C_b(\mathbb{R}^d)$ . Pour la réciproque, on prend  $h_n \in C_c(\mathbb{R}^d)$  suite croissante à valeur  $[0, 1]$  tendant vers 1 (par exemple  $h_n(x) = \max(1 - d(x, B(0, n)), 0)$  qui vaut 1 sur la boule  $B(0, n)$ .) comme  $h_m \phi$  est à support compact  $\mathbf{E}((h_m \phi)(X_n)) \rightarrow_{n \rightarrow \infty} \int_{\mathbb{R}^d} \varphi(x) h_m(x) d\mu(x)$ .

Or par positivité de  $1 - h_m$  :

$$|\mathbf{E}((1 - h_m)(X_n)\varphi(X_n))| \leq \|\varphi\| \mathbf{E}((1 - h_m)(X_n))$$

et de même

$$\left| \int (1 - h_m)\varphi d\mu \right| \leq \|\varphi\| \int (1 - h_m) d\mu.$$

Comme d'habitude, on utilise l'inégalité triangulaire :

$$\begin{aligned} \left| \mathbf{E}[\varphi(X_n)] - \int_{\mathbb{R}^d} \varphi(x) d\mu(x) \right| &\leq \left| \mathbf{E}[(h_m \varphi)(X_n)] - \int_{\mathbb{R}^d} h_m \varphi(x) d\mu(x) \right| \\ &\quad + |\mathbf{E}((1 - h_m)(X_n)\varphi(X_n))| + \left| \int (1 - h_m)\varphi d\mu \right| \end{aligned}$$

donc comme le premier terme tend vers 0, en utilisant la sous additivité de la lim sup et les deux bornes précédentes pour les 2 derniers termes, on a :

$$\limsup_{n \rightarrow \infty} \left| \mathbf{E}[\varphi(X_n)] - \int_{\mathbb{R}^d} \varphi(x) d\mu(x) \right| \leq \limsup_{n \rightarrow \infty} \|\varphi\| \left( E((1 - h_m)(X_n)) + \int (1 - h_m) d\mu \right).$$

L'utilisation de la limsup permet la convergence de  $X_n$  de nouveau sur la borne car  $(E((1 - h_m)(X_n)) = 1 - E(h_m(X_n)) \rightarrow 1 - \int h_m d\mu = \int (1 - h_m) d\mu$ . Donc

$$\limsup_{n \rightarrow \infty} \left| \mathbf{E}[\varphi(X_n)] - \int_{\mathbb{R}^d} \varphi(x) d\mu(x) \right| \leq 2\|\varphi\| \int (1 - h_m) d\mu \xrightarrow{m \rightarrow \infty} 0,$$

et la dernière convergence par convergence monotone est fait alors que le premier terme ne dépend pas de  $m$  et le second pas de  $n$ .

Pour 2.  $P(Y_n = x) = \mathbf{E}[1_{\{x\}}(Y_n)] = \mathbf{E}(h_\epsilon(Y_n))$  pour une fonction  $h_\epsilon$  continue égale à 1 sur  $x$  et 0 en dehors de  $B(x, \epsilon)$  donc 0 sur les autres points de  $\mathbb{Z}^d$  si  $\epsilon \leq 1/2$ ). Donc pour l'implication directe  $P(Y_n = x) \rightarrow \mu(h_\epsilon) \xrightarrow{\epsilon \rightarrow 0} \mu(\{x\})$  car en faisant tendre  $\epsilon \rightarrow 0$ , par convergence monotone on obtient la limite souhaitée. Pour la réciproque Il suffit de noter pour  $h$  a support compacte  $E(h(Y_n)) = \sum_{x \in \mathbb{Z}^d, h(x) \neq 0} h(x) P(Y_n = x)$  qui comme la somme est finie tend vers  $\sum_{x \in \mathbb{Z}^d, h(x) \neq 0} h(x) \mu(\{x\})$ . Ceci montre que  $Y_n$  converge en loi vers  $\sum_{x \in \mathbb{Z}^d} \mu(\{x\}) \delta_x = \mu$  car  $\mu$  est une probabilité sur  $\mathbb{Z}^d$ . (si on ne savait pas cela, on ne pourrait exclure une perte de masse) ■

Le résultat principal est la caractérisation en terme de fonctions caractéristiques :

**Théorème 11** (Lévy). *Soit  $Y_n : \Omega_n \rightarrow \mathbb{R}^d$  une suite de variables aléatoires et  $Y : \Omega \rightarrow \mathbb{R}^d$  une autre v.a., alors  $Y_n$  converge en loi vers  $Y$  si et seulement si :*

$$\forall t \in \mathbb{R}^d, \Phi_{Y_n}(t) \xrightarrow{n \rightarrow \infty} \Phi_Y(t).$$

**Exemple 4.**  $(X_i)$  i.i.d  $\mathcal{N}(0, 1)$  limite en loi de  $\frac{1}{n^{5/2}} \sum_{k=1}^n k^2 X_k$ .

**Exemple 5.** Soit  $(X_n)$  i.i.d de loi de bernoulli  $\mathcal{B}(1/2)$ . Mq  $\sum_{k=1}^n \frac{1}{2^k} X_k \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \mathcal{U}([0, 1])$ .

*Remarque 1.* Comme la convergence est partout absolue, la limite  $U$  redonne  $X_k = Ent(2^k U) - 2Ent(2^{k-1} U)$  sont indépendants pour  $U$  uniforme. La mesure de Lebesgue donne donc une suite de v.a. de Bernoulli indépendante. et on construit à partir de là n'importe quelle suite de v.a. indépendantes.

*Démonstration.* Comme  $x \rightarrow e^{itx}$  est une fonction continue bornée, l'implication suit directement de la définition. Réciproquement, soit  $h \in C_c(\mathbb{R}^d)$ , on a vu dans la preuve du théorème d'injectivité de la transformée de Fourier que

$$\mathbf{E}((h * g_\sigma)(X)) = \int_{\mathbb{R}^{2d}} dx dv \frac{h(x)}{(2\pi)^d} \exp\left(-\frac{\sigma^2 \|v\|^2}{2} - i\langle x, v \rangle\right) \Phi_Y(v)$$

et de même pour  $Y_n$  comme  $h$  est à support compacte  $h(x) \exp(-\frac{\sigma^2 \|v\|^2}{2})$  est intégrable par rapport à la mesure de Lebesgue et donne une domination (vu  $|\Phi_{Y_n}(v)| \leq 1$ ). Or par hypothèse  $\Phi_{Y_n}(v) \rightarrow \Phi_Y(v)$  donc la convergence dominée :

$$\mathbf{E}((h * g_\sigma)(Y_n)) \rightarrow \mathbf{E}((h * g_\sigma)(Y)).$$

Or, d'après la preuve du lemma 9,  $(h * g_\sigma)$  converge uniformément vers  $h$  car  $h$  est uniformément continue (car à support compact et par le Théorème de Heine) donc pour  $\epsilon$  fixé on a  $\|(h * g_\sigma) - h\| \leq \epsilon$

D'où  $\limsup_{n \rightarrow \infty} |\mathbf{E}(h(Y_n)) - \mathbf{E}(h(Y))| \leq 2\epsilon + \limsup_{n \rightarrow \infty} |\mathbf{E}((h * g_\sigma)(Y_n)) - \mathbf{E}((h * g_\sigma)(Y))| = 2\epsilon \rightarrow_{\epsilon \rightarrow 0} 0$ . Donc  $\mathbf{E}(h(Y_n)) \rightarrow \mathbf{E}(h(Y))$  pour tout  $h$  continue à support compact et, par la proposition précédente, alors  $Y_n$  converge en loi vers  $Y$ .  $\square$

## 2.1 Version forte du Théorème de Paul Lévy [Preuve facultative]

Dans la version forte, on n'a pas besoin de supposer que la limite des fonctions caractéristiques est une fonction caractéristique, on obtient ce résultat. On a aussi le résultat utile pour les transformées de Laplace.

**Théorème 12** (Lévy). 1. Soit  $Y_n : \Omega_n \rightarrow \mathbb{R}^d$  une suite de vecteurs aléatoires, on suppose que

$$\forall t \in \mathbb{R}^d, \Phi_{Y_n}(t) \rightarrow_{n \rightarrow \infty} \varphi(t)$$

et que sa limite  $\varphi$  est continue en  $t = 0$ .

Alors,  $Y_n$  converge en loi vers une probabilité  $\mu$  sur  $\mathcal{B}(\mathbb{R}^d)$  et  $\varphi(t) = \int_{\mathbb{R}^d} e^{i\langle x, t \rangle} d\mu(x)$ .

2. Soit  $Y_n : \Omega_n \rightarrow \mathbb{R}_+$  une suite de variables aléatoires positives, on suppose que la transformée de Laplace  $\mathcal{L}_{Y_n}(t) = \mathbf{E}[\exp(-Y_n t)]$  converge :

$$\forall t \in \mathbb{R}_+, \mathcal{L}_{Y_n}(t) \rightarrow_{n \rightarrow \infty} h(t)$$

et que sa limite  $h$  est continue en  $t = 0$ .

Alors,  $Y_n$  converge en loi vers une probabilité  $\mu$  sur  $\mathcal{B}([0, \infty[)$  dont la transformée de Laplace est  $h(t) = \int_0^\infty e^{-tx} d\mu(x)$ .

**Exemple 6.** Si  $Y_n \sim \mathcal{N}(0, n)$ ,  $\Phi_{Y_n}(t) \rightarrow 1_{\{0\}}(t)$  mais  $Y_n$  ne converge pas en loi,  $\frac{Y_n}{1+|Y_n|} \xrightarrow{\mathcal{L}}_{k \rightarrow \infty} \frac{\delta_1 + \delta_{-1}}{2}$  ( $Y_n$  perd donc moitié de sa masse en  $+\infty$ , moitié de sa masse en  $-\infty$ ,)

La preuve est assez technique et donc FACULTATIVE (la suite de cette section, définitions comprises est donc facultative). L'idée est en 2 étapes, de montrer que la condition de continuité implique qu'aucune masse ne peut s'échapper vers l'infini. On rend ceci précis avec la notion de tension (suite tendue). Ensuite, on voit comment la tension implique une notion de compacité pour la convergence faible et on conclut.

**Définition 9.** Une suite  $X_n : \Omega \rightarrow \mathbb{R}^d$  de vecteurs aléatoires est **tendue** si pour tout  $\epsilon > 0$  il existe  $K$  compact tel que, pour tout  $n : P(X_n \in K) \geq 1 - \epsilon$ .

**Lemme 13.** Pour toute v.a.  $X : \Omega \rightarrow \mathbb{R}$ , resp. tout v.a. positive  $Y : \Omega \rightarrow \mathbb{R}_+$ ,

$$P(|X| \geq \frac{2}{u}) \leq \frac{1}{u} \int_{-u}^u (1 - \operatorname{Re} \Phi_X(t)) dt, \quad P(Y \geq \frac{1}{u}) \leq \frac{e}{u} \int_0^u (1 - \mathcal{L}_Y(t)) dt.$$

En conséquence, si une suite satisfait l'hypothèse du Thm 12.1 ou 2., elle est tendue.

PREUVE : Un calcul donne  $\int_{-u}^u du(1 - e^{itx}) dt = 2u - \frac{2 \sin(ux)}{x}$  donc en utilisant Fubini (et introduisant la partie réelle sans risque puisque l'expression est déjà réelle) :

$$\frac{1}{u} \int_{-u}^u du(1 - \operatorname{Re} \Phi_X(t)) dt \geq 2 \int_{\mathbb{R}} \left(1 - \frac{\sin(ux)}{ux}\right) dP_X(x) \geq 2 \int_{|x|u \geq 2} \left(1 - \frac{\sin(ux)}{ux}\right) dP_X(x) \geq P(|X| \geq \frac{2}{u})$$

car  $1 - \frac{\sin(ux)}{ux} \geq 1/2$  si  $|x|u \geq 2$  et est toujours positif. Pour l'application aux coordonnées de la suite  $Y_n$ , on obtient pour  $u = 2/M$  :  $\limsup_{n \rightarrow \infty} P(|(Y_n)_i| \geq M) \leq \limsup_{n \rightarrow \infty} \frac{M}{2} \int_{-2/M}^{2/M} (1 - \operatorname{Re} \Phi_{Y_n}(0, \dots, t_i, 0, \dots, 0)) dt_i = \frac{M}{2} \int_{-2/M}^{2/M} (1 - \operatorname{Re} \varphi(0, \dots, t_i, 0, \dots, 0)) dt_i$ , en utilisant le théorème de convergence dominée (par une constante) pour le calcul de la limite. Pour  $\epsilon$  fixé dans la définition de la tension, on trouve,  $M$  tel que pour tout  $i$ , l'intégrande (donc l'intégrale) est uniformément bornée par  $\epsilon/2d$ , donc par définition de la limsup, on trouve  $k$  tel que  $\forall n \geq k, P(\exists i, |(Y_n)_i| \geq M) \leq d(\epsilon/d) = \epsilon$ . En choisissant  $M$  plus grand on réalise de même pour le nombre fini des premières coordonnées (un nombre fini de v.a. est toujours tendue dans  $\mathbb{R}^d$  par convergence dominée  $P(\exists i, |(Y_n)_i| \geq M) \rightarrow_{M \rightarrow \infty} 0$ ). On conclut vu que la boule de rayon  $M$  pour la norme infini est compacte.

Dans le cas positif,  $\frac{1}{u} \int_0^u du(1 - e^{-tx}) = \frac{e^{-ux} - 1 + ux}{ux} \geq e^{-1} 1_{ux \geq 1}$  (par un tableau de variation) et le raisonnement est le même. ■

Le résultat clé qui donne de la compacité est le suivant (voir Barbe Ledoux p 128 pour une preuve). Il est surtout basé sur un résultat d'analyse fonctionnelle : le Théorème de Banach-Alaoglu sur la compacité dans les espaces duaux (dont on verra un cas particulier pour les espaces  $L^p$  dans la section 1.3 du chapitre 5, elle-même facultative).

**Théorème 14** (Prokhorov). *Soit  $X_n : \Omega \rightarrow \mathbb{R}^d$  une suite tendue de vecteurs aléatoires. Alors il existe une sous-suite  $X_{n_k}$  et une mesure de probabilité  $\mu$  telle que  $X_{n_k} \xrightarrow{L}_{k \rightarrow \infty} \mu$ .*

PREUVE DU THM 12 : 1. Par le théorème et le lemme, toute sous-suite de  $Y_n$  a une sous-suite convergeant en loi vers une mesure de proba. Par le sens simple de l'implication du premier Thm de Lévy, sa transformée de Fourier est  $\varphi$ . Donc on a maintenant l'hypothèse du sens difficile de ce premier théorème pour conclure la convergence en loi vers cette proba  $\mu$  de transformée de Fourier  $\varphi$ .

2. On montre de même la tension par le lemme, on en déduit que toute sous-suite de  $Y_n$  a une sous-suite convergeant en loi vers une mesure de proba  $\mu$ . Par la réciproque simple, la transformée de Laplace de cette proba est  $h$ . Donc toutes les sous-suites ont la même limite en loi (on utilise théorème d'inversion pour la transfo de Laplace, cf TD 2 pour déduction à partir de Fourier). Prenons  $f$  continue bornée  $E(f(Y_n))$  a une sous-suite convergente par compacité dans  $\mathbb{R}$ , et toutes ces sous-suites convergent vers la même limite  $\int f d\mu$ , on en déduit aisément par l'absurde que la suite de départ doit converger vers cette limite (sans quoi on construit une sous-suite loin de cette limite qui n'a donc pas de sous-suites convergeant vers cette limite.) Cette convergence donne la convergence en loi. ■

## 2.2 Suppléments sur la convergence en loi.

On a aussi les caractérisations équivalentes suivantes (admisses) de la convergence en loi.

**Théorème 15.** 1. *Soit  $Y_n : \Omega_n \rightarrow \mathbb{R}$  une suite de variables aléatoires et  $\mu$  une probabilité.  $Y_n$  converge en loi vers  $\mu$  si et seulement si pour tout*

$$\forall x \in \mathbb{R}, \mu(\{x\}) = 0 \Rightarrow \lim_{n \rightarrow \infty} P(Y_n \leq x) = \mu(-\infty, x].$$

2. (Lemme de Skohokhod)/[facultatif] Soient  $Y_n : \Omega_n \rightarrow \mathbb{R}^d$  une suite de vecteurs aléatoires et  $Y$  un vecteur aléatoire,  $Y_n \xrightarrow{L}_{n \rightarrow \infty} Y$  si et seulement si il existe un espace probabilisé  $(\Omega, \mathcal{T}, P)$ ,  $X_n, X : \Omega \rightarrow \mathbb{R}^d$  tels que  $P_{X_n} = P_{Y_n}, P_X = P_Y$  et  $X_n \xrightarrow{p.s.}_{n \rightarrow \infty} X$ .

Dans le premier énoncé, il est important que l'on ne considère les limites que pour les points tels que  $\mu(\{x\}) = 0$ .

**Exemple 7.** Par exemple, Soit  $P_{Y_n} = \mathcal{N}(0, 1/n)$ , on a vu que  $Y_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \delta_0$ . Mais  $P(Y_n \leq 0) = 1/2$  par symétrie qui ne converge pas vers  $\delta_0([-\infty, 0]) = 1$  car 0 est un point tel que  $\delta_0(\{0\}) = 1 \neq 0$ .

**Exemple 8.**  $x_n \rightarrow x$  si et seulement si  $x_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \delta_x$  (moins évident avec Fourier...)

IDÉES DE PREUVE : 1. On peut prendre  $Y$  de loi  $\mu$ . Pour l'implication directe, on prend  $f_m(x) = 1$  si  $x \leq y - 1/m$ , 0 si  $x \geq y$  continue linéaire entre les deux. de même  $g_m(x) = 1$  si  $x \leq y$ , 0 si  $x \geq y + 1/m$  continue linéaire entre les deux. on a  $\mathbf{E}(f_m(Y_n)) \rightarrow \mathbf{E}(f_m(Y))$   $\mathbf{E}(g_m(Y_n)) \rightarrow \mathbf{E}(g_m(Y))$ .

Puis on utilise le bon comportement de la limsup et liminf avec les inégalités et  $f_m \leq 1_{]-\infty, y]} \leq g_m$  d'où

$$\begin{aligned} \limsup_{n \rightarrow \infty} P(Y_n \leq y) &\leq \limsup_{n \rightarrow \infty} \mathbf{E}(g_m(Y_n)) = \mathbf{E}(g_m(Y)) \rightarrow_{m \rightarrow \infty} P(Y \leq y) \\ \liminf_{n \rightarrow \infty} P(Y_n \leq y) &\geq \liminf_{n \rightarrow \infty} \mathbf{E}(f_m(Y_n)) = \mathbf{E}(f_m(Y)) \rightarrow_{m \rightarrow \infty} P(Y < y) \end{aligned}$$

les limites en  $m$  par convergence monotone ou dominée (par une constante). Les deux limites sont égales si  $P(Y = y) = \mu(\{y\}) = 0$ .

Pour la réciproque, on utilise que 2 n'utilise que cette conséquence, et on utilise que convergence p.s. implique convergence en loi.

Pour 2,  $d = 1$  on utilise un modèle canonique concret (sur  $[0,1]$ ) utilisant l'inverse de la fonction de répartition. ■

### 3 Une application : le théorème central limite

Pour le cas vectoriel du théorème central limite, on a besoin de la définition d'un vecteur gaussien que l'on étudiera plus au prochain chapitre. Pour  $X = (X_1, \dots, X_d)$  vecteur de variable dans  $L^2(\Omega)$  (i.e.  $E(X_i^2) < \infty$ ) on note  $\mathbf{Cov}(X)$  la matrice de covariance  $(\mathbf{Cov}(X))_{i,j} = Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$ . On rappelle qu'une matrice symétrique réelle est positive si ces valeurs propres sont toutes positives. Il est facile de voir que les covariances sont des matrices positives. On note aussi  $\mathbf{E}(X) = (\mathbf{E}(X_1), \dots, \mathbf{E}(X_d))$  le vecteur des moyennes.

**Définition 10.** Soit  $C \in M_d(\mathbb{R})$  une matrice symétrique positive. Un vecteur aléatoire  $X_i : \Omega \rightarrow \mathbb{R}^d$  avec  $X_i \in L^2$  est appelé **vecteur gaussien centré de covariance**  $C$ , dite loi  $\mathcal{N}(0, C)$ , si

$$\forall t \in \mathbb{R}^d, \Phi_X(t) = \mathbf{E}[e^{i\langle t, X \rangle}] = \exp\left(-\frac{\langle t, Ct \rangle}{2}\right).$$

On rappelle que  $\langle t, s \rangle = \sum_i t_i s_i$  est le produit scalaire et donc que  $\langle t, Ct \rangle = \sum_{i,j} t_i C_{ij} t_j$ . On remarquera que cela coïncide avec la loi normale usuelle dans le cas  $d = 1$ .

On verra l'existence des vecteurs gaussiens au prochain chapitre et une construction simple. Sinon, l'application du théorème de Levy fort dans le théorème ci-dessous donnera aussi l'existence dès que  $C$  sera la matrice de covariance d'une loi de probabilité. (Et on verra encore au prochain chapitre que toute matrice réelle symétrique positive peut être obtenue). L'utilisation du théorème de Paul Lévy va être basée sur un développement limité des fonctions caractéristiques.

**Lemme 16.** Soit un vecteur aléatoire  $X : \Omega \rightarrow \mathbb{R}^d$  avec  $X_i \in L^2$ . Alors  $\Phi_X$  est de classe  $\mathcal{C}^2$  et son développement de Taylor en 0 est :

$$\Phi_X(t) = 1 + i \sum_{j=1}^d t_j E[X_j] - \frac{1}{2} \sum_{k,j=1}^d t_k t_j E[X_k X_j] + o(\|t\|^2).$$

PREUVE :  $\varphi : (t, \omega) \rightarrow e^{it\langle t, X(\omega) \rangle}$  est intégrable en  $\omega$  et  $\mathcal{C}^\infty$  en  $t$ , et ces deux premières dérivées sont :

$$\frac{\partial \varphi}{\partial t_j} = iX_j e^{it\langle t, X(\omega) \rangle}, \quad \frac{\partial^2 \varphi}{\partial t_k \partial t_j} = -X_k X_j e^{it\langle t, X(\omega) \rangle},$$

qui sont dominés par

$$\left| \frac{\partial \varphi}{\partial t_j} \right| \leq |X_j|, \quad \left| \frac{\partial^2 \varphi}{\partial t_k \partial t_j} \right| \leq |X_k X_j|$$

qui sont intégrables car  $X_i \in L^2$  et en particulier par Cauchy-Schwartz,  $\mathbf{E}[|X_j X_k|] \leq \|X_j\|_2 \|X_k\|_2$ . Donc d'après le théorème de dérivation avec condition de domination  $\Phi_X(t) = \int \varphi(t, \omega) dP(\omega)$  est  $\mathcal{C}^2$  et ces dérivées sont :

$$\frac{\partial \Phi_X}{\partial t_j}(t) = \mathbf{E}[iX_j e^{it\langle t, X \rangle}], \quad \frac{\partial^2 \Phi_X}{\partial t_k \partial t_j}(t) = -\mathbf{E}(X_k X_j e^{it\langle t, X \rangle}).$$

L'énoncé vient donc de la formule de Taylor-Young en 0. ■

**Théorème 17.** Soit  $(X_n)$  une suite de vecteurs aléatoires  $X : \Omega \rightarrow \mathbb{R}^d$  indépendants de même loi et dans  $L^2$ . Alors

$$\frac{(X_1 + \dots + X_n - n\mathbf{E}[X_1])}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \mathbf{Cov}(X_1)).$$

PREUVE : En remplaçant  $X_i$  par  $X_i - \mathbf{E}[X_i]$  on peut supposer et on suppose  $\mathbf{E}[X_1] = 0$ . On calcule la fonction caractéristique en utilisant d'abord l'indépendance (cf lemme 8) puis le fait que les  $X_i$  ont même loi.

$$\Phi_{\frac{X_1 + \dots + X_n}{\sqrt{n}}}(t) = \mathbf{E} \left[ \exp(i\langle t, \frac{X_1 + \dots + X_n}{\sqrt{n}} \rangle) \right] = \prod_{j=1}^n \mathbf{E} \left[ \exp(i\langle t, \frac{X_j}{\sqrt{n}} \rangle) \right] = \left( \Phi_{X_1} \left( \frac{t}{\sqrt{n}} \right) \right)^n.$$

En utilisant le dL du lemme précédent, on obtient une suite (complexe)  $\epsilon_n \rightarrow 0$  telle que

$$\Phi_{X_1} \left( \frac{t}{\sqrt{n}} \right) = \left( 1 - \frac{\langle t, \mathbf{Cov}(X_1)t \rangle}{2n} + \frac{\epsilon_n}{n} \right) = \left( 1 - \frac{\langle t, \mathbf{Cov}(X_1)t \rangle}{2n} \right) \left( 1 + \frac{\epsilon_n}{n} \left( 1 - \frac{\langle t, \mathbf{Cov}(X_1)t \rangle}{2n} \right)^{-1} \right).$$

Si on pose  $\eta_n = \epsilon_n \left( 1 - \frac{\langle t, \mathbf{Cov}(X_1)t \rangle}{2n} \right)^{-1} \rightarrow 0$ , vérifions que  $(1 + \frac{\eta_n}{n})^n \rightarrow 1$  (ce qui est bien connu pour le cas réel en utilisant le log mais peut être moins dans le cas  $\eta_n$  complexe.) = On utilise la formule du binôme :

$$\left| \left( 1 + \frac{\eta_n}{n} \right)^n - 1 \right| \leq \sum_{k=1}^n C_n^k \frac{|\eta_n|^k}{n^k} = \left| \left( 1 + \frac{|\eta_n|}{n} \right)^n - 1 \right| \leq |\eta_n| \left( 1 + \frac{|\eta_n|}{n} \right)^{n-1} \rightarrow 0,$$

la deuxième inégalité venant du théorème des accroissement fini pour  $x > 1$ ,

$$|f(x) - f(1)| = |(1+x)^n - 1| \leq |x-1| \sup f'(x) = |x-1| n(1+x)^{n-1}.$$

En conclusion

$$\Phi_{\frac{X_1 + \dots + X_n}{\sqrt{n}}}(t) \sim_{n \rightarrow \infty} \left( 1 - \frac{\langle t, \mathbf{Cov}(X_1)t \rangle}{2n} \right)^n = \exp[n \ln \left( 1 - \frac{\langle t, \mathbf{Cov}(X_1)t \rangle}{2n} \right)] \rightarrow_{n \rightarrow \infty} \exp\left(-\frac{\langle t, \mathbf{Cov}(X_1)t \rangle}{2}\right).$$

Comme la limite est la fonction caractéristique de la loi gaussienne centrée de covariance  $\mathbf{Cov}(X_1)$  le théorème de Paul Lévy conclut. ■

**Exemple 9.** Mq  $u_n = e^{-n} \sum_{k=0}^n \frac{n^k}{k!} \rightarrow 1/2$ .

# Chapitre 2

## Vecteurs Gaussiens.

### 1 Définition générale et transformée de Fourier

**Définition 11.** Un vecteur aléatoire  $X : \Omega \rightarrow \mathbb{R}^d$  est un **vecteur gaussien** si toute combinaison linéaire  $\sum_{i=1}^d \lambda_i X_i$  avec  $\lambda_i \in \mathbb{R}$  est de loi normale sur  $\mathbb{R}$ .

(nécessairement  $\mathcal{N}(\sum_{i=1}^d \lambda_i \mathbf{E}(X_i), \sum_{i,j=1}^d \lambda_i \lambda_j \text{Cov}(X_i, X_j))$ .)

**Théorème 18.** Un vecteur aléatoire  $X : \Omega \rightarrow \mathbb{R}^d$  est un **vecteur gaussien** si et seulement si sa fonction caractéristique est

$$\Phi_X(t) = \exp(i\langle m, t \rangle - \frac{1}{2}\langle t, Ct \rangle)$$

et alors  $m = E(X)$ ,  $C = \mathbf{Cov}(X)$ . Dans ce cas on note  $X \sim \mathcal{N}(m, C)$

La notation et la définition sont donc cohérentes avec la définition 10 d'un vecteur gaussien centré intervenant dans le TCL.

PREUVE : Par définition  $\Phi_X(t) = \mathbf{E}[e^{i\langle X, t \rangle}] = \Phi_{\langle X, t \rangle}(1)$  or par le lemme 5 (et injectivité de la transformée de Fourier),  $\langle X, t \rangle = \sum_{i=1}^d t_i X_i$  est de loi  $\mathcal{N}(\sum_{i=1}^d t_i \mathbf{E}(X_i), \sum_{i,j=1}^d t_i t_j \text{Cov}(X_i, X_j))$  si et seulement si sa transformée de Fourier est  $\Phi_{\langle X, t \rangle}(l) = \exp(i\langle m, lt \rangle - \frac{l^2}{2}\langle t, Ct \rangle)$  soit la valeur dans la caractérisation. ■

**Corollaire 19.** Deux vecteurs gaussiens ont la même loi si et seulement si ils ont même espérance et même matrice de covariance.

**Corollaire 20.** Si un vecteur aléatoire  $X : \Omega \rightarrow \mathbb{R}^d$  est gaussien, alors pour toute application linéaire  $A : \mathbb{R}^d \rightarrow \mathbb{R}^m$  et tout vecteur  $n \in \mathbb{R}^m$ ,  $Y = n + A(X) : \Omega \rightarrow \mathbb{R}^m$  est un vecteur gaussien. Plus précisément, si  $X \sim \mathcal{N}(m, C)$  alors  $Y \sim \mathcal{N}(n + A(m), ACA^T)$ .

On rappelle que  $(A^T)_{ij} = A_{ji}$  désigne la transposée de  $A$ .

PREUVE : Pour tout  $\lambda_1, \dots, \lambda_m$ ,  $\sum_{i=1}^m \lambda_i Y_i = \sum_{i=1}^m \lambda_i m_i + \sum_{i=1}^m \sum_{j=1}^d \lambda_i A_{ij} X_j = c + \sum_{j=1}^d \mu_j X_j$  avec  $\mu_j = \sum_{i=1}^m \lambda_i A_{ij}$  c'est donc la translaté d'une v.a. gaussienne donc une v.a. gaussienne de  $\mathbb{R}$ . Comme les  $\lambda_i$  sont arbitraires,  $Y$  est gaussien. Il ne reste qu'à calculer son espérance et sa variance. Par linéarité  $\mathbf{E}(Y) = n + A\mathbf{E}(X)$  et par bilinéarité pour la covariance

$$\mathbf{Cov}(Y_i, Y_j) = \sum_{k=1}^d \sum_{l=1}^d A_{ik} \text{Cov}(X_k, X_l) A_{jl} = (ACA^T)_{ij}$$

■

**Exemple 10.** Soit  $(X_1, X_2, X_3)$  un vecteur gaussien de loi  $\mathcal{N}(m, \Gamma)$  avec  $m = (4, 1, 0)$  et  $\Gamma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix}$ . Loi de  $Y = (X_2, 3X_3 - 2X_2)$  ?

**Exemple 11.** Soit  $(X_1, X_2, X_3, Y_1, Y_2, Y_3)$  de loi  $\mathcal{N}(0, I_6)$  quelle est la loi de  $Z = (\frac{\sum_{i=1}^3 X_i Y_i}{\sqrt{\sum_{i=1}^3 Y_i^2}}, Y_1, Y_2, Y_3)$  ? (attention la transformation n'est PAS linéaire et on ne peut pas appliquer le résultat précédent.) Montrer que  $Z \sim \mathcal{N}(0, I_4)$ .

**Exemple 12.** (cf TD) si  $(X, Y) \sim \mathcal{N}(0, I_2)$ ,  $P_{X/Y}(dz) = \frac{1}{\pi(1+z^2)}dz$  est la loi de Cauchy.

On va donner une construction canonique d'un vecteur  $\mathcal{N}(m, C)$  en utilisant la forme du corollaire précédent.

## 2 Construction et simulation.

On appelle *variable gaussienne standard* notée  $\gamma_d$  une variable  $\mathcal{N}(0, I_d)$  sur  $\mathbb{R}^d$ . On rappelle que pour une matrice positive  $C = ODO^T$  dont on a une diagonalisation ( $D = \text{diag}(\lambda_1, \dots, \lambda_d)$  diagonale à coefficient positif,  $O$  orthogonale) la racine carrée de  $C$  est  $\sqrt{C} = O\sqrt{D}O^T$  avec  $D = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})$ . On a donc  $\sqrt{C}$

**Théorème 21.** Si  $X = (X_1, \dots, X_d)$  variable gaussienne standard  $\gamma_d$ ,  $m \in \mathbb{R}^d$  et  $C \in M_d(\mathbb{R})$  une matrice positive. Soit  $B = \sqrt{C}$  la racine carrée de  $C$ . Alors  $Y = m + BX$  est un vecteur gaussien de loi  $\mathcal{N}(m, C)$ .

Cela donne une construction d'un vecteur  $Y \sim \mathcal{N}(m, C)$ . La preuve est une application directe du dernier corollaire.

Pour simuler un vecteur gaussien  $\mathcal{N}(m, C)$ , il suffit donc de savoir simuler un vecteur gaussien standard  $X$  et de considérer  $Y = m + \sqrt{C}X$ . On peut soit construire  $\sqrt{C}$  par diagonalisation comme ci-dessus. Il existe cependant une méthode plus efficace de construction de vecteur gaussiens basé sur le résultat suivant.

**Proposition 22** (Décomposition de Cholesky). Soit  $C$  une matrice symétrique positive, il existe une matrice réelle triangulaire inférieure  $L$  tel que  $C = LL^T$ . En conséquence, si  $X = (X_1, \dots, X_d)$  variable gaussienne standard  $\gamma_d$ ,  $m \in \mathbb{R}^d$   $Y = m + LX$  est un vecteur gaussien de loi  $\mathcal{N}(m, C)$ .

*Démonstration.* On suppose d'abord que  $C$  est inversible. La relation  $LL^T = C$  est équivalent au système d'équation :  $C_{i,j} = \sum_{k=1}^{\min(i,j)} L_{ik}L_{jk}$  qu'il suffit de vérifier pour  $i \leq j$  puisque  $C$  est symétrique, ce qui donne :

$$L_{i,i}L_{j,i} = C_{i,j} - \sum_{k=1}^{i-1} L_{ik}L_{jk}$$

ce qui se résout par récurrence sur  $i$  en obtenant à chaque étape  $L_{mi}$  pour tout  $m \geq i$ .

Il suffit donc de définir inductivement

$$L_{1,1} = \sqrt{C_{1,1}}, L_{j,1} = C_{1,j}/L_{1,1}, j > 1$$

vu  $L_{1,1} > 0$  (comme  $C$  inversible). En général au rang  $i$  on veut poser :

$$L_{ii} = \sqrt{C_{i,i} - \sum_{k=1}^{i-1} L_{ik}^2}$$



Prenons d'abord une racine carré complexe et voyons qu'on peut prendre la racine carré strictement positive (en supposant avoir obtenu cela par récurrence). C'est possible car en se restreignant à la sous matrice  $i \times i$   $C^{(i)}$  formé des  $i$  premières colonnes des  $i$  premières lignes, on a obtenu  $C^{(i)} = L^{(i)}(L^{(i)})^T$  avec  $L_{k,l}^{(i)} = L_{k,l}$  tant que  $k, l \leq i$ . En prenant le déterminant  $\det(C^{(i)}) = \det(L^{(i)})^2 = \prod_{j=1}^i L_{j,i}^2$  et les deux côtés ont des facteurs strictement positifs (car  $C^{(i)}$  définie positive et par l'hypothèse de récurrence). Donc  $L_{i,i} > 0$  et on continue la résolution par

$$L_{i,j} = (C_{i,j} - \sum_{k=1}^{i-1} L_{ik}L_{j,k})L_{i,i}^{-1}$$

pour  $j \geq i$  ce qui donne les  $i$  premières colonnes de  $L$ .

Si  $C$  n'est pas défini-positive, on obtient  $L_n$  associé à  $C_n = C + I_n/n$  ( $I_n$  la matrice identité). On a  $\|L_n\|^2 = \|L_n L_n^T\| \leq \|C\| + 1$  est une suite borné donc par compacité (boule unité fermé est compact en dimension finie) on extrait une sous suite  $L_{n_k} \rightarrow L$  la relation  $C_n = L_n L_n^T$  donne en passant à la limite  $C = LL^T$  (on utilise la continuité du produit pour la norme d'opérateur). La conséquence pour les variables gaussiennes est évidente.  $\square$

On rappelle maintenant 2 façons de simuler les vecteurs gaussiens standard.

**Exemple 13** (Box Muller). Si  $U \sim \mathcal{E}(1/2)$ ,  $V \sim \mathcal{U}([0, 1])$  indépendantes,  $\mathbf{Mq}$ ,  $X = \sqrt{U} \cos(2\pi V)$ ,  $Y = \sqrt{U} \sin(2\pi V)$  sont indépendantes et de loi normales centrés réduites.

Si on prend donc  $V, W \sim \mathcal{U}([0, 1])$  indépendantes, alors

$$X = \sqrt{-2 \log(V)} \cos(2\pi W), \quad Y = \sqrt{-2 \log(V)} \sin(2\pi W)$$

donne un vecteur gaussien standard  $(X, Y) \sim \gamma_2$ . En pratique l'évaluation de  $\cos, \sin, \log$  rend cette méthode couteuse et soumise à erreurs numériques. Une méthode évitent l'évaluation du cosinus et du sinus utilise la loi uniforme sur le disque :

**Proposition 23** (Loi uniforme sur le disque). SI  $(U, V)$  un couple de variable distribué selon la loi uniforme sur  $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ . Soit  $(U, V) = (\rho U_0, \rho V_0)$  avec  $\rho = \sqrt{U^2 + V^2}$ . Alors les variables

$$X = 2\sqrt{-\log(\rho)}U_0, \quad Y = 2\sqrt{-\log(\rho)}V_0$$

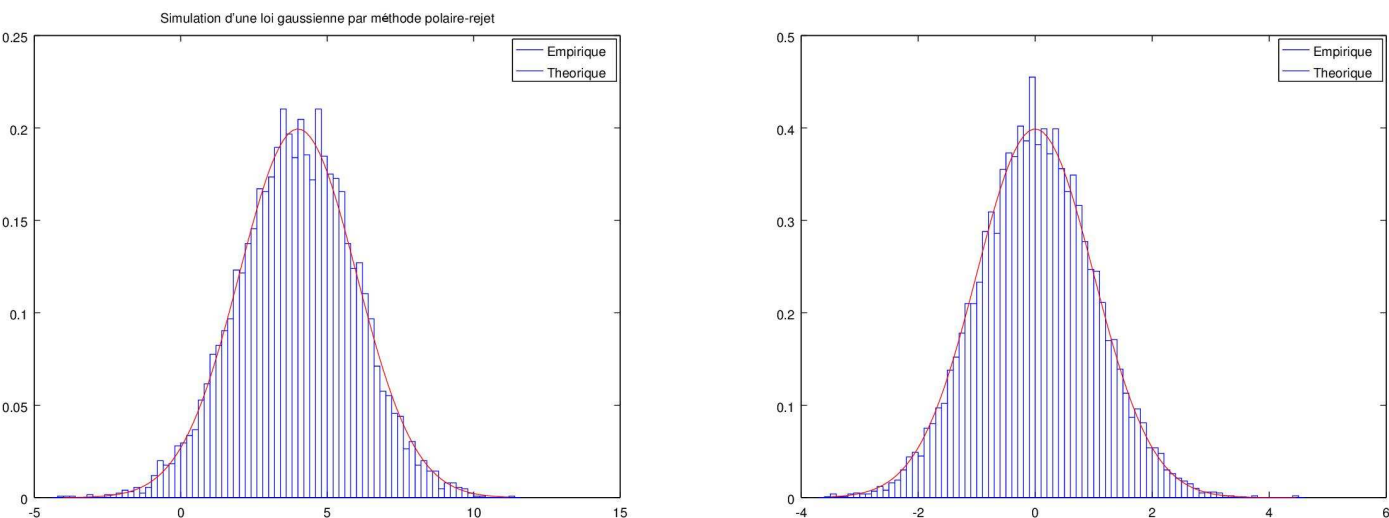
sont distribués selon la loi normale standard :  $(X, Y) \sim \gamma_2$

*Démonstration.* On écrit  $(U_0, V_0) = (\cos(\theta), \sin(\theta))$  avec  $\theta \in [0, 2\pi[$ . Par l'exemple précédent il suffit de voir que  $(\sqrt{-\log(\rho)}, \theta)$  sont indépendants de loi respective  $-\log(\rho) \sim \mathcal{E}(2)$ ,  $\theta \sim \mathcal{U}([0, 2\pi])$ . Le fait que  $\theta$  est de loi uniforme vient de l'invariance par rotation de la loi uniforme sur le disque De plus  $\theta$  est déterminé par  $\cos(\theta), \epsilon(\theta)$  où se nombre est le signe du sinus. Il suffit donc de voir que  $-\log(\rho), \cos(\theta), \epsilon(\theta)$  indépendant et que le premier a la bonne loi. SOit  $f$  mesurable positive, on calcule par transfert :

$$\begin{aligned} E(f(-\log(\rho), \cos(\theta), \epsilon(\theta))) &= \int dP_{(U,V)}(u, v) f(-\log(u^2 + v^2)/2, u/\sqrt{u^2 + v^2}, \text{signe}(v)) \\ &= \int_{-1}^1 du \int_0^{\sqrt{1-u^2}} dv f(r(u, v), u/\sqrt{u^2 + v^2}, 1) + \int_{-1}^1 du \int_{-\sqrt{1-u^2}}^0 dv f(r(u, v), u/\sqrt{u^2 + v^2}, -1) \\ &= \int_{-1}^1 \frac{du_0}{\sqrt{1-u_0^2}} \int_0^\infty dr e^{-2r} f(r, u_0, 1) + \int_{-1}^1 \frac{du_0}{\sqrt{1-u_0^2}} \int_0^\infty dr e^{-2r} f(r, u_0, -1) \end{aligned}$$

ou on a fait le changement de variable  $r = r(u, v) := -\log(u^2 + v^2)/2, u_0 = u/\sqrt{u^2 + v^2}$  en utilisant  $u^2 + v^2 = e^{-2r}, u = u_0 e^{-r}, v = \pm e^{-r} \sqrt{1 - u_0^2}$ , d'où le jacobien positif  $|J(u, v)(u_0, r)| = e^{-r} \sqrt{1 - u_0^2} e^{-r} + u_0 e^{-r} e^{-r} u_0 / \sqrt{1 - u_0^2} = e^{-2r} / \sqrt{1 - u_0^2}$ . On a donc obtenu (vu la densité produit), l'indépendance souhaitée et la densité de  $-\log(\rho)$  à savoir  $Ce^{-2r} 1_{r>0}$  qui implique bien la loi exponentielle voulue.  $\square$

On sait (Thm de la simulation par méthode de rejet) que si  $(U_n, V_n)$  une suite de vecteurs indépendants de loi uniforme sur  $[-1, 1]^2, T = \inf\{n, (U_n, V_n) \in D\}$  alors  $(U, V) = (U_T, V_T)$  est de loi uniforme sur  $D$  et peut servir de point de départ à la simulation de loi gaussienne par le théorème précédent. Voici une simulation sur un échantillon de 10000 d'une normale  $\mathcal{N}(4, 4)$ (à gauche). A comparer à une simulation par la version par défaut de Matlab/Stixbox 1.29 d'une normale  $\mathcal{N}(0, 1)$  (à droite), (algorithme de Ziggurat).



### 3 Critère d'absolue continuité

**Théorème 24.** Soit  $Y : \Omega \rightarrow \mathbb{R}^d$  un vecteur gaussien de loi  $\mathcal{N}(m, C)$ , alors  $Y$  admet une densité par rapport à la mesure de Lebesgue si et seulement si  $C$  est définie positive (comme  $C$  est toujours positive ssi  $\det(C) \neq 0$ ) et alors  $P_Y(dx) = f_Y(x)dx$  avec sa densité :

$$f_Y(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det(C)}} \exp\left(-\frac{1}{2} \langle C^{-1}(x - m), (x - m) \rangle\right).$$

De plus si  $C$  n'est pas définie positive,  $Y$  est p.s. à valeur dans  $m + \text{Im}(C)$ .

PREUVE : Si  $\det(C) \neq 0$ , on reprend la notation du théorème précédent, et il suffit de montrer que  $Y = m + BX$  a la densité cherchée par changement de variable. En effet, par la forme de la transformée de Fourier de  $X$ , les coordonnées de  $X$  sont indépendantes et la formule pour la densité découle du calcul des transformées de Fourier une variable lemme 5 et du théorème de caractérisation de l'indépendance comme mesure produit :

$$f_X(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|x\|_2^2}{2}\right).$$

On fait donc le changement de variables  $y = m + Bx$  le jacobien est  $B$ ,  $x = B^{-1}(y - m)$  et  $B^{-1}$  existe puisque  $\det(B) = \sqrt{\det(C)}$ , on obtient donc pour  $h$  mesurable positive par transfert puis le changement de variables ci-dessus :

$$\mathbf{E}(h(Y)) = \int_{\mathbb{R}^d} h(m+Bx) \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|x\|_2^2}{2}\right) dx = \int_{\mathbb{R}^d} h(y) \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|B^{-1}(y-m)\|_2^2}{2}\right) \frac{1}{\det(B)} dy.$$

Ce qui donne par identification la densité une fois que l'on a remarqué que  $\|B^{-1}(y-m)\|_2^2 = \langle B^{-1}(y-m), B^{-1}(y-m) \rangle = \langle C^{-1}(y-m), (y-m) \rangle$  (car  $(B^{-1})^T B^{-1} = B^{-2} = C^{-1}$  et  $\det(B) = \sqrt{\det(C)}$ ).

Comme  $C$  est une matrice carré elle est inversible si et seulement si elle est surjective donc il suffit de prouver le dernier résultat pour avoir la contraposée de la réciproque (si  $C$  non inversible, vu son support ayant mesure de Lebesgue 0 comme tout e.v. de plus petite dimension, elle ne peut pas avoir de densité.) Mais comme  $\text{Im}(C) = \text{Im}(B)$ , le modèle concret donne l'image de valeur p.s.  $m + \text{Im}(B)$  (formellement on prend n'importe quelle fonction  $h$  supportée à l'extérieur, par transfert l'intégrale de  $h$  est  $\mathbf{E}(h(m + BX))$  et  $h(m + BX) = 0$  par hypothèse sur le support). ■

On peut décrire explicitement la mesure gaussienne dans tous les cas mais le résultat est moins important donc facultatif.

**Corollaire 25.** (Facultatif) Si  $X = (X_1, \dots, X_d) \sim \mathcal{N}(m, C)$  variable gaussienne avec  $m \in \mathbb{R}^d$  et  $C \in M_d(\mathbb{R})$  une matrice positive. Soit  $I$  l'injection de  $\mathbb{R}^m \simeq \text{Im}(C) \subset \mathbb{R}^d$  telle que  $I(e_i)$  base orthonormale de  $\text{Im}(C)$  et  $D_{ij} = \langle CI(e_i), I(e_j) \rangle$  la matrice induite. On suppose  $m \geq 1$ . Alors pour tout  $h$  mesurable positive

$$\mathbf{E}(h(X)) = \int_{\mathbb{R}^m} dy_1 \dots dy_m \frac{1}{(2\pi)^{m/2} \sqrt{\det(D)}} h(I(y) + m) \exp\left(-\frac{\langle D^{-1}y, y \rangle}{2}\right).$$

PREUVE : En identifiant  $\text{Im}(C)$  à un  $\mathbb{R}^m$   $Y_i = \langle I(e_i), X - m \rangle$  donne en effet un vecteur gaussien centré  $(Y_1, \dots, Y_m)$  de covariance  $D$ , qui est la restriction/projection orthogonale de  $C$  à  $\text{Im}(C)$  est de déterminant non nul (car l'orthogonal de  $\text{Im}(C)$  est  $\text{Ker}(C^T) = \text{Ker}(C)$ , donc  $D$  restriction à  $\text{Im}(C)$  est injective (le noyau est dans l'orthogonal) donc bijective par le théorème de rang).  $Y$  nous ramène donc au cas précédent d'une densité. ■

## 4 Indépendance

Le résultat suivant est important. Au contraire des variables générales, pour laquelle l'indépendance n'est pas caractérisée par l'annulation de la covariance, l'indépendance dans un vecteur gaussien est caractérisée de cette façon. On va voir que c'est une conséquence simple de la forme de la fonction caractéristique.

**Théorème 26.** Soit  $(X_1, \dots, X_n, Y_1, \dots, Y_m) : \Omega \rightarrow \mathbb{R}^{m+n}$  un vecteur gaussien. Alors  $X = (X_1, \dots, X_n)$  (ou  $\sigma(X_1, \dots, X_n)$ ) est indépendant de  $Y = (Y_1, \dots, Y_m)$  (ou  $\sigma(Y_1, \dots, Y_m)$ ) si et seulement si

$$\forall i \in [1, n], \forall j \in [1, m], \text{Cov}(X_i, Y_j) = 0.$$

Il est crucial que l'ensemble des vecteurs  $(X_1, \dots, X_n, Y_1, \dots, Y_m)$  soit un vecteur gaussien, et PAS seulement  $(X_1, \dots, X_n)$  et  $(Y_1, \dots, Y_m)$  !!

PREUVE : L'implication, indépendance entraîne covariance nulle, a déjà été notée  $E(X_i Y_j) = E(X_i)E(Y_j)$  si  $X_i, Y_j$  indépendantes. Pour la réciproque, comme  $(X_1, \dots, X_n, Y_1, \dots, Y_m)$  gaussien on peut utiliser la formule de la fonction caractéristique et l'annulation de la partie des bloc reliant les  $X$  et les  $Y$  :

$$\begin{aligned}\Phi_{(X_1, \dots, X_n, Y_1, \dots, Y_m)}(t_1, \dots, t_n, s_1, \dots, s_m) &= \exp(i\langle \mathbf{E}(X), t \rangle + i\langle \mathbf{E}(Y), s \rangle - \frac{1}{2}\langle (t, s), \mathbf{Cov}(X, Y)(t, s) \rangle) \\ &= \exp(i\langle \mathbf{E}(X), t \rangle + i\langle \mathbf{E}(Y), s \rangle - \frac{1}{2}\langle t, \mathbf{Cov}(X)t \rangle - \frac{1}{2}\langle s, \mathbf{Cov}(Y)s \rangle) \\ &= \Phi_X(t)\Phi_Y(s)\end{aligned}$$

Par la caractérisation 7, on obtient l'indépendance souhaitée. ■

**Exemple 14.** On voit que  $Y_1, Y_2$  de l'ex 26 sont indépendantes par le calcul de leur covariance.

## 5 Mesures, espaces et processus gaussiens

Une famille de variables aléatoires  $(X_i)_{i \in I}$  est souvent appelé processus. C'est particulièrement le cas quand  $I = \mathbb{N}$ , ou  $I = \mathbb{R}, I = \mathbb{R}_+$  (il y a alors une intuition temporelle liée à l'ordre qui est l'origine de la terminologie processus).

**Définition 12.** Un processus  $X_t, t \in I$  est un **processus gaussien** (centré) si pour tout  $t_1, \dots, t_n \in I$   $(X_{t_1}, \dots, X_{t_n})$  est un vecteur gaussien (centré).

Un sous-espace vectoriel  $H$  de  $L^2(\Omega, \mathcal{T}, P)$  est un **espace gaussien** (centré), si  $(X_1, \dots, X_n)$  est un vecteur gaussien (centré) pour tout  $X_1, \dots, X_n \in H$  (c'est-à-dire si pour tout  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ ,  $\sum \lambda_i X_i$  est une variable gaussienne).

**Exemple 15.** Une suite  $(X_n)$  de variables  $\mathcal{N}(0, 1)$  i.i.d est un processus gaussien.

**Exemple 16.** Si  $X_i, i \in I$  est un processus gaussien,  $Vect(X_i, i \in I)$  est un espace gaussien. Un espace gaussien  $H$  est lui même un processus gaussien (indiqué par lui même)  $(X)_{X \in H}$ .

**Théorème 27.** Pour tout espace gaussien  $H$ , la fermeture  $\overline{H}^{L^2}$  est encore un espace gaussien.

PROOF : Il suffit de voir que si  $Y = (Y_1, \dots, Y_m) = \lim(X_1^{(n)}, \dots, X_m^{(n)})$  dans  $L^2$  et que les  $X^{(n)} = (X_1^{(n)}, \dots, X_m^{(n)})$  sont des vecteurs gaussiens, il en est de même de  $(Y_1, \dots, Y_m)$ . En effet, la moyenne  $E(X^{(n)}) \rightarrow m$  et la matrice de covariance  $C_{ij}^{(n)} = Cov(X_i^{(n)}, X_j^{(n)}) \rightarrow C$  d'après la convergence dans  $L^2$  (et par Cauchy-Schwartz pour le second). De plus la convergence  $L^2$  implique la convergence en loi, donc par le thm de Paul Lévy, les fonctions caractéristiques convergent

$$\Phi_Y(t) = \lim_{n \rightarrow \infty} \Phi_{X^{(n)}}(t) = \lim_{n \rightarrow \infty} \exp(i\langle t, E(X^{(n)}) \rangle - \frac{\langle C^{(n)}t, t \rangle}{2}) = \exp(i\langle t, m \rangle - \frac{\langle Ct, t \rangle}{2}),$$

donc par la caractérisation en terme de transformée de Fourier  $Y \sim \mathcal{N}(m, C)$ . ■

**Définition 13.** On appelle covariance d'un processus  $X_t, t \in I$ , la fonction sur  $I^2$ ,  $C(t, s) = Cov(X_t, X_s)$ .

**Théorème 28** (Kolmogorov). *(facultatif) Pour toute matrice de covariance  $C$  sur  $I$ , c'est à dire fonction sur  $I^2$  telle que pour tout  $t_1, \dots, t_n \in I$   $(C(t_i, t_j))_{i,j \in [1,n]}$  est une matrice symétrique positive. Il existe un unique processus gaussien centré  $X_t, t \in I$  tel que  $C(t, s) = \text{Cov}(X_t, X_s)$ .*

Enfin, on parle parfois de mesure gaussienne pour généraliser à un espace de Banach séparable le cas des mesures associés aux vecteurs gaussiens de  $\mathbb{R}^n$ . On verra une mesure gaussienne et un processus gaussien particulier au chapitre 5, le mouvement brownien.

**Définition 14.** Une mesure  $\gamma$  sur la tribu des boréliens d'un espace de Banach  $E$  est appelée **mesure gaussienne** si pour toute forme linéaire continue  $L \in E^*$ , sa mesure image  $L_*\gamma(B) = \gamma(L^{-1}(B))$  sur  $\mathbb{C} = \mathbb{R}^2$  est une mesure gaussienne.

# Chapitre 3

## Espérance conditionnelle

### 1 Rappels sur l'espérance et les espaces $L^p$

#### 1.1 Premiers rappels

**Définition 15.** Soit  $X : \Omega \rightarrow \mathbb{R}$  une v.a., l'espérance de  $X$  est notée

$$\mathbf{E}[X] = \int_{\Omega} X(\omega) dP(\omega).$$

Elle est définie dans deux cas :

1. si  $X \geq 0$ , et alors  $\mathbf{E}(X) \in [0, \infty]$ .
2. si  $X \in L^1(\Omega, dP)$  i.e. si  $\mathbf{E}(|X|) = \int |X| dP < \infty$ .

Si  $X : \Omega \rightarrow \mathbb{R}^n$ , on note  $\mathbf{E}(X) = (\mathbf{E}(X_1), \dots, \mathbf{E}(X_n))$ .

La moyenne est une forme linéaire croissante sur l'espace vectoriel  $L^1(\Omega, \mathcal{T}, P)$  des v.a. intégrables. En d'autres termes, on a toutes les propriétés usuelles de l'intégrale, dont les résultats de convergence :

**Proposition 29.** Si  $X$  et  $Y$  sont deux v.a. intégrables, alors

1.  $L^1(\Omega, \mathcal{T}, P)$  est un espace vectoriel et  $\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y)$ .
2.  $X \leq Y \Rightarrow \mathbf{E}[X] \leq \mathbf{E}[Y]$ . En particulier,  $X \geq 0 \Rightarrow \mathbf{E}[X] \geq 0$ .
3.  $|Z| \leq |X|$  implique  $Z$  intégrable.
4.  $|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$ .
5.  $\mathbf{E}[c] = c$ ,  $a \leq X \leq b \Rightarrow a \leq \mathbf{E}[X] \leq b$ .
6. (Convergence monotone TCM) Si  $X_n \geq 0$ ,  $X_n$  suite de v.a.r croissante et tend simplement vers  $X$  alors  $\mathbf{E}(X_n) \rightarrow \mathbf{E}(X)$ .
7. (Lemme de Fatou) Si  $X_n \geq 0$   $\mathbf{E}(\liminf X_n) \leq \liminf \mathbf{E}(X)$ .
8. (Convergence dominée TCD) Si  $|X_n| \leq Y$ ,  $\mathbf{E}[Y] < \infty$  et  $P(X_n \rightarrow X) = 1$  (on dit  $X_n \rightarrow X$  p.s.) alors  $\mathbf{E}(X_n) \rightarrow \mathbf{E}(X)$ .

Vous avez vu durant le cours d'intégration, les espaces

$$L^p(\Omega, \mathcal{T}, P) = \{X : \Omega \rightarrow \mathbb{R} \text{ v.a. } |\mathbf{E}(|X|^p) < \infty\},$$

pour  $p \in [1, \infty[$  (ou plutôt l'espace des classes d'équivalence de fonctions égales sur un ensemble de proba 1), espace vectoriel normé complet pour la norme

$$\|X\|_p = \mathbf{E}(|X|^p)^{1/p}.$$

On définit  $\|X\|_\infty = \lim_{p \rightarrow \infty} \|X\|_p$  et  $L^\infty(\Omega, \mathcal{T}, P) = \{X : \Omega \rightarrow \mathbb{R} \text{ v.a. } \|X\|_\infty < \infty\}$ . On rappelle l'inégalité de Hölder : si  $p, q \in [1, \infty]$  tels que  $1/p + 1/q = 1/r \leq 1$ ,  $X \in L^p, Y \in L^q$  alors  $XY \in L^r$  et

$$\|XY\|_r \leq \|X\|_p \|Y\|_q.$$

On rappelle aussi qu'en prenant  $Y = 1$ , on obtient :

$$L^p(\Omega, \mathcal{T}, P) \subset L^q(\Omega, \mathcal{T}, P) \text{ si } p \geq q.$$

**Définition 16.** Soit  $k \in \mathbb{N}$  Soit  $X$  une v.a. dans  $L^k(\Omega, P)$ . Le *moment d'ordre  $k$*  de  $X$  est défini par la quantité

$$m_k = \mathbf{E}[X^k] = \int_{-\infty}^{\infty} x^k dP_X(x).$$

Il est important de retenir la conséquence suivante du théorème de Fubini :

**Proposition 30.** Soit  $X \geq 0$  une variable aléatoire positive, alors pour  $p \in ]0, \infty[$

$$E(X^p) = \int_0^\infty dt p t^{p-1} P(X > t).$$

PREUVE : On remarque que  $X^p = \int 0^X p t^{p-1} dt = \int 0^\infty p t^{p-1} 1_{\{X > t\}} dt$ . Or par Fubini-Tonelli,

$$E(X^p) = \int dP \int_0^\infty dt p t^{p-1} 1_{\{X > t\}} = \int_0^\infty dt p t^{p-1} \int dP 1_{\{X > t\}} = \int_0^\infty dt p t^{p-1} P(X > t). \quad \blacksquare$$

Tous les espaces  $L^p$  sont implicitement des espaces de (classes d'équivalences p.s. de) fonctions à valeur COMPLEXE !

## 1.2 L'espace de Hilbert $L^2(\Omega)$

La formule

$$\langle f, g \rangle = \mathbf{E}[\bar{f}g]$$

définit un produit scalaire tel que  $\|f\|_2^2 = \langle f, f \rangle$ . On peut donc appliquer à  $L^2(\Omega)$  le résultat vu en MASS 51 (Topologie et convexité) sur les espaces de Hilbert, dont le théorème de projection sur un convexe fermé. On rappelle ici l'énoncé du cas d'un sous-espace vectoriel fermé. Ce sera la base que l'on généralisera aux autres espaces  $L^p$  avec l'espérance conditionnelle dans la section suivante.

**Théorème 31.** Soit  $K \subset L^2(\Omega, \mathcal{T}, P)$  un sous espace vectoriel fermé. Pour tout  $f \in L^2(\Omega, \mathcal{T}, P)$  il existe un unique  $u = P_K(f)$  tel que

$$\|f - u\|_2 = \inf_{v \in K} \|f - v\|_2.$$

De plus c'est l'unique vecteur  $u \in K$  tel que

$$\forall v \in K, \quad \mathbf{E}(\bar{v}(f - u)) = 0$$

c'est à dire tel que  $f - u$  est orthogonal à  $K$  ou encore

$$\forall v \in K, \quad \mathbf{E}(\bar{v}f) = \mathbf{E}(\bar{v}P_K(f)). \quad (PC)$$

$P_K$  est une application linéaire bornée appelée **projection orthogonale sur  $K$** .

On en déduit maintenant le calcul de l'espace vectoriel des formes linéaires continues sur  $L^2$  (voir sous-section 1.2 pour des rappels).

**Théorème 32.** *Soit  $\phi$  une forme linéaire continue sur  $L^2(\Omega, \mathcal{T}, P)$  alors il existe un unique  $f \in L^2(\Omega, \mathcal{T}, P)$  tel que*

$$\forall v \in L^2(\Omega, \mathcal{T}, P), \phi(v) = \mathbf{E}(fv).$$

*De plus, on a l'expression duale pour la norme :*

$$\|f\|_2 = \sup_{\|v\|_2 \leq 1} |\mathbf{E}(fv)|.$$

PREUVE : Soit  $K = \phi^{-1}(0)$  le noyau de  $\phi$ . Si  $K = L^2(\Omega, \mathcal{T}, P)$  alors  $f = 0$  convient. On suppose donc  $K \neq L^2(\Omega, \mathcal{T}, P)$ . Soit donc  $g_0 \in L^2(\Omega, \mathcal{T}, P)$  et  $g = \frac{g_0 - P_K(g_0)}{p} \|g_0 - P_K(g_0)\|_2$  un vecteur de norme 1 et orthogonal à  $H$ . Comme  $\phi$  est une forme linéaire, on s'attend que  $H$  et  $g$  engendrent  $L^2$ , sorte de généralisation du théorème du rang. En effet, soit  $v \in L^2(\Omega, \mathcal{T}, P)$ ,  $w = v - \frac{\phi(v)}{\phi(g)}g$  vérifie  $\phi(w) = \phi(v) - \frac{\phi(v)}{\phi(g)}\phi(g) = 0$  donc  $w \in H$  et  $v = \lambda g + w$  avec  $\lambda = \frac{\phi(v)}{\phi(g)}$ .

On montre donc que  $f = \phi(g)\bar{g}$  convient en montrant l'égalité sur  $v$  quelconque précédent :

$$\mathbf{E}(fv) = \phi(g)\langle g, v \rangle = \phi(g)\langle g, \lambda g + w \rangle = \phi(g)\lambda \|g\|_2^2 = \phi(g)\lambda = \phi(v).$$

L'égalité des normes vient de Hölder, cf proposition 33 ci-dessous. ■

### 1.3 Dualité dans les espaces $L^p$ [Facultatif]

On rappelle du cours MASS 31 (prop 23 p 12) que les formes linéaires continues sur un espace vectoriel normé coïncident avec les formes linéaires lipschitziennes. Pour  $E$  un e.v.n, on note  $E^*$  l'espace des formes linéaires continues avec pour norme la constante de lipschitzianité :

$$\|\phi\|_{E^*} = \sup\{|\phi(x)|; \|x\|_E \leq 1\}.$$

On appelle  $E^*$  l'espace de Banach dual de  $E$  (il est facile de voir que cet espace vectoriel normé est toujours complet).

La conséquence facile de Hölder que l'on utilisera fréquemment est la proposition suivante :

**Proposition 33.** *Soit  $p \in [1, \infty]$ ,  $q$  tel que  $1/p + 1/q = 1$  le coefficient conjugué, alors  $L^q(\Omega, P) \subset (L^p(\Omega, P))^*$  en identifiant  $g \in L^q(\Omega, P)$  l'application linéaire à  $I : f \mapsto \int fgdP$  et en particulier :*

$$\|g\|_q = \sup\left\{\left|\int fgdP\right|; \|f\|_p \leq 1\right\}.$$

PREUVE : Par Hölder,  $fg \in L^1$  donc l'intégrale est définie et

$$\left|\int fgdP\right| \leq \|fg\|_1 \leq \|f\|_p \|g\|_q$$

d'où  $\|g\|_q$  est plus grand que le sup de l'énoncé. Mais, pour  $p \in ]1, \infty[$ , si on prend  $f = \bar{g}|g|^{q-2}/\|g\|_q^{q-1}$  on a  $|f|^p = |g|^{p(q-1)}/\|g\|_q^{p(q-1)} = |g|^q/\|g\|_q^q$  car  $p(q-1) = qp(1-1/q) = q$ , donc  $f \in L^p$  et  $\|f\|_p^p = \mathbf{E}(|f|^p) = \|g\|_q^q/\|g\|_q^q = 1$  d'où le sup est supérieur à  $|\int fgdP| = \int |g|^q dP/\|g\|_q^{q-1} = \|g\|_q$ . On déduit donc l'égalité énoncée. En conséquence on déduit que l'identification est bien injective (si la forme linéaire est nulle, le sup aussi, donc  $\|g\|_q = 0$  donc  $g = 0$  p.s. c'est à dire dans  $L^q$ .)



Si  $p = 1, q = \infty$ ,  $g$  est dans tout  $L^r$  et en prenant le  $f = \overline{g}|g|^{r-2}/\|g\|_r^{r-1}$  précédant qui est de norme  $L^1$  inférieure à 1, on voit que le sup est inférieur à  $\|g\|_r$  donc à la limite quand  $r \rightarrow \infty$  à savoir  $\|g\|_\infty$ .

Si  $p = \infty, q = 1$ , il suffit de prendre  $f = 1_{g \neq 0} \frac{\overline{g}}{|g|}$  de sorte que  $fg = |f|$  et la norme  $\|f\|_\infty \leq 1$ . ■

On a même le théorème suivant (on notera que  $p < \infty$  contrairement au cas de la proposition précédente) :

**Théorème 34.** Dans l'identification précédente,  $L^q(\Omega, P) = (L^p(\Omega, P))^*$ , pour  $p \in [1, \infty[$ ,  $q$  tel que  $1/p + 1/q = 1$ .

## 2 Définition et propriétés de l'espérance conditionnelle

En L2, vous avez vu la notion de probabilité conditionnelle  $P(A|B) = P(A \cap B)/P(B)$  de deux événements. Cela donne une probabilité  $P(\cdot|B)$  il est donc facile d'intégrer par rapport à cette mesure pour obtenir un nombre. On veut maintenant pouvoir considérer à la place d'un événement supposé connu, avoir une tribu d'événements. Ce sera l'objet de base pour pouvoir formaliser un processus avec notre connaissance qui augmente avec le temps (cf. exemple  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$  déjà rencontré pour la loi du 0 – 1.) Le prochain résultat veut donc définir une espérance "sachant une tribu" qui sera une v.a. mesurable par rapport à cette tribu et devra avoir les propriétés de l'espérance. On va s'appuyer sur la projection orthogonale dans  $L^2$  de la section précédente et l'étendre au cas des variables intégrables.

**Théorème 35.** Soit  $(\Omega, \mathcal{T}, P)$  un espace probabilisé et  $\mathcal{A} \subset \mathcal{T}$  une sous-tribu. Soit  $X \in L^1(\Omega, \mathcal{T}, P)$  une v.a. réelle, il existe une unique (p.s.) v.a. réelle  $\mathcal{A}$  mesurable notée  $\mathbf{E}(X|\mathcal{A}) \in L^1(\Omega, \mathcal{A}, P) \subset L^1(\Omega, \mathcal{T}, P)$  vérifiant la propriété caractéristique :

$$\forall f \in L^\infty(\Omega, \mathcal{A}, P), \mathbf{E}(fX) = \mathbf{E}(f\mathbf{E}(X|\mathcal{A})) \quad (PC).$$

Remarque 2. (PC) est équivalente à :

$$\forall A \in \mathcal{A}, \mathbf{E}(1_A X) = \mathbf{E}(1_A \mathbf{E}(X|\mathcal{A})) \quad (PC')$$

comme on va voir dans la preuve. (PC') suffit à voir l'unicité et (PC) implique (PC') car  $1_A \in L^\infty(\Omega, \mathcal{A}, P)$ . L'existence de la variable avec la propriété la plus forte implique donc que la v.a. vérifiant (PC') qui est unique doit aussi vérifier (PC) (sinon on raisonne comme pour la construction de l'intégrale pour aller au delà des fonctions étagées.

PREUVE :

**Unicité** Soit  $g, h$   $\mathcal{A}$ -mesurables vérifiant (PC'). On pose pour  $\epsilon > 0$ ,  $A = A_\epsilon = \{\omega, (g\omega) - h(\omega) \geq \epsilon\} = (g - h)^{-1}([\epsilon, \infty[)$  qui est donc  $\mathcal{A}$ -mesurable comme  $h - g$ , donc par (PC')  $E(1_A h) = E(1_A g) = E(1_A g)$  soit en utilisant  $1_A(g - h) \geq \epsilon 1_A$  :

$$0 = E(1_A(g - h)) \geq \epsilon P(A)$$

donc  $P(A) = 0$  pour tout  $\epsilon$  donc par union croissante

$$P(g > h) = \lim_{\epsilon \rightarrow 0} P(A_\epsilon) = 0.$$

Par symétrie  $P(h > g) = 0$  et donc  $g = hp.s.$

**Existence** Si  $X \in L^2(\Omega, \mathcal{T}, P)$ , on pose  $\mathbf{E}(X|\mathcal{A}) = P_{L^2(\Omega, \mathcal{A}, P)}(X)$  la projection sur le sous-espace fermé  $L^2(\Omega, \mathcal{A}, P)$ . Par définition elle vérifie pour (PC) avec  $f \in L^2(\Omega, \mathcal{A}, P)$  donc a fortiori avec  $f \in L^\infty(\Omega, \mathcal{A}, P)$ .

Montrons que pour  $X \in L^2(\Omega, \mathcal{T}, P)$ ,

$$|\mathbf{E}(X|\mathcal{A})| \leq \mathbf{E}(|X||\mathcal{A}), p.s.$$

Vu que  $|\mathbf{E}(X|\mathcal{A})| = \max(\mathbf{E}(X|\mathcal{A}), \mathbf{E}(-X|\mathcal{A}))$ , il suffit de voir et par linéarité il suffit de voir  $\mathbf{E}(|X| - X|\mathcal{A}) \geq 0$  p.s. Mais si  $A = \{\omega : \mathbf{E}(|X| - X|\mathcal{A}) < -\epsilon\}$ ,  $\epsilon > 0$  qui est  $\mathcal{A}$  mesurable,  $\mathbf{E}(1_A(|X| - X)) = \mathbf{E}(1_A \mathbf{E}(|X| - X|\mathcal{A}))$  la première espérance est positive comme espérance de fonctions positives la seconde est négative par définition de  $A$ , donc elle est nulle. Or  $\mathbf{E}(1_A \mathbf{E}(|X| - X|\mathcal{A})) \leq -\epsilon P(A)$  donc  $P(A) = 0$  En prenant  $\epsilon = 1/n$  aussi par union dénombrable  $P(\mathbf{E}(|X| - X|\mathcal{A}) < 0) = 0$  ce qu'il fallait démontrer.

En intégrant l'inégalité que l'on vient de montrer, on obtient :

$$\|\mathbf{E}(X|\mathcal{A})\|_1 \leq \mathbf{E}(\mathbf{E}(|X||\mathcal{A})) = \mathbf{E}(|X|) = \|X\|_1$$

par la propriété caractéristique. On étend donc  $\mathbf{E}(\cdot|\mathcal{A}) : L^1(\Omega, \mathcal{T}, P) \rightarrow L^1(\Omega, \mathcal{A}, P)$  par continuité et par densité de  $L^2(\Omega, \mathcal{T}, P) \subset L^1(\Omega, \mathcal{T}, P)$ . (On verra plus loin une construction dans le cas positif qui évite cette prolongation par continuité.). (PC) est étendu par densité vu que  $L^\infty \subset (L^1)^*$  (et on a vu qu'on avait même égalité). ■

**Exemple 17.** Si  $\mathcal{A} = \{\Omega, \emptyset\}$  est la tribu triviale :

$$\mathbf{E}(X|\mathcal{A}) = \mathbf{E}(X).$$

Plus généralement on a le même résultat si  $\sigma(X)$  et  $\mathcal{A}$  sont indépendantes (quelque soit  $\mathcal{A}$ ).

**Exemple 18.** Si  $A$  est un événement,  $\mathcal{A} = \sigma(A) = \{\Omega, A, A^c, \emptyset\}$  tq  $P(A) \neq 0, P(A^c) \neq 0$  et si  $B \in \mathcal{T}$ ,

$$\mathbf{E}(1_B|\mathcal{A}) = P(B|A)1_A + P(B|A^c)1_{A^c}.$$

Plus généralement, on a :

$$\mathbf{E}(X|\mathcal{A}) = \frac{\mathbf{E}(X1_A)}{P(A)}1_A + \frac{\mathbf{E}(X1_{A^c})}{P(A^c)}1_{A^c}.$$

On commence par donner les propriétés fondamentales similaires à celles de l'intégrale et on donne ensuite les exemples fondamentaux (cas discrets, continus et gaussiens).

Dans les propriétés suivantes, on sous-entend les égalités/inégalités p.s. puisque les égalités sont entre objets de  $L^1$ , définis comme classes d'équivalence p.s.

**Théorème 36.** Soit  $(\Omega, \mathcal{T}, P)$  un espace probabilisé,  $\mathcal{A} \subset \mathcal{T}$  une sous-tribu alors et  $X, Y \in L^1(\Omega, \mathcal{T}, P)$

1. (Linéarité)  $\mathbf{E}(aX + bY + c|\mathcal{A}) = a\mathbf{E}(X|\mathcal{A}) + b\mathbf{E}(Y|\mathcal{A}) + c$ .
2. (monotonie)  $X \leq Y \Rightarrow \mathbf{E}[X|\mathcal{A}] \leq \mathbf{E}[Y|\mathcal{A}]$ . En particulier,  $X \geq 0 \Rightarrow \mathbf{E}[X|\mathcal{A}] \geq 0$ .
3. (Jensen) Si  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  est convexe  $X$  v.a. réelle, et  $\phi(X) \in L^1(\Omega, \mathcal{T}, P)$  alors

$$\phi(\mathbf{E}(X|\mathcal{A})) \leq \mathbf{E}(\phi(X)|\mathcal{A}).$$

4. (Cas  $L^p$ ) Si  $p \in [1, \infty]$ , si  $X \in L^p(\Omega, \mathcal{T}, P)$ , alors  $\mathbf{E}(X|\mathcal{A}) \in L^p(\Omega, \mathcal{A}, P)$  et

$$\|\mathbf{E}(X|\mathcal{A})\|_p \leq \|X\|_p.$$

5. (Cas  $L^2$ ) Si  $X \in L^2$ ,  $E(X|\mathcal{A}) = P_{L^2(\Omega, \mathcal{A}, P)}(X)$ .

6. (Cas positif) Si  $Z \geq 0$ ,

$$\mathbf{E}(Z|\mathcal{A}) = \lim_{n \rightarrow \infty} \mathbf{E}(\min(Z, n)|\mathcal{A})$$

limite croissante à valeur  $[0, \infty]$  est l'unique v.a. positive  $\mathcal{A}$  mesurable vérifiant (PC'), ou par

$$\forall f \geq 0, \sigma(f) \subset \mathcal{A}, \mathbf{E}(fX) = \mathbf{E}(f\mathbf{E}(X|\mathcal{A})) \quad (PC''').$$

7. (Convergence monotone TCM) Si  $Z_n \geq 0$ ,  $Z_n$  suite de v.a.r croissante et tend simplement vers  $Z$  alors  $\mathbf{E}(Z_n|\mathcal{A}) \rightarrow \mathbf{E}(Z|\mathcal{A})$  p.s. en croissant.

8. (Lemme de Fatou) Si  $Z_n \geq 0$   $\mathbf{E}(\liminf Z_n|\mathcal{A}) \leq \liminf \mathbf{E}(Z_n|\mathcal{A})$  p.s.

9. (Convergence dominée TCD) Si  $|Z_n| \leq Y$ ,  $\mathbf{E}[Y] < \infty$  et  $Z_n \rightarrow Z$  p.s. alors

$$\mathbf{E}(Z_n|\mathcal{A}) \rightarrow \mathbf{E}(Z|\mathcal{A}) \text{ p.s et dans } L^1.$$

10. (Modularité) Si  $Y$  est  $\mathcal{A}$  mesurable et soit  $Y, X \geq 0$  soit  $X, XY \in L^1$  alors  $E(XY|\mathcal{A}) = YE(X|\mathcal{A})$ .

11. (Conditionnement successif) Si  $\mathcal{B} \subset \mathcal{A}$ ,  $\mathbf{E}(\mathbf{E}(X|\mathcal{A})|\mathcal{B}) = \mathbf{E}(X|\mathcal{B})$ .

12. (Indépendance) Si  $X, T : \Omega \rightarrow \mathbb{R}^m$  indépendantes de  $Z : \Omega \rightarrow \mathbb{R}^d$  alors  $\mathbf{E}(X|\sigma(Z, T)) = \mathbf{E}(X|T)$ .

PREUVE : 2. et 5. on été vérifiées durant la construction. Toutes les égalités vont venir de (PC).

- Pour 1. soit  $f \in L^\infty(\Omega, \mathcal{A}, P)$ , par linéarité de l'intégrale et application de (PC) à chaque membre !

$$\mathbf{E}(f(a\mathbf{E}(X|\mathcal{A}) + b\mathbf{E}(Y|\mathcal{A}) + c)) = \mathbf{E}(fa\mathbf{E}(X|\mathcal{A})) + \mathbf{E}(fb\mathbf{E}(Y|\mathcal{A})) + \mathbf{E}(fc) = \mathbf{E}(faX + fbY + fc)$$

ce qui donne le résultat par (PC).

-De même pour 11. si  $f \in L^\infty(\Omega, \mathcal{B}, P) \subset L^\infty(\Omega, \mathcal{A}, P)$ , deux applications successives de (PC) pour les deux espérances conditionnelles donnent :

$$\mathbf{E}[f\mathbf{E}(\mathbf{E}(X|\mathcal{A})|\mathcal{B})] = \mathbf{E}[f\mathbf{E}(X|\mathcal{A})] = \mathbf{E}[fX]$$

ce qui donne l'égalité par (PC).

-Pour montrer 3., on utilise que le graphe de  $\phi$  convexe est l'intersection des demi-plans de pentes rationnelles dessous lui. Si  $E_\phi = \{(a, b) \in \mathbb{Q}^2 : \forall x \in \mathbb{R} \phi(x) \geq ax + b\}$ , on a donc

$$\phi(x) = \sup_{(a,b) \in E_\phi} (ax + b).$$

Par 2.  $\mathbf{E}(\phi(X)|\mathcal{A}) \geq \mathbf{E}(aX + b|\mathcal{A})$  p.s pour tout  $(a, b) \in E_\phi$  c'est à dire sur  $\Omega_{a,b}$  avec  $P(\Omega_{a,b}) = 1$  Donc si  $\Omega' = \bigcap_{(a,b) \in E_\phi} \Omega_{a,b}$  on a  $P(\Omega') = 1$  et

$$\forall \omega \in \Omega', \mathbf{E}(\phi(X)|\mathcal{A}) \geq \sup_{(a,b) \in E_\phi} \mathbf{E}(aX + b|\mathcal{A}) = \phi(\mathbf{E}(aX + b|\mathcal{A})),$$

donc l'identité est vrai p.s.

- Pour 4. comme  $\phi(x) = |x|^p$ ,  $p \in ]1, \infty[$  est convexe, on applique 3. et on prend l'espérance

$$\mathbf{E}(|\mathbf{E}(X|\mathcal{A})|^p) \leq \mathbf{E}(\mathbf{E}(|X|^p|\mathcal{A})) = \mathbf{E}(|X|^p) = \|X\|_p^p < \infty$$

l'égalité venant de (PC) avec  $f = 1$ . Cela donne l'énoncé et l'égalité.

- Pour 6., (PC) usuelle appliquée à  $\min(f, n)$  et le TCM 2 fois donne (PC''')

$$\mathbf{E}(fX) = \lim_{n \rightarrow \infty} \mathbf{E}(\min(f, n) \min(X, n)) = \lim_{n \rightarrow \infty} \mathbf{E}(\min(f, n) \mathbf{E}(\min(X, n) | \mathcal{A})) = \mathbf{E}(f \mathbf{E}(X | \mathcal{A})).$$

L'unicité utilise le même  $A_\epsilon = \{g - h \geq \epsilon\}$  que dans le théorème, pour  $g, h$  vérifiant (PC')  $E(h1_{A_\epsilon} 1_{h < n}) = E(g1_{A_\epsilon} 1_{h < n}) \geq E(h1_{A_\epsilon} 1_{h < n}) + \epsilon P(A_\epsilon \cap \{h < n\})$  donc comme les deux termes sont finis  $\epsilon P(A_\epsilon \cap \{h < n\}) = 0$  en faisant tendre  $n \rightarrow \infty$ ,  $P(A_\epsilon) = 0$  et on conclut comme avant.

- Pour 7.  $E(Z_n | \mathcal{A})$  tend p.s. vers  $Y$  par monotonie il suffit de voir que  $Y$  vérifie (PC'') par TCM des intégrales. Pour  $f$  positive, (c'est crucial pour garder monotonie),  $\mathcal{A}$  mesurable,  $\mathbf{E}(fY) = \lim_{n \rightarrow \infty} \mathbf{E}(fE(Z_n | \mathcal{A})) = \lim_{n \rightarrow \infty} \mathbf{E}(fZ_n) = \mathbf{E}(fZ)$ .

-8.9 se démontrent alors comme en théorie de la mesure.  $\inf_{k \geq n} Z_k \rightarrow \liminf Z_n$  en croissant donc on peut utiliser 7. et il suffit de remarquer que  $\mathbf{E}[\inf_{k \geq n} Z_k | \mathcal{A}] \leq \inf_{k \geq n} \mathbf{E}[Z_k | \mathcal{A}]$  en utilisant 2. et en utilisant la dénombrabilité pour avoir le résultat p.s.

Donc

$$\mathbf{E}[\liminf Z_n | \mathcal{A}] = \lim_{k \geq n} \mathbf{E}[\inf_{k \geq n} Z_k | \mathcal{A}] \leq \liminf_{k \geq n} \mathbf{E}[Z_k | \mathcal{A}] = \liminf \mathbf{E}[Z_n | \mathcal{A}].$$

-Pour 9. vu  $Z_n - Y, Z_n + Y \geq 0$  on applique Fatou

$$\mathbf{E}(Z - X | \mathcal{A}) = \mathbf{E}(\liminf (Z - X_n) | \mathcal{A}) \leq \mathbf{E}(Z | \mathcal{A}) - \limsup \mathbf{E}(X_n | \mathcal{A})$$

$$\mathbf{E}(Z + X | \mathcal{A}) = \mathbf{E}(\liminf (Z + X_n) | \mathcal{A}) \leq \mathbf{E}(Z | \mathcal{A}) + \liminf \mathbf{E}(X_n | \mathcal{A})$$

donc en soustrayant le terme en  $Z$ ,  $\mathbf{E}(X | \mathcal{A}) \leq \liminf \mathbf{E}(X_n | \mathcal{A}) \leq \limsup \mathbf{E}(X_n | \mathcal{A}) \leq \mathbf{E}(X | \mathcal{A})$  p.s. ce qui donne l'égalité p.s. et la convergence p.s. la convergence  $L^1$  vient du TCD vu que  $\mathbf{E}(Z | \mathcal{A})$  sert de domination.

- Pour 10, le cas positif est évident par (PC'') et le cas  $L^1$  s'en déduit en décomposant en parties positives et négatives.

- Pour 12  $\mathbf{E}(X | T)$  est  $\sigma(Z, T)$  mesurable, et par le lemme de Doob-Dynkin une fonction  $Y \in L^\infty(\Omega, \sigma(Z, T))$  s'écrit  $Y = h(Z, T)$  avec  $h$  borélienne bornée. Or de même  $\mathbf{E}(X | T) = g(T)$  avec  $g \in L^1(\mathbb{R}^m, P_T)$  donc par transfert, indépendance, Fubini et (PC) :

$$\begin{aligned} \mathbf{E}(\mathbf{E}(X | T) h(Z, T)) &= \int dP_T(t) \int dP_Z(z) g(t) h(z, t) \\ &= \int dP_Z(z) \mathbf{E}(\mathbf{E}(X | T) h(z, T)) = \int dP_Z(z) \mathbf{E}(X h(z, T)) \\ &= \int dP_{(X, T)}(x, t) \int dP_Z(z) x h(z, t) = \mathbf{E}(X h(Z, T)) \end{aligned}$$

ce qui conclut par (PC). ■

*Remarque 3.* Pour des v.a.  $X_1, \dots, X_n$   $Y$  positive ou intégrable, on note

$$\mathbf{E}(Y | X_1, \dots, X_n) := \mathbf{E}(Y | \sigma(X_1, \dots, X_n)).$$

## 2.1 Cas discret

Il y a deux façons de le décrire avec les systèmes presque complets d'événements (definition ??) et les variables aléatoires.

**Proposition 37.** 1. Soit  $(A_n)_{n \geq 1}$  un système presque complet d'événements et  $\mathcal{A} = \sigma(A_n, n \geq 1)$  alors pour  $X \in L^1(\Omega, \mathcal{T}, P)$  :

$$\mathbf{E}(X|\mathcal{A}) = \sum_{n: P(A_n) \neq 0} \frac{\mathbf{E}(X1_{A_n})}{P(A_n)} 1_{A_n}.$$

2. Si  $Y$  v.a. à valeur dans un ensemble dénombrable  $E$  et  $X \in L^1(\Omega, \mathcal{T}, P)$ , soit  $\phi(y) = \frac{\mathbf{E}(X1_{Y=y})}{P(Y=y)}$  si  $P(Y=y) \neq 0$  et  $\phi(y) = 0$  sinon, alors

$$\mathbf{E}(X|Y) = \phi(Y).$$

## 2.2 Cas Continue

**Théorème 38.** Soit  $(X_1, \dots, X_n, Y_1, \dots, Y_m)$  un vecteur aléatoire de densité  $p$  sur  $\mathbb{R}^{n+m}$ . Soit  $h : \mathbb{R}^n \rightarrow [0, \infty[$  mesurable positive et  $g$  définie par

$$g(y) = \frac{1}{q(y)} \int_{\mathbb{R}^n} h(x)p(x, y)dx$$

si  $q(y) := \int_{\mathbb{R}^n} p(x, y)dx$  la densité marginale est non nulle et  $g(y) = 0$  (ou n'importe quelle valeur) si  $q(y) = 0$  alors :

$$\mathbf{E}[h(X)|Y] = g(Y).$$

PREUVE : Il suffit de vérifier la propriété caractéristique (PC<sup>m</sup>) du cas positif en appliquant le transfert. Par le lemme de Doob-Dynkin (proposition 4), prendre  $F$   $\sigma(Y)$  mesurable positive revient à prendre  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  borélienne positive et  $F = f(Y)$ . Par transfert, pour appliquer (PC) on calcule donc :

$$\begin{aligned} \mathbf{E}(f(Y)g(Y)) &= \int_{\mathbb{R}^m} f(y)g(y)q(y)dy \\ &= \int_{q(y) \neq 0} dy f(y)q(y) \frac{1}{q(y)} \int_{\mathbb{R}^n} h(x)p(x, y)dx \\ &= \int_{\mathbb{R}^m} dy f(y) \int_{\mathbb{R}^n} h(x)p(x, y)dx \\ &= \mathbf{E}(f(Y)h(X)). \end{aligned}$$

car  $\int_{\mathbb{R}^n} h(x)p(x, y)dx = 0$  si  $q(y) = 0$  (qui implique  $p(x, y) = 0$  pour presque tout  $x$ ) ce qui a permis de ramener l'intégrale à une intégrale sur  $\mathbb{R}^m$  qui est de nouveau la valeur donnée par transfert. Par PC cas positif cela conclut à l'identité. ■

*Remarque 4.* La donnée de  $\mathbf{E}(h(X)|Y)$  pour tout  $h$  positive mesurable est souvent appelée donnée de la loi conditionnelle de  $X$  sachant  $Y$ .

## 2.3 Cas Indépendant

**Proposition 39.** Soient  $X : \Omega \rightarrow \mathbb{R}^n, Y : \Omega \rightarrow \mathbb{R}^m$  des v.a. indépendantes et soit  $\Phi : \mathbb{R}^{n+m} \rightarrow \mathbb{R}_+$  mesurable et alors  $\varphi(x) = \mathbf{E}[\Phi(x, Y)] \in \mathbb{R}_+$ , est mesurable et

$$\mathbf{E}[\Phi(X, Y)|X] = \varphi(X) \text{ p.s.}$$

PREUVE : La mesurabilité vient de Fubini-Tonelli. On est dans le cas positif où on utilise (PC<sup>m</sup>) pour l'espérance conditionnelle positive. Soit  $Z \geq 0$   $\sigma(X)$ -mesurable, par le lemme de Doob-Dynkin 4,  $Z = h(X)$  avec  $h : \mathbb{R}^n \rightarrow \mathbb{R}_+$ .

Or par transfert et indépendance

$$\begin{aligned} \mathbf{E}[\Phi(X, Y)Z] &= \mathbf{E}[\Phi(X, Y)h(X)] = \int dP_X(x) \int dP_Y(y) \Phi(x, y)h(x) \\ &= \int dP_X(x) \mathbf{E}(\Phi(x, Y))h(x) = \int dP_X(x) \varphi(x)h(x) = \mathbf{E}(\varphi(X)Z) \end{aligned}$$

ce qui établit par (PC<sup>m</sup>)  $\mathbf{E}[\Phi(X, Y)|X] = \varphi(X)$ . ■

### 3 Lois conditionnelles

On va maintenant étendre le principe du conditionnement des espérances aux lois (en s'appuyant sur le lemme de Doob-Dynkin : Proposition 4). On sait que si on a  $X = (X_1, \dots, X_m)$ ,  $Y = (Y_1, \dots, Y_n)$  alors pour  $h$  mesurable borné,  $\mathbf{E}(h(X)|\sigma(Y))$  est une fonction  $\sigma(Y)$ -mesurable, donc une fonction  $g(Y)$  pour  $g$  mesurable bornée. On veut décrire  $g$  en terme d'une intégrale de  $h$  par rapport à la loi conditionnelle de  $X$  sachant  $Y$ .

**Définition 17.** On appelle noyau de transition, toute fonction  $K : \mathbb{R}^n \times \mathcal{B}(\mathbb{R}^m) \rightarrow [0, 1]$  telle que

1. pour tout  $B \in \mathcal{B}(\mathbb{R}^m)$ ,  $y \mapsto K(y, B) = K^y(B)$  est mesurable,
2. pour tout  $y \in \mathbb{R}^n$ ,  $B \mapsto K(y, B) = K^y(B)$  est une mesure de probabilité.

**Théorème 40** (admis). Soit  $(X, Y)$  un vecteur aléatoire dans  $\mathbb{R}^m \times \mathbb{R}^n$ , de loi  $P$  considérée sur la tribu complétée  $\mathcal{T}$  (engendré par les boréliens et les ensembles contenus dans un borélien de  $P$ -mesure nulle). Il existe un noyau de transition  $K$  telle que pour toute fonction borélienne bornée  $h$  :

$$\mathbf{E}(h(X)|\sigma(Y)) = \int h dK^Y \text{ p.s.}$$

La mesure  $K^y(dx)$  est appelée loi conditionnelle de  $X$  sachant  $Y = y$  et notée  $\mathcal{L}(X|Y = y)$ .

(Une idée de preuve est donné dans le Barbes-Ledoux et la preuve complète dans *Real Analysis and Probability* de R. Dudley Theorems 10.2.2 and 10.2.5)

**Exemple 19.** La proposition 37 dit que si  $(X, Y)$  est discrète alors  $\mathcal{L}(X|Y = y) = \sum_{x \in X(\Omega)} P(X = x|Y = y)\delta_x$  si  $P(Y = y) \neq 0$  et  $\delta_0$  sinon.

**Exemple 20.** Le Théorème 38 dit que si  $(X, Y)$  a une densité,  $p$  alors  $\mathcal{L}(X|Y = y) = \frac{p(x, y)dx}{\int_{\mathbb{R}^m} p(x, y)dx}$  quand l'intégrale est non nulle et disons  $\delta_0$  sinon.

**Exemple 21.** La proposition 39 dit que si  $(X, Y)$  sont indépendante alors  $\mathcal{L}(X|Y = y) = \mathcal{L}(X)$ .

### 4 Espérances conditionnelles dans le Cas Gaussien

Pour les vecteurs gaussiens, on a deux simplifications, d'abord le calcul de l'espérance conditionnelle est ramené à un problème d'algèbre linéaire :

**Théorème 41.** Soit  $(X, Y_1, \dots, Y_m)$  un vecteur gaussien centré (i.e.  $\mathbf{E}(X) = \mathbf{E}(Y_j) = 0$ ) et  $Y = (Y_1, \dots, Y_m)$ , alors

$$\mathbf{E}(X|Y) = \sum_{j=1}^m \lambda_j Y_j$$

est la projection orthogonale sur  $\text{Vect}(Y_1, \dots, Y_m) \subset L^2(\Omega, \sigma(Y), P)$ .

Cette projection est donc sur un espace beaucoup plus petit que celle a priori pour tous les vecteurs  $L^2$  à savoir sur  $L^2(\Omega, \sigma(Y), P)$ .

PREUVE :

Soit  $\lambda_j$  avec  $\sum_{j=1}^m \lambda_j Y_j$  la projection orthogonale sur  $\text{Vect}(Y_1, \dots, Y_m)$ . Voyons la propriété caractéristique pour la projection orthogonale sur  $L^2(\Omega, \sigma(Y), P)$ , qu'on sait être  $\mathbf{E}(X|Y)$ . On veut donc voir que  $X - \sum_{j=1}^m \lambda_j Y_j$  est orthogonale à  $L^2(\Omega, \sigma(Y), P)$ .

Or  $Z = X - \sum_{j=1}^m \lambda_j Y_j$  orthogonale à  $\text{Vect}(Y_1, \dots, Y_m)$  donc  $\text{Cov}(Z, Y_j) = E(ZY_j) = 0$  (car les vecteurs sont centrés). De plus  $(Z, Y_1, \dots, Y_m)$  est un vecteur gaussien comme application linéaire de  $(X, Y_1, \dots, Y_m)$  donc par le critère d'indépendance 26,  $\sigma(Z)$  et  $\sigma(Y_1, \dots, Y_m)$  sont indépendants, donc pour tout  $f \in L^2(\Omega, \sigma(Y), P)$ ,  $E(Zf) = E(Z)E(f) = 0$ . Donc  $Z$  est bien orthogonal à tout  $L^2(\Omega, \sigma(Y), P)$ . Donc  $Z = \mathbf{E}(X|Y)$ . ■

**Exemple 22.** Soit  $(X_1, X_2, X_3)$  un vecteur gaussien de loi  $\mathcal{N}(0, \Gamma)$  et  $\Gamma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix}$ . Montrer que  $\mathbf{E}(X_2|X_1, X_3) = \frac{X_1 + X_3}{2}$ .

Le résultat suivant donne en fait la "loi conditionnelle".

**Théorème 42.** Si  $(X_1, \dots, X_n, Y_1, \dots, Y_m)$  un vecteur gaussien centré,  $Y = (Y_1, \dots, Y_m)$ . Soit  $h : \mathbb{R}^n \rightarrow [0, \infty[$  mesurable positive. Soit  $C$  la matrice de covariance conditionnelle

$$C_{ij} = \mathbf{E}[(X_i - \mathbf{E}(X_i|Y))(X_j - \mathbf{E}(X_j|Y))]$$

et soit  $g : \mathbb{R}^n \rightarrow [0, \infty[$  définie par

$$g(m) = \mathbf{E}(h(Z(m))),$$

avec  $Z(m) \sim \mathcal{N}(m, C)$ . Alors

$$\mathbf{E}(h(X)|Y) = g(\mathbf{E}(X|Y)).$$

**Remarque 5.** Le résultat s'interprète en terme de loi conditionnelle et disant que  $\mathcal{L}(X|Y) = \mathcal{N}(\mathbf{E}(X|Y), C)$ . On remarque aussi que

$$C = \text{Cov}(X, X) - \text{Cov}(X, Y)\text{Cov}(Y, Y)^{-1}\text{Cov}(Y, X).$$

En effet, on a d'abord :

$$\mathbf{E}(X|Y) = \text{Cov}(X, Y)\text{Cov}(Y, Y)^{-1}Y$$

car il suffit qu'il est la bonne covariance avec  $Y$  et  $\text{Cov}(\text{Cov}(X, Y)\text{Cov}(Y, Y)^{-1}Y, Y) = \text{Cov}(X, Y)\text{Cov}(Y, Y)^{-1}\text{Cov}(X, Y)$  comme voulu. De plus on obtient donc  $\mathbf{E}[(X_i - \mathbf{E}(X_i|Y))(X_j - \mathbf{E}(X_j|Y))] = E(X_i X_j) - \mathbf{E}[\mathbf{E}(X_i|Y)\mathbf{E}(X_j|Y)] = \text{Cov}(X_i, X_j) - \mathbf{E}[\text{Cov}(X, Y)\text{Cov}(Y, Y)^{-1}Y]_i[\text{Cov}(X, Y)\text{Cov}(Y, Y)^{-1}Y]_j = \text{Cov}(X_i, X_j) - \text{Cov}(X, Y)\text{Cov}(Y, Y)^{-1}\text{Cov}(Y, Y)\text{Cov}(Y, Y)^{-1T}\text{Cov}(X, Y)^T]_{ij}$ .

**Exemple 23.** Dans l'exemple précédent,  $Var(X_2 - \mathbf{E}(X_2|X_1, X_3)) = (-1/2, 1, -1/2)\Gamma(-1/2, 1, -1/2)^T = (-1/2, 1, -1/2)(0, 3/2, 0)^T = 3/2$ . Donc

$$E(X_2^4|X_1, X_3) = 27/4 + \left(\frac{X_1 + X_3}{2}\right)^4 + 9\left(\frac{X_1 + X_3}{2}\right)^2.$$

PREUVE : On note  $g_h$  le  $g$  de l'énoncé. Par la caractérisation du cas positif et le lemme de Doob-Dynkin, il suffit de montrer pour  $f : \mathcal{T}^m \rightarrow [0, \infty[$  mesurable :

$$\mathbf{E}(f(Y)g_h(\mathbf{E}(X|Y))) = \mathbf{E}(f(Y)h(X))$$

Si  $\mathbf{E}(f(Y)) = 0$   $f(Y) = 0$  p.s. donc les deux cotés sont nuls. On suppose le contraire, soit

$$\mu(h) = \frac{\mathbf{E}(f(Y)h(X))}{\mathbf{E}(f(Y))}, \nu(h) = \frac{\mathbf{E}(f(Y)g_h(\mathbf{E}(X|Y)))}{\mathbf{E}(f(Y))}.$$

$\mu(1) = 1$  et  $g_1 = 1$  donc  $\nu(1) = 1$ .  $\mu(B) = \mu(1_B), \nu(B) = \nu(1_B)$  définissent des mesures de probabilités sur  $\mathcal{B}(\mathbb{R}^n)$  et il est facile de voir que  $\mu(h) = \int h d\mu, \nu(h) = \int h d\nu$ .

Donc pour voir  $\mu(h) = \nu(h)$  il suffit de voir  $\mu = \nu$  et par le théorème d'injectivité de la transformée de Fourier il suffit de voir pour tout  $t \in \mathbb{R}^n$  :

$$\mathbf{E}(f(Y)E(e^{i\langle t, Z(\mathbf{E}(X|Y))\rangle})) = \mathbf{E}(f(Y)e^{i\langle t, X \rangle}).$$

Soit encore

$$E(e^{i\langle t, X \rangle}|Y) = E(e^{i\langle t, Z(\mathbf{E}(X|Y))\rangle}) = \exp(i\langle t, \mathbf{E}(X|Y) \rangle - \frac{\langle Ct, t \rangle}{2})$$

la dernière identité par la transformée de Fourier du vecteur gaussien  $Z(m)$ . Or  $(\langle t, X \rangle, Y_1, \dots, Y_m)$  est gaussien, et  $\langle t, \mathbf{E}(X|Y) \rangle = \mathbf{E}(\langle t, X \rangle|Y)$  par linéarité de l'espérance conditionnelle donc par la preuve du résultat précédent  $\langle t, X \rangle - \langle t, \mathbf{E}(X|Y) \rangle$  est indépendant de  $Y_1, \dots, Y_m$ , donc

$$E(e^{i\langle t, X \rangle}|Y) = E(e^{i\langle t, X \rangle - \langle t, \mathbf{E}(X|Y) \rangle}|Y)\exp(i\langle t, \mathbf{E}(X|Y) \rangle) = E(e^{i\langle t, X \rangle - \langle t, \mathbf{E}(X|Y) \rangle})\exp(i\langle t, \mathbf{E}(X|Y) \rangle),$$

par la modularité (Thm 36.10) et la relation à l'indépendance (ex 39). En utilisant la transformée de Fourier d'une variable gaussienne et  $\langle Ct, t \rangle = Var(\langle t, X \rangle - \langle t, \mathbf{E}(X|Y) \rangle)$ , on obtient le résultat d'égalité des transformées de Fourier qu'il fallait démontrer. ■



# Chapitre 4

## Chaînes de Markov

### 1 Définitions et premiers exemples

Toutes les variables seront dans un ensemble fini ou dénombrable  $E$ , appelé espace d'états.

#### 1.1 Matrices de transition

**Définition 18.** On appelle *matrice de transition sur  $E$*  une collection de nombres réels positifs  $P = (P_{i,j})_{i,j \in E}$  telle que :

$$\forall i \in E, \sum_{j \in E} P_{ij} = 1$$

*Remarque 6.* Une matrice de transition donne des probabilités  $\sum_{j \in E} P_{ij} \delta_j$  qui donnera la loi de probabilité sous  $E$  partant de  $i$ . Il sera pertinent de noter  $x \rightarrow y$  si  $P_{x,y} > 0$  pour dire qu'il y a une probabilité non nulle de passer de  $x$  à  $y$ .

Pour deux matrices de transition,  $P, Q$ , on note

$$(P.Q)_{ij} = \sum_{k \in E} P_{i,k} Q_{k,j}.$$

Cette somme de termes positifs est bien définie (a priori dans  $[0, +\infty]$  mais comme  $Q_{k,j} \leq 1$ , on a  $(P.Q)_{ij} \leq \sum_{k \in E} P_{i,k} \leq 1$  donc dans  $[0, 1]$ .)

De plus par Fubini-Tonelli, on a

$$\sum_{j \in E} (P.Q)_{ij} = \sum_{k \in E} \sum_{j \in E} P_{i,k} Q_{k,j} = \sum_{k \in E} P_{i,k} 1 = 1.$$

Donc en particulier tous les termes sont finis et  $PQ$  est une matrice de transition.

On notera  $P^n = (P \cdots P)$   $n$  fois. Sur un espace d'états finis, c'est le produit matriciel usuel.

#### 1.2 Chaînes de Markov

**Définition 19.** Soit  $P$  une matrice de transition sur  $E$ . Une suite de variables aléatoires à valeur dans  $E$ ,  $(X_n)_{n \geq 0}$ , est une *chaîne de Markov de matrice de transition  $P$*  si pour tout  $n$ , dès que  $P(X_n = i_n, \dots, X_0 = i_0) > 0$ , on a :

$$P(X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_0 = i_0) = P_{i_n, i_{n+1}}.$$

Remarque 7. On a donc

$$P(X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_0 = i_0) = P(X_{n+1} = i_{n+1} | X_n = i_n).$$

Autrement dit le futur  $X_{n+1}$  ne dépend du passé qu'à travers le présent  $X_n$ , ceci se réécrit :

$$\mathcal{L}((X_{m+n}, \dots, X_{n+1}, X_n) | (X_n, \dots, X_0)) = \mathcal{L}((X_{m+n}, \dots, X_{n+1}, X_n) | X_n).$$

Certains auteurs nomment chaîne de Markov toute suite de variables aléatoires satisfaisant cette relation. Ils appellent alors chaînes de Markov homogènes notre définition, ou en plus on a :

$$P(X_{n+1} = y | X_n = x) = P_{x,y} = P(X_1 = y | X_0 = x).$$

Toutes les chaînes de Markov de ce cours seront homogènes, la probabilité de transition ne dépendra pas de l'instant  $n$ . On a alors

$$\mathcal{L}((X_{m+n}, \dots, X_{n+1}, X_n) | (X_n, \dots, X_0) = (x_n, \dots, x_0)) = \mathcal{L}((X_{m+n}, \dots, X_n) | X_n = x_n) = \mathcal{L}((X_n, \dots, X_0) | X_0 = x_n). \quad (4.1)$$

On a la caractérisation évidente :

**Lemme 43.** Une suite  $(X_n)_{n \geq 0}$  est une chaîne de Markov de matrice de transition  $P$  si et seulement si pour tout  $i_0, \dots, i_n \in E$  :

$$P(X_n = i_n, \dots, X_1 = i_1, X_0 = i_0) = P(X_0 = i_0) P_{i_0, i_1} \dots P_{i_{n-1}, i_n}.$$

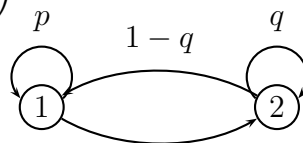
Ce lemme montre que la loi des suites finies  $(X_0, \dots, X_n)$  est déterminée et un théorème de Kolmogorov dit alors que la loi de la suite infinie  $(X_n)_{n \geq 0}$  est déterminée. Il y a donc au plus une loi de chaîne de Markov de matrice de transition  $P$  partant de  $X_0$  v.a. de loi  $P_{X_0} = \mu$  fixé. **On la note  $\text{CM}(\mathbf{E}, \mu, \mathbf{P})$ .**

### 1.3 Exemples

**Exemple 24. Variables i.i.d** Une suite de variables i.i.d de loi  $\mu = \sum_i p_i \delta_i$  est une chaîne de Markov de matrice de transition à colonnes constantes  $P_{i,j} = p_j$ .

**Exemple 25. Cas à 2 états** Toutes les chaînes à deux états ont pour  $p, q \in [0, 1]$  une matrice de transition de la forme :  $P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$ .

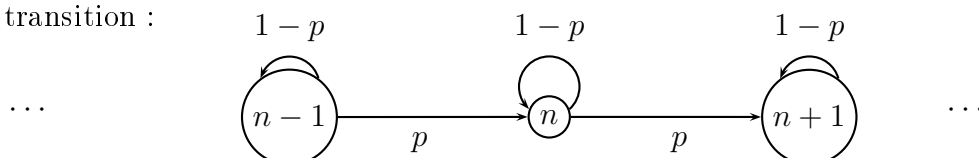
On la représentera par le graphe :



**Exemple 26. Un Cas à 3 états** de matrice de transition  $P = \begin{pmatrix} 1-p & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

**Exemple 27. Processus de Bernoulli**

C'est la chaîne de matrice de transition  $P_{i,i} = (1-p)$ ,  $P_{i,i+1} = p$  et  $p_{i,j} = 0$  sinon, pour  $p \in [0, 1]$  de graphe de transition :



C'est le modèle discret dont une certaine limite donnera le processus de Poisson du chapitre 5.

**Exemple 28. Suites récurrentes aléatoires** Soit  $X_0$  une v.a. à valeur  $E$ ,  $(U_n)_{n \geq 0}$  une suite i.i.d. dans un espace de proba  $F$ , et  $g : E \times F \rightarrow E$  mesurable, on peut définir  $X_{n+1} = g(X_n, U_{n+1})$ . C'est une chaîne de Markov de matrice de transition  $P_{i,j} = P(g(i, U_n) = j)$  qui ne dépend pas de  $n$  puisque  $(U_n)_{n \geq 0}$  est i.i.d.

**Exemple 29. Marches aléatoires** C'est le cas précédent avec  $E = F = E$  un groupe fini (par exemple  $E = \mathbb{Z}/n\mathbb{Z}$ ) ou dénombrable ( $E = \mathbb{Z}^d$ ) et  $g(x, y) = x + y$ , de sorte que  $X_n = X_0 + U_1 + \dots + U_n$ .

**Exemple 30. Marches aléatoires simples (MAS)** C'est le cas précédent avec  $E = \mathbb{Z}^d$  et  $U_i$  i.i.d de loi de sauts aux plus proches voisins partant de  $X_0 = 0$ ,  $P_{U_i} = \frac{1}{2d} \sum_{j=1}^d \delta_{\pm e_j}$  où  $e_j = (0, \dots, 0, 1, 0, \dots, 0)$  est le vecteur coordonné avec un 1 sur la  $j$ -ème coordonnée et la somme contient  $2d$  masse de Dirac. Alors la MAS est  $X_n = U_1 + \dots + U_n$ . C'est le modèle dont une certaine limite donnera le mouvement Brownien du chapitre 6.

## 1.4 Filtration et Temps d'arrêt

On fixe  $(\Omega, \mathcal{T}, P)$  un espace probabilisé.

**Définition 20.** Une *filtration*  $(\mathcal{F}_n)_{n \in \mathbb{N}}$  est une suite croissante de sous-tribus de  $\mathcal{T}$ .

Intuitivement  $\mathcal{F}_n$  est la tribu des événements connus à l'instant  $n$ . A partir de maintenant **toutes les tribus  $\mathcal{F}_n, \mathcal{T}$  sont supposées dénombrablement engendrées**, même si ce n'est pas strictement nécessaire pour les énoncés, seulement pour simplifier les preuves.

**Exemple 31.** Si  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires, on considère souvent la filtration engendrée par ce processus, définie par  $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ .

Réciproquement, on a une notion correspondant à un processus connu à l'instant  $n$  pour une filtration.

**Définition 21.** Une suite  $(X_n)_{n \in \mathbb{N}}$  est dite *adaptée* à la filtration  $(\mathcal{F}_n)_{n \in \mathbb{N}}$  si pour tout  $n$ ,  $X_n$  est  $\mathcal{F}_n$ -mesurable.

**Exemple 32.** Une suite  $(X_n)_{n \in \mathbb{N}}$  est toujours adaptée à la filtration engendrée par le processus  $(\mathcal{F}_n = \sigma(X_0, \dots, X_n))_{n \in \mathbb{N}}$  et elle est adaptée à  $\mathcal{F}_n$  si et seulement si  $\sigma(X_0, \dots, X_n) \subset \mathcal{F}_n$

On fixe à présent une filtration  $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ .

**Définition 22.** Soit  $(\mathcal{F}_n)_{n \in \mathbb{N}}$  une filtration. Une variable aléatoire  $T : (\Omega, \mathcal{T}) \rightarrow \mathbb{N} \cup \{\infty\}$  est un *temps d'arrêt* si pour tout  $n \in \mathbb{N}$ ,  $\{T \leq n\} \in \mathcal{F}_n$ . On note alors la tribu en  $T$

$$\mathcal{F}_T := \{A \in \mathcal{T}; \forall n, A \cap \{T \leq n\} \in \mathcal{F}_n\}.$$

Une fonction constante  $T = n$  est un temps d'arrêt et alors  $\mathcal{F}_T = \mathcal{F}_n$  est une notation consistante.

**Exemple 33 (Temps d'atteinte).** Soit  $A$  un ensemble mesurable dans  $\mathbb{R}^d$  et  $X_n : \Omega \rightarrow \mathbb{R}^d$  un processus adapté, alors

$$T = \inf\{n \in \mathbb{N}, X_n \in A\}$$

est un temps d'arrêt. En effet on a  $\{T \leq n\} = \{\exists m \leq n, X_m \in A\} \in \mathcal{F}_n$  car  $\sigma(X_0, \dots, X_n) \subset \mathcal{F}_n$ .

*Exercice 1.* Si  $T, S$  sont des temps d'arrêts alors leur minimum  $S \wedge T$  aussi.

*Exercice 2.* Si  $T$  temps d'arrêt p.s. fini pour  $\mathcal{F}_n$  et  $(X_n)$  adapté, montrer que  $X_T$  est  $\mathcal{F}_T$ -mesurable.

## 2 Construction, Propriété de Markov forte et Simulation

### 2.1 Construction et Simulation récurrente

On va commencer par construire une chaîne de Markov à partir d'une suite de variables indépendantes uniformes sur  $[0, 1]$  (ce qui donnera l'existence puisque qu'une telle suite peut être construite sur  $[0, 1]$  avec la mesure de Lebesgue). Cela donne aussi une méthode de simulation naïve des chaînes de Markov. A partir d'une construction canonique on obtiendra un renforcement de la propriété de Markov de la définition qui donnera une construction plus efficace en terme de temps d'attente et de sauts, comme au chapitre suivant pour le processus de Poisson.

**Proposition 44.** *Pour toute matrice de transition  $P$  sur  $E$  et mesure de probabilité  $\mu$  sur  $E$ , il existe des fonctions  $f : [0, 1] \rightarrow E, g : E \times [0, 1] \rightarrow E$  telles que si  $(U_n)_{n \geq 0}$  est une suite de variables aléatoires i.i.d. uniformes sur  $[0, 1]$  alors la suite  $X_0 = f(U_0)$  et  $X_{n+1} = g(X_n, U_{n+1})$  est une chaîne de Markov  $CM(E, \mu, P)$ .*

Comme  $f, g$  sont explicites dans la preuve, cela donne une méthode de simulation à partir d'une suite de nombres aléatoires.

*Démonstration.* On suppose sans perte de généralité  $E = \mathbb{N}$  et on pose :

$$f(x) = \min\{n \in \mathbb{N} : \mu(\{0\}) + \dots + \mu(\{n\}) \geq x\},$$

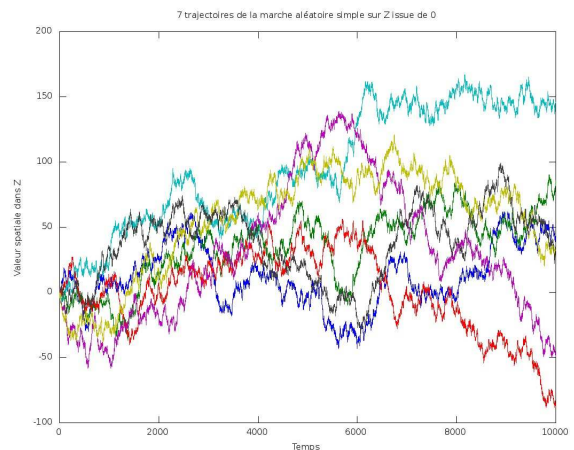
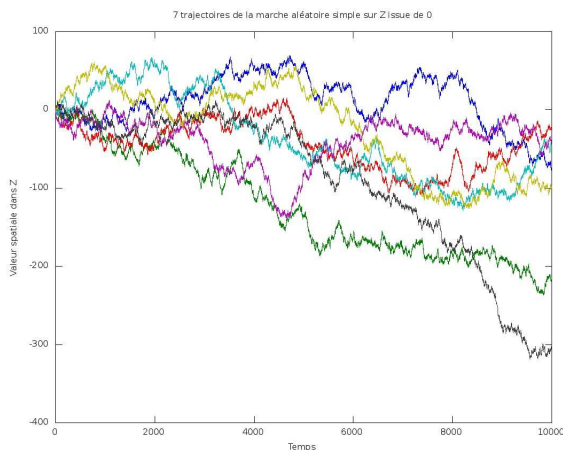
$$g(n, x) = \min\{m \in \mathbb{N} : P_{n,0} + \dots + P_{n,m} \geq x\}.$$

Le fait que  $f(U_0)$  ait la loi  $\mu$  est standard (méthode de simulation par l'inverse de la fonction de répartition). Rappelons le  $f(x) \leq n$  si et seulement si ( $n$  est dans l'ensemble définissant le min soit)  $\mu(\{0\}) + \dots + \mu(\{n\}) \geq x$ , donc  $f(U_0)$  est à valeur dans  $\mathbb{N}$  et  $P(f(U_0) \leq n) = P(U_0 \leq \mu(\{0\}) + \dots + \mu(\{n\})) = \mu(\{0\}) + \dots + \mu(\{n\})$  comme voulu pour la fonction de répartition de  $\mu$ . De même, la loi de  $g(m, U_{n+1})$  est de loi  $\sum_{n \geq 0} P_{m,n} \delta_n$ . Pour voir qu'on a bien une chaîne de Markov, on utilise le lemme 43 et l'indépendance des  $U_n$  :

$$\begin{aligned} P(X_n = i_n, \dots, X_1 = i_1, X_0 = i_0) &= P(f(U_0) = i_0, g(i_0, U_1) = i_1, \dots, g(i_{n-1}, U_n) = i_n) \\ &= P(f(U_0) = i_0)P(g(i_0, U_1) = i_1) \cdots P(g(i_{n-1}, U_n) = i_n) \\ &= P(X_0 = i_0)P_{i_0, i_1} \cdots P_{i_{n-1}, i_n}. \end{aligned}$$

□

Un exemple de simulation appliquée à la Marche aléatoire simple sur  $\mathbb{Z}$  :



## 2.2 Chaîne canonique

**Théorème 45** (Construction canonique des chaînes de Markov). *Soit  $\mu$  une probabilité sur  $E$  et  $P$  une matrice de transition.*

*On définit sur l'ensemble des suites à valeur dans  $E : \Omega = E^{\mathbb{N}}$  une filtration  $\mathcal{F}_n = \{A \times E^{\mathbb{N}} : A \subset E^n\}$  de la tribu canonique  $\mathcal{F} = \sigma(\cup \mathcal{F}_n)$ . Il existe sur  $\Omega$  une unique mesure  $\mathbf{P}_\mu$  tel que  $X_n(\omega) = \omega_n$  ( $n$ -ième coordonnée) soit une chaîne de Markov  $(X_n)_{n \geq 0}$  de matrice de transition  $P$  et de loi initiale  $\mu = P_{X_0}$ , soit  $\mathbf{P}_\mu \sim CM(E, \mu, P)$ .*

*Démonstration.* Par la proposition 44, on a  $Y : \Omega' \rightarrow E^{\mathbb{N}}$  la chaîne de Markov obtenue. Il est facile de voir quelle est  $\mathcal{F}$  mesurable, car ceci équivaut à  $X_n \circ Y = Y_n$  mesurable pour tout  $n$  (en effet vérifier cela pour  $n \leq N$  donne  $\mathcal{F}_N$  mesurable).  $P_\mu = P_Y$  est alors la mesure image, loi de  $Y$ . Pour l'unicité, il suffit que deux chaînes de Markov coïncident sur les  $\mathcal{F}_n$  par définition, et  $\cup_n \mathcal{F}_n$  soit stable par intersection finie, donc le lemme de classe monotone conclut à l'unicité (égalité de deux mesures ayant la même loi de chaîne de Markov).  $\square$

## 2.3 Propriété de Markov forte

La propriété de Markov permet de reformuler la propriété fondamentale des chaînes de Markov (4.2) en terme d'indépendance conditionnelle et de renouvellement de la structure de chaîne de Markov dans l'avenir. La version forte exprime la même propriété au niveau de tout temps d'arrêt. On note  $\theta_n : \Omega = E^{\mathbb{N}} \rightarrow \Omega$  l'opérateur de translation  $(\theta_n(\omega))_m = \omega_{n+m}$ .

**Lemme 46** (Propriété de Markov faible). *Soit  $\mu$  une loi de probabilité sur  $E$ ,  $P$  une matrice de transition et  $P_\mu \sim CM(E, \mu, P)$  la construction sur l'espace canonique  $\Omega$ ,  $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ . On note  $x \mapsto \mathbf{E}_x = E_{\delta_x}$  l'espérance pour la loi  $P_{\delta_x} \sim CM(E, \delta_x, P)$ . Pour tout  $F$  mesurable (positive ou bornée), on peut calculer l'espérance conditionnelle sur le passé du processus de la façon suivante :*

$$\mathbf{E}_\mu(F \circ \theta_n | \mathcal{F}_n) = \mathbf{E}_{X_n}(F) \quad (= \mathbf{E}.(F) \circ X_n). \quad (4.2)$$

*Autrement dit  $(X_{n+m})_{m \geq 0}$  est de loi  $CM(E, P_{X_n}, P)$  et la loi conditionnelle  $\mathcal{L}((X_{n+m})_{m \geq 0} | (X_0, \dots, X_n)) = P_{X_n}$ . En particulier, conditionnellement à l'événement  $\{X_n = x\}$  cette chaîne de Markov  $(X_{n+m})_{m \geq 0} \sim CM(E, \delta_x, P)$  et est indépendante de  $\sigma(X_0, \dots, X_n)$ .*

On se souvient du résultat en terme de loi conditionnelle mais c'est la formule (4.2) qui est son interprétation concrète qu'on utilise dans les exercices et autres applications.

*Démonstration.* Par la propriété caractéristique de l'espérance conditionnelle, il s'agit de vérifier pour tout  $A \in \mathcal{F}_n = \sigma(X_0, \dots, X_n)$  :

$$\mathbf{E}_\mu(F \circ \theta_n 1_A) = \mathbf{E}_\mu(\mathbf{E}_{X_n}(F) 1_A).$$

Vu la forme des tribus pour les variables discrètes, on peut prendre  $A = \{X_0 = x_0, \dots, X_n = x_n\}$ . Commençons par  $F = 1_{X_0=y_0, X_1=y_1, \dots, X_m=y_m}$ . Alors, on a :

$$\begin{aligned} \mathbf{E}_\mu(F \circ \theta_n 1_A) &= \mathbf{P}_\mu(X_0 = x_0, \dots, X_n = x_n = y_0, X_{n+1} = y_1, \dots, X_{n+m} = y_m) \\ &= 1_{x_n=y_0} \mu(\{x_0\}) P_{x_0, x_1} \dots P_{x_{n-1}, y_0} P_{y_0, y_1} \dots P_{y_{m-1}, y_m} \end{aligned}$$

De même, on a  $\mathbf{E}_y(F) = 1_{y=y_0} P_{y_0, y_1} \dots P_{y_{m-1}, y_m}$  donc

$$\mathbf{E}_\mu(\mathbf{E}_{X_n}(F) 1_A) = P_{y_0, y_1} \dots P_{y_{m-1}, y_m} \mathbf{P}_\mu(X_n = y_0, X_0 = x_0, \dots, X_n = x_n)$$

ce qui donne bien la même valeur que ci-dessus. Le lemme de classe monotone conclut à l'égalité pour tout  $F = 1_B$  vu que  $\nu_1(B) = \mathbf{E}_\mu((1_B) \circ \theta_n 1_A)$ ,  $\nu_2(A) = \mathbf{E}_\mu(\mathbf{E}_{X_n}(1_B)1_A)$  définissent deux mesures de probabilités qui s'accordent sur tous les  $B = \{X_0 = y_0, X_1 = y_1, \dots, X_m = y_m\}$  qui forment une classe stable par intersection fine qui engendre  $\mathcal{F}$ . Enfin, comme dans la construction de l'intégration ou le lemme de transfert, on étend par linéarité à  $F$  étagée puis par convergence monotone à  $F$  mesurable positive, puis à  $F$  intégrable (par exemple mesurable bornée).

Pour obtenir le résultat conditionnellement à  $\{X_n = x\}$  (qui est  $\mathcal{F}_n$  mesurable, on remarque

$$\mathbf{E}_\mu(F \circ \theta_n 1_A 1_{\{X_n=x\}}) = \mathbf{E}_\mu(\mathbf{E}_{X_n}(F)1_A 1_{\{X_n=x\}}) = \mathbf{E}_x(F)\mathbf{E}_\mu(1_A 1_{\{X_n=x\}}).$$

Donc en divisant par  $P(X_n = x)$  si c'est non nul :

$$\mathbf{E}_\mu(F \circ \theta_n 1_A | X_n = x) = \mathbf{E}_x(F)\mathbf{E}_\mu(1_A | X_n = x) = \mathbf{E}_\mu(F \circ \theta_n | X_n = x)\mathbf{E}_\mu(1_A | X_n = x).$$

ce qui est l'indépendance souhaitée (la dernière égalité vient du cas  $A = \Omega$ ) et donne la loi de  $(X_{n+m})_{m \geq 0} = ((X_m)_{m \geq 0}) \circ \theta_n$  qui est  $\mathbf{P}_x \sim CM(E, \delta_x, P)$ .  $\square$

On rappelle que la tribu  $\mathcal{F}_T$  d'un temps d'arrêt a été définie en sous-section 1.4.

**Théorème 47** (Propriété de Markov forte). *Soit  $\mu$  une loi de probabilité sur  $E$ ,  $P$  une matrice de transition et  $P_\mu \sim CM(E, \mu, P)$  la construction sur l'espace canonique  $\Omega$ ,  $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ . On note  $x \mapsto \mathbf{E}_x = E_{\delta_x}$  l'espérance pour la loi  $P_{\delta_x} \sim CM(E, \delta_x, P)$ . Pour tout  $F$  mesurable (positive ou bornée), tout temps d'arrêt  $T$  par rapport à la filtration  $(\mathcal{F}_n)$ , on a*

$$\mathbf{E}_\mu(1_{\{T < \infty\}} F \circ \theta_T | \mathcal{F}_T) = 1_{\{T < \infty\}} \mathbf{E}_{X_T}(F). \quad (4.3)$$

*Démonstration.* Pour utiliser la propriété caractéristique, il faut d'abord noter que par l'exercice 2,  $1_{\{T < \infty\}} X_T$  est  $\mathcal{F}_T$  mesurable de sorte que par composition et produit,  $1_{\{T < \infty\}} \mathbf{E}_{X_T}(F)$  est aussi  $\mathcal{F}_T$  mesurable.

Comme dans la preuve de la propriété de Markov Faible, par la propriété caractéristique de l'espérance conditionnelle, il suffit de vérifier, pour  $A \in \mathcal{F}_T$  mesurable :

$$\mathbf{E}_\mu(1_{\{T < \infty\}} F \circ \theta_T 1_A) = \mathbf{E}_\mu(1_{\{T < \infty\}} \mathbf{E}_{X_T}(F) 1_A).$$

En décomposant  $\{T < \infty\} = \cup_n \{T = n\}$  comme union disjointe et par  $\sigma$ -additivité, il suffit de vérifier :

$$\mathbf{E}_\mu(1_{\{T=n\}} F \circ \theta_T 1_A) = \mathbf{E}_\mu(1_{\{T=n\}} F \circ \theta_n 1_A) = \mathbf{E}_\mu(1_{\{T=n\}} \mathbf{E}_{X_n}(F) 1_A) = \mathbf{E}_\mu(1_{\{T=n\}} \mathbf{E}_{X_T}(F) 1_A).$$

Mais l'égalité au milieu est en effet vraie puisque par définition de  $\mathcal{F}_T$ ,  $A \cap \{T = n\} \in \mathcal{F}_n$  de sorte que le résultat se réduit à l'identité de Markov faible.  $\square$

**Corollaire 48.** *Sous les hypothèses du théorème, conditionnellement à l'événement  $\{X_T = x, T < \infty\}$  on a une chaîne de Markov  $(X_{T+m})_{m \geq 0} \sim CM(E, \delta_x, P)$  qui est indépendante de  $\mathcal{F}_T$ .*

*Démonstration.* C'est la même preuve que pour Markov simple, on le laisse en exercice.  $\square$

## 2.4 Simulation utilisant les sauts

Soit  $(X_n)$  une chaîne de Markov  $CM(E, \mu, P)$  et  $S_n$  la suite des temps de sauts, définie par  $S_0 = 0$  et

$$S_{n+1} = \inf\{k > S_n : X_k \neq X_{S_n}\}.$$

Il est facile de voir que les  $S_n$  sont des temps d'arrêts (exo) à valeurs dans  $\mathbb{N} \cup \{+\infty\}$  (bien sûr  $\inf \emptyset = +\infty$ ), et par définition  $X_m = X_{S_n}$  pour tout temps  $m$  élément de  $\{m \in \mathbb{N} : S_n \leq m < S_{n+1}\}$ . On peut donc reconstruire  $(X_n)$  à partir de la suite  $(S_n, X_{S_n}) \in F = \mathbb{N} \times E \cup \{(\infty, 0)\}$  (la valeur  $X_\infty = 0$  étant fixée par convention).

Nous allons voir que c'est une chaîne de Markov que l'on peut simuler comme en section 2.1. Comme on évite ainsi les sauts constants, cela est plus efficace, d'autant plus que ceux-ci sont nombreux (si les coefficients diagonaux de  $P$  sont proches de 1).

**Proposition 49.** *Soit  $(X_n)$  une chaîne de Markov  $CM(E, \mu, P)$ . Avec les notations précédentes,  $(S_n, X_{S_n})_{n \geq 0}$  est une chaîne de Markov  $CM(F, \delta_0 \times \mu, Q)$  avec  $F = \mathbb{N} \times E \cup \{(\infty, 0)\}$  et matrice de transition  $Q$  donnée par :*

$$Q((s, x), (t, y)) = P_{x,x}^{t-s-1} P_{x,y}, x \neq y, t > s, Q((s, x), (\infty, 0)) = 1_{\{P_{x,x}=1\}},$$

$$Q((\infty, 0), (\infty, 0)) = 1, Q((s, x), (t, y)) = 0 \text{ sinon.}$$

*Démonstration.* cf TD. □

## 3 Classification des états

La compréhension qualitative d'une chaîne de Markov va passer par la classification de l'espace des états en terme du retour possible ou non de la chaîne à un état donné et donc des régions que la chaîne peut ou non visiter selon son point de départ.

Dans cette perspective, les définitions suivantes vont être pertinentes. On fixe toujours une chaîne de Markov canonique  $(X_n)_{n \geq 0}$

**Définition 23.** Le temps de premier passage en  $x$  est le temps :

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

et le nombre de visites en  $x$  est le nombre :

$$N_x = \sum_{k=0}^{\infty} 1_{\{X_k=x\}}.$$

On note  $\rho(x, y) = \mathbf{P}_x(\tau_y < +\infty)$  et  $U(x, y) = \mathbf{E}_x(N_y)$ .

**Définition 24.** 1. Un état  $x \in E$  est dit *récurrent* si  $\rho(x, x) = \mathbf{P}_x(\tau_x < +\infty) = 1$ . Il est dit au contraire *transitoire* si  $\rho(x, x) = \mathbf{P}_x(\tau_x < +\infty) < 1$ .

2. On dit que  $x$  est connecté à  $y$ , qu'on note  $x \rightsquigarrow y$ , si  $\rho(x, y) = \mathbf{P}_x(\tau_y < +\infty) > 0$  ou de façon équivalente si il existe  $n$  tel que  $P_{x,y}^n > 0$ . (cf 3 du thm montrant aussi l'équivalence à  $U(x, y) > 0$ .)

On dit que  $x$  et  $y$  communiquent si  $x \rightsquigarrow y$  et  $y \rightsquigarrow x$  ce que l'on note  $x \longleftrightarrow y$ .

3. Un ensemble  $A \subset E$  est dit *fermé* si pour tout  $x \in A$ , si  $x \rightsquigarrow y$  alors  $y \in A$
4. Un ensemble  $A \subset E$  est dit *unre classe irréductible* si pour tout  $x, y \in A$  alors  $x \rightsquigarrow y$  et pour tout  $x \in A, y \notin A$  alors  $x \not\rightsquigarrow y$ .

Le résultat suivant résume les principales propriétés permettant d'utiliser ces notions :

- Théorème 50** (De classification des états). 1. Un état  $x \in E$  est récurrent si et seulement si  $U(x, x) = +\infty$  si et seulement si  $\mathbf{P}_x(N_x = +\infty) = 1$ .  
 - Dans ce cas, pour tout  $y \in E$   $\rho(y, x) = \mathbf{P}_y(N_x = +\infty)$ .  
 - Si de plus  $y \rightsquigarrow x$ , on a  $U(y, x) = +\infty$   
 - Si enfin  $x \rightsquigarrow y$  alors  $y$  est récurrent,  $\mathbf{P}_x(N_y = +\infty) = 1$  et  $y \rightsquigarrow x$ .
2. Un état  $x \in E$  est transitoire si et seulement si  $U(x, x) < +\infty$  si et seulement si  $\mathbf{P}_x(N_x = +\infty) < 1$  si et seulement si  $\mathbf{P}_x(N_x = +\infty) = 0$ .  
 Alors on a  $U(y, x) = \frac{\rho(y, x)}{1 - \rho(x, x)}$  si  $y \neq x$  et  $U(x, x) = \frac{1}{1 - \rho(x, x)}$ .
3. En général pour tout  $x, y$ , on a

$$U(x, y) = \sum_{n=0}^{\infty} P_{x, y}^n,$$

et pour  $x \neq y$ ,  $U(x, y) = \rho(x, y)U(y, y) \geq \rho(x, y)$ .

4. Enfin,  $E = R \cup T$  se partitionne en classes d'équivalence (les classes irréductibles) pour la relation de communication  $x \rightsquigarrow y$ . Chaque classe est constituée soit seulement d'états récurrents, soit seulement d'états transitoires. L'ensemble  $R$  des états récurrents est ainsi partitionné en sous-ensembles fermés irréductibles  $R_i$ , indicés par un ensemble  $I$  (dit autrement toute classe d'ensembles récurrents est automatiquement fermée irréductible). L'ensemble  $T = T_a \cup T_f$  des états transitoires se partitionne entre états (absorbés)  $T_a$  qui communiquent avec des états récurrents et des états  $x \in T_f$  ("fuyant à l'infini") qui ne communiquent avec aucun état récurrent et pour lesquels pour tout  $y$ ,  $\mathbf{P}_x(N_y < +\infty) = 1$ . Donc on a la partition :

$$E = T_f \cup T_a \cup \left( \bigcup_{i \in I} R_i \right).$$

**Classifier les états d'une chaîne de Markov, c'est trouver les classes irréductibles (savoir si elles sont closes ou non), les états récurrents et les composantes irréductibles de  $R$ , et d'autre part les états transitoires, ceux communicants vers des états récurrents et les autres.**

### 3.1 Exemples de Classifications

On consacrerà une section plus tard aux cas où l'espace d'états est fini. On traite ici quelques exemples infinis.

**Exemple 34. Variables indépendantes de loi  $\mu$ .** On sépare le support  $S = \{x : \mu(\{x\}) \neq 0\}$  de  $\mu$ . Pour tout  $x \in E, y \in S$ , on a  $x \rightarrow y$  donc  $x \rightsquigarrow y$ . Donc  $S$  est bien une classe irréductible. Si  $x \notin S$ ,  $U(x, x) = \sum P_{x, x}^n = I_{x, x} = 1 < +\infty$  donc  $x$  est transitoire.

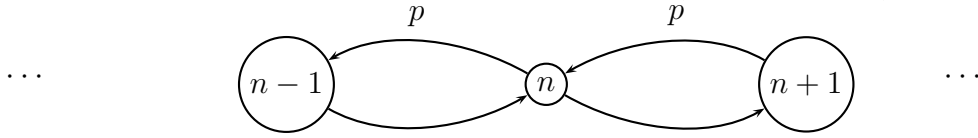
$S$  est la seule classe irréductible récurrente. En effet,  $P^2(x, y) = \sum_{k \in S} P_{x, k} P_{k, y} = \sum_{k \in S} \mu(\{k\}) \mu(\{y\}) = \mu(\{y\})$ , donc tous ces éléments sont bien récurrents car si  $y \in E$   $U(y, y) = \sum_{n=0}^{\infty} P^n(y, y) \geq \sum_{n=1}^{\infty} \mu(\{y\}) = +\infty$ .

Enfin, les états transitoires sont dans  $T_a$  car ils sont tous reliés à  $S$  en un saut, et leurs classes irréductibles sont les singletons  $\{y\}, y \in S^c$  car ils n'ont aucun  $x$  avec  $x \rightarrow y$ .



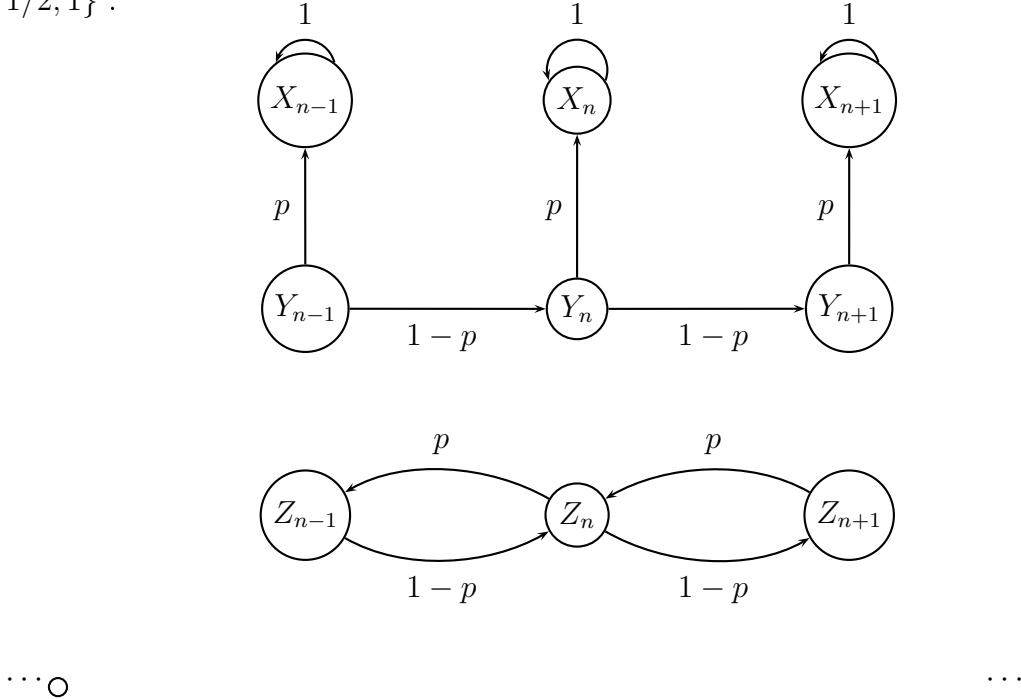
**Exemple 35.** On reprend l'exemple 27. Pour tout  $n$  on a  $U(x, x) = \sum_{n=0}^{\infty} P_{x,x}^n = \sum_{n=0}^{\infty} (1-p)^n = \frac{1}{p}$  (vu que le seul chemin de  $x$  à  $x$  est le chemin constant et en sommant ensuite la série géométrique). Donc si  $p > 0$  tous les états sont transitoires, et forment une classe close. Remarquez que  $\rho(x, x) \geq (1-p) > 0$  même si l'état est transitoire. Chaque singleton est une classe irréductible, car si  $m > n$  on a  $U(m, n) = 0$  (le processus ne peut décroître donc un état n'est équivalent qu'à lui-même). Si  $p = 0$  tous les singletons sont absorbants et forment une classe irréductible à part...

**Exemple 36. Une marche aléatoire asymétrique** sur  $\mathbb{Z}$   $P_{n,n+1} = 1-p, P_{n,n-1} = p$  avec  $p \neq 1/2$ .



$X_n = \sum_{i=1}^n U_i$  avec  $E(U_i) = -p + (1-p) = 1-2p \neq 0$  Par la LGN  $X_n/n \rightarrow E(U_1)$  on déduit  $X_n \rightarrow \pm\infty$  p.s. donc pour tout  $n$  la proba de retour en  $n$  est nulle. Tous les états sont donc transitoires. Cependant  $\rho(n, m) \geq (1-p)^{m-n}$  (chemin direct) si  $m > n$ ,  $\rho(n, m) \geq p^{n-m}$  si  $n > m$ . Donc tous les états communiquent et on a une seule classe irréductible transitoire (contenu dans  $T_f$ ).

**Exemple 37.** On considère la chaîne de Markov sur  $E = \{X_n, Y_n, Z_n, n \in \mathbb{Z}\}$  de diagramme suivant avec  $p \notin \{0, 1/2, 1\}$  :



La ligne des  $Z$  reproduit l'exemple précédent, c'est donc une classe irréductible de  $T_f$ .  $\rho(X_n, X_n) \geq P_{X_n, X_n} = 1$  donc tous les  $X_n$  sont absorbants (et a fortiori récurrent). Si  $Y_n$  était récurrent comme  $\rho(Y_n, X_n) \geq p > 0$  donc  $Y_n \rightsquigarrow X_n$ , on aurait  $X_n \rightsquigarrow Y_n$ , ce qui n'est pas le cas comme  $X_n$  absorbant, donc  $Y_n$  est transitoire. Comme les  $Y_n$  ne communiquent que vers des  $Y_m, m > n$ , les  $\{Y_n\}$  sont des classes irréductibles transitoires de  $T_a$  car elles sont absorbées par les  $X_n$ . Cet exemple a donc tous les types de classes du théorème de classification et une infinité de classes irréductibles récurrentes, une infinité de classes irréductibles transitoires, et une classe irréductible transitoire infinie.

### 3.2 Preuve (facultative) du théorème de classification

3/ On commence par prouver le point 3. Par définition et Fubini-Tonelli

$$U(x, y) = \sum_{n=0}^{\infty} \mathbf{E}_x(1_{X_n=y}) = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y) = \sum_{n=0}^{\infty} P_{x,y}^n,$$

comme voulu. Le second point vient de la propriété de Markov forte car pour que  $N_y$  soit non nul il faut d'abord que  $y$  soit atteint et  $N_y = N_y \circ \theta_{\tau_y}$  dès qu'on part de  $x \neq y$  (pour que  $n = 0$  ne soit pas une visite de  $y$  puisque  $n = 0$  est exclu de  $\tau_y \geq 1$  par définition) :

$$U(x, y) = \mathbf{E}_x(1_{\tau_y < \infty} N_y) = \mathbf{E}_x(1_{\tau_y < \infty} N_y \circ \theta_{\tau_y}) = \mathbf{E}_x(1_{\tau_y < \infty} E_{X_{\tau_y}}(N_y)) = \mathbf{E}_x(1_{\tau_y < \infty}) E_y(N_y) = \rho(x, y) U(y, y).$$

1a/ On montre ensuite 1 bien sûr  $\mathbf{P}_x(N_x = +\infty) = 1$  implique  $U(x, x) = +\infty$  qui implique  $x$  récurrent (car si  $x$  transitoire le 2 donnera  $U(x, x) < +\infty$ ). On suppose donc  $x$  récurrent et on calcule  $\mathbf{P}_x(N_x = +\infty)$ . On pose  $\tau_x^{(1)} = \tau_x$  et  $\tau_x^{(k+1)} = \inf\{n > \tau_x^{(k)} : X_n = x\} = \tau_x^{(k)} \circ \theta_{\tau_x}$ .

En appliquant la propriété de Markov forte au temps d'arrêt  $\tau_x$  on obtient vu  $X_{\tau_x} = x$  :

$$\mathbf{P}_x(\tau_x^{(k+1)} < +\infty) = \mathbf{E}_x(1_{\tau_x < \infty} 1_{\{\tau_x^{(k)} < +\infty\}} \circ \theta_{\tau_x}) = \mathbf{E}_x(1_{\tau_x < \infty} E_{X_{\tau_x}}(1_{\{\tau_x^{(k)} < +\infty\}})) = \mathbf{E}_x(1_{\tau_x < \infty}) E_x(1_{\{\tau_x^{(k)} < +\infty\}}).$$

On obtient donc par une récurrence immédiate :  $\mathbf{P}_x(\tau_x^{(k+1)} < +\infty) = [\mathbf{P}_x(\tau_x < \infty)]^{k+1}$ . Or  $N_x = 1_{\{X_0=x\}} + \sum_{k=1}^{\infty} 1_{\{\tau_x^k < +\infty\}}$  donc

$$U(x, x) = 1 + \sum_{k=1}^{\infty} \rho(x, x)^k.$$

En particulier si  $\rho(x, x) = 1$  on obtient bien l'infini et plus par intersection croissante  $\mathbf{P}_x(N_x = +\infty) = \mathbf{P}_x(\forall k \tau_x^k < +\infty) = \lim_{k \rightarrow \infty} \mathbf{P}_x(\tau_x^{(k+1)} < +\infty) = 1$ . Ceci conclut l'équivalence du 1.

2/ De plus si  $\rho(x, x) < 1$ , on obtient  $U(x, x) = \frac{1}{1-\rho(x, x)}$ , comme annoncé en 2. Ceci conclut la preuve de l'équivalence du 1 et des formules du 2 (la première formule en combinant avec 3.) On voit aussi que  $x$  transitoire implique  $U(x, x) < +\infty$  donc  $\mathbf{P}_x(N_x = \infty) = 0$  (sinon un calcul d'intégrale donnerait  $U(x, x) = +\infty$ ) donc  $\mathbf{P}_x(N_x = \infty) < 1$  et la contraposé de ce qu'on vient de montrer donne que  $x$  est transitoire. Ceci conclut la preuve du 2.

1b/ On reprend la preuve des assertions du 1. Si  $y = x$  et  $x$  récurrent, on vient de voir  $\mathbf{P}_x(N_x = \infty) = 1 = \rho(x, x)$ . On applique de nouveau la propriété de Markov forte pour  $y \neq x$  :

$$\mathbf{P}_y(N_x = \infty) = \mathbf{E}_y(1_{\tau_x < \infty} 1_{N_x = \infty} \circ \theta_{\tau_x}) = \mathbf{E}_y(1_{\tau_x < \infty}) \mathbf{P}_x(N_x = \infty) = \mathbf{E}_y(1_{\tau_x < \infty})$$

la dernière égalité vient de  $x$  récurrent.

1c/ Si  $x$  récurrent on a  $U(x, x) = +\infty$  et la formule du 3 donne  $U(y, x) = \rho(y, x) U(x, x)$  et si  $y \rightsquigarrow x$ , donc  $\rho(y, x) > 0$  donc  $U(y, x) = +\infty$ .

1d/ De même, on peut supposer que  $y \neq x$ , on applique de nouveau la propriété de Markov forte après avoir remarqué que  $\{N_x < \infty\} \supset \{\tau_y < \infty, \tau_x \circ \theta_{\tau_y} = +\infty\}$  correspondant à la stratégie d'aller en  $y$  et ne plus revenir en  $x$  ce qui garantit un nombre fini de passages en  $x$  (avant d'atteindre  $y$ ). Donc si  $x$  récurrent

$$0 = \mathbf{P}_x(N_x < \infty) \geq \mathbf{E}_x(1_{\{\tau_y < \infty\}} 1_{\{\tau_x = +\infty\}} \circ \theta_{\tau_y}) = \mathbf{P}_x(\tau_y < \infty) \mathbf{P}_y(\tau_x = \infty)$$

donc si  $x \rightsquigarrow y$  le premier terme est positif et donc  $\mathbf{P}_y(\tau_x = \infty) = 0$  soit  $\rho(y, x) = \mathbf{P}_y(\tau_x < \infty) = 1$  et donc  $y \rightsquigarrow x$ . Donc  $U(x, y) = \rho(x, y) U(y, y) \geq \rho(x, y) > 0$  et de même  $U(y, x) > 0$  donc soit  $n, m$

avec  $P_{x,y}^n > 0, P_{y,x}^m > 0$ , donc :

$$U(y, y) = \sum_{k=0}^{\infty} P_{y,y}^k \geq \sum_{k=0}^{\infty} P_{y,x}^m P_{x,x}^k P_{x,y}^n = P_{y,x}^m U(x, x) P_{x,y}^n = +\infty$$

donc on trouve  $y$  récurrent. De là l'assertion précédente montre  $\mathbf{P}_x(N_y = +\infty) = \rho(x, y)$  et on a vu  $\rho(x, y) = 1$  (par symétrie  $y/x$ . Ceci conclut la preuve du 1.

4/ la relation de communication entre états est évidemment symétrique et réflexive, elle est aussi transitive (car la relation de connexion à partir de laquelle elle est définie est clairement transitive). C'est donc une relation d'équivalence qui partage donc  $E$  en classe d'équivalence, qui sont par définition des composantes irréductibles. Par 1/ dès qu'un membre d'une classe d'équivalence est récurrent, tous les autres le sont. Donc de même si un membre est transitoire, tous les autres le sont. Par 1. encore, les classes récurrentes sont closes car si  $x$  récurrents et  $x \rightsquigarrow y$ , on a vu que  $x, y$  sont dans la même classe. D'où la décomposition de  $R$  en  $R_i$ .

Enfin par 1. si  $y$  récurrent  $\mathbf{P}_x(N_y < +\infty) = 1$  si et seulement si  $\rho(x, y) = 0$  soit si et seulement si  $x$  n'est pas connecté à  $y$ . Si un état transitoire n'est connecté à aucun état récurrent (soit  $x \notin T_a$ ) alors  $\mathbf{P}_x(N_y < +\infty) = 1$  pour tout  $y$  récurrent. Si maintenant  $y$  transitoire on a  $U(y, y) < +\infty$  et par 3  $U(x, y) = \rho(x, y)U(y, y) < +\infty$  donc  $\mathbf{P}_x(N_y < +\infty) = 1$  (sans quoi  $U(x, y) \geq \mathbf{P}_x(N_y = +\infty)\infty$  serait infini). Dans tous les cas on a bien  $\mathbf{P}_x(N_y < +\infty) = 1$  comme voulu. Et ceci pour tout  $y$  entraîne aussi réciproquement que  $x$  n'est relié à aucun récurrent.  $T_a$  et  $T_f$  forment bien une partition de  $T$ .

## 4 Mesures invariantes

On cherche à déterminer la loi asymptotique de  $X_n$  quand  $n \rightarrow \infty$ . Il se trouve qu'elle va être reliée au temps moyen passé sur chaque sommet et qu'elle va donc nous donner accès à un moyen de calculer ce temps.

**Définition 25.** Soit  $(X_n)_{n \geq 0}$ , une chaîne de Markov de matrice de transition  $P$  sur  $E$ . Une mesure  $\mu$  sur  $E$  est une *mesure invariante* si dès que  $X_0$  a loi  $\mu$ ,  $X_1$  a aussi loi  $\mu$  (et donc  $X_n$  a loi  $\mu$  pour tout  $n$ ) : c'est-à-dire si :

$$\forall x \in E, \mu(\{x\}) = \sum_{y \in E} \mu(\{y\}) P_{y,x}.$$

*Remarque 8.* Lorsque  $E$  est fini on obtient un système linéaire à  $|E|$  inconnus et le même nombre d'équations qu'il est facile de résoudre.

On rappelle qu'une chaîne de Markov est dite *irréductible récurrente* si elle est irréductible et que sa seule classe d'équivalence est composée d'états récurrents.

**Théorème 51** (des mesures invariantes). 1. Toute chaîne de Markov irréductible récurrente admet une mesure invariante non nulle.

2. Si  $\mu_1, \mu_2$  sont deux mesures invariantes non nulles sur une chaîne de Markov irréductible récurrente, alors il existe  $c > 0$  tel que  $\mu_1 = c\mu_2$ .

3. Pour une chaîne de Markov irréductible récurrente on a deux possibilités.

(a) Soit toutes les mesures invariantes vérifient  $\mu(E) = +\infty$  et dans ce cas pour tout  $x \in E$  :  $\mathbf{E}_x(\tau_x) = +\infty$ . On dit alors que la chaîne est *récurrente nulle*. De plus, on a pour tout  $y \in E$  :

$$\frac{1}{n} \sum_{k=0}^n 1_{\{X_k=x\}} \rightarrow 0, \mathbf{P}_y - p.s.$$

(b) Soit il existe une unique mesure de probabilité invariante  $\pi$  et alors pour tout  $x \in E$  :  $\mathbf{E}_x(\tau_x) = \frac{1}{\pi(\{x\})}$ . On dit alors que la chaîne est récurrente positive. De plus, on a pour tout  $y \in E$  :

$$\frac{1}{n} \sum_{k=0}^n 1_{\{X_k=x\}} \rightarrow \pi(\{x\}), \mathbf{P}_y - p.s.$$

*Remarque 9.* On utilise le théorème pour calculer  $\mathbf{E}_x(\tau_x)$  à partir de  $\pi$ , que l'on obtient en résolvant une équation, rarement l'inverse.

En général, on n'a pas

$$\lim_{n \rightarrow \infty} \mathbf{P}_x(X_n = y) \rightarrow_{n \rightarrow \infty} \mu(\{x\})$$

avec  $\mu$  mesure invariante. C'est vrai avec  $\mu$  la probabilité invariante quand la chaîne est irréductible récurrente positive et APERIODIQUE, c'est à dire que pour un  $x \in E$  (de façon équivalent pour tout  $x \in E$ )  $\text{pgcd}(n \geq 1 : P_{x,x}^n > 0) = 1$ .

**Définition 26.** Soit  $(X_n)_{n \geq 0}$ , une chaîne de Markov de matrice de transition  $P$  sur  $E$ . Une mesure  $\mu$  sur  $E$  est une *mesure réversible* si

$$\mu(\{y\})P_{y,x} = \mu(\{x\})P_{x,y}.$$

*Remarque 10.* Il est facile de voir qu'une mesure réversible est invariante. Mais c'est une condition plus restrictive à vérifier, la réciproque est fausse.

## 4.1 Preuve facultative (et admise en cours) du théorème d'existence 51

1. Par récurrence, on sait  $\tau_x < \infty$  o.s. On peut donc définir la mesure

$$\mu_x(\{y\}) = E_x\left(\sum_{k=1}^{\tau_x} 1_{\{X_k=y\}}\right) = E_x\left(\sum_{k=0}^{\tau_x-1} 1_{\{X_k=y\}}\right).$$

Montrons que  $\mu_x$  est invariante.

On note d'abord que par Markov faible au temp  $k-1$  vu  $\{\tau_x \geq k\} = \{\tau_x \leq k-1\}^c \in \mathcal{F}_{k-1}$ , on a :

$$E_x(1_{\{X_k=y, \tau_x \geq k, X_{k-1}=z\}}) = E_x(E(1_{\{X_k=y\}} | \mathcal{F}_{k-1}) 1_{\{\tau_x \geq k, X_{k-1}=z\}}) = P(z, y) E_x(1_{\{\tau_x \geq k, X_{k-1}=z\}}). \quad (4.4)$$

On utilise Fubini-Tonelli à la première ligne

$$\begin{aligned} \mu_x(\{y\}) &= E_x\left(\sum_{k=1}^{+\infty} 1_{\{X_k=y, \tau_x \geq k\}}\right) = \sum_{k=1}^{+\infty} E_x(1_{\{X_k=y, \tau_x \geq k\}}) \\ &= \sum_{k=1}^{+\infty} \sum_{z \in E} E_x(1_{\{X_k=y, \tau_x \geq k, X_{k-1}=z\}}) \\ &= \sum_{z \in E} \sum_{k=1}^{+\infty} P(z, y) E_x(1_{\{\tau_x \geq k, X_{k-1}=z\}}) \\ &= \sum_{z \in E} P(z, y) E_x\left(\sum_{k=1}^{\tau_x} 1_{\{X_{k-1}=z\}}\right) = \sum_{z \in E} P(z, y) E_x\left(\sum_{k=0}^{\tau_x-1} 1_{\{X_k=z\}}\right) \\ &= \sum_{z \in E} P(z, y) \mu_x(\{z\}) \end{aligned}$$

En deuxième ligne on a introduit une partition de  $\Omega$ , puis en troisième ligne utilisé le fait précédent. Ensuite, on reprend les calculs à l'envers. La dernière ligne est exactement l'invariance.

2/ Pour montrer l'unicité, on montre que toute mesure invariante  $\nu$  vérifie  $\nu(\{y\}) = \nu(\{x\})\mu_x(\{y\})$  (c'est possible car par définition de  $\tau_x$ ,  $\mu_x(\{x\}) = 1$ )

On voit d'abord par récurrence sur  $p$  :

$$\nu(\{y\}) \geq \nu(\{x\})E_x \sum_{k=0}^{\inf(\tau_x-1,p)} 1_{\{X_k=y\}} \quad (4.5)$$

si  $x = y$  c'est évident car  $\nu(\{y\}) = \nu(\{x\})\mu_x(\{x\})$ . On prend  $x \neq y$ . Le cas  $p = 0$  est évident car  $E_x(1_{\{X_0=y\}}) = 0 \leq \nu(\{y\})$

Si c'est vrai au rang  $p$ , on écrit l'invariance et l'hypothèse de récurrence

$$\begin{aligned} \nu(\{y\}) &= \sum_{z \in E} \nu(\{z\})P(z, y) \\ &\geq \sum_{z \in E} P(z, y)\nu(\{x\})E_x \left( \sum_{k=0}^{\inf(\tau_x-1,p)} 1_{\{X_k=z\}} \right) \\ &= \sum_{z \in E} P(z, y)\nu(\{x\}) \sum_{k=0}^p E_x(1_{\{X_k=z, \tau_x-1 \geq k\}}) \\ &= \sum_{z \in E} \nu(\{x\}) \sum_{k=0}^p E_x(1_{\{X_k=z, X_{k+1}=y, \tau_x-1 \geq k\}}) \\ &= \nu(\{x\}) \sum_{k=0}^p E_x(1_{\{X_{k+1}=y, \tau_x-1 \geq k\}}) \\ &= \nu(\{x\})E_x \left( \sum_{k=0}^{\inf(\tau_x-1,p)} 1_{\{X_{k+1}=y\}} \right) \\ &= \nu(\{x\})E_x \left( \sum_{k=1}^{\inf(\tau_x,p+1)} 1_{\{X_k=y\}} \right) \\ &= \nu(\{x\})E_x \left( \sum_{k=0}^{\inf(\tau_x-1,p+1)} 1_{\{X_k=y\}} \right) \end{aligned}$$

A la quatrième ligne on utilise (4.4), puis une partition après Fubini-Toneli, enfin un changement d'indice à l'avant dernière ligne et à la dernière ligne  $\tau_x < \infty$  p.s. donc  $X_0 \neq yP_x$  p.s et  $X_{\tau_x} \neq yP_x$ .

Cela termine la récurrence et par convergence monotone, on prend la limite  $p \rightarrow \infty$  et on obtient :

$$\nu(\{y\}) \geq \nu(\{x\})E_x \sum_{k=0}^{\tau_x-1} 1_{\{X_k=y\}} = \nu(\{x\})\mu_x(\{y\}) \quad (4.6)$$

Un récurrence évidente montre que  $\nu, \mu_x$  sont invariantes pour  $P^n$ .

On peut donc écrire :

$$\nu(\{x\}) = \sum_{y \in E} \nu(\{y\})P^n(y, x) \geq \sum_{y \in E} \nu(\{x\})\mu_x(\{y\})P^n(y, x) = \nu(\{x\})\mu_x(\{x\}) = \nu(\{x\}).$$

Comme on a inégalité terme à terme, on a égalité partout soit :  $\nu(\{y\})P^n(y, x) = \nu(\{x\})\mu_x(\{y\})P^n(y, x)$  mais pour tout  $y$  il existe  $n$  avec  $P^n(y, x) \neq 0$  par irréductibilité, donc  $\nu(\{y\}) = \nu(\{x\})\mu_x(\{y\})$ . CQFD.

3/ On voit par Fubini-Toneli et une partition :

$$\mu_x(E) = \sum_{y \in E} E_x \left( \sum_{k=0}^{\tau_x-1} 1_{\{X_k=y\}} \right) = E_x \left( \sum_{k=0}^{\tau_x-1} \sum_{y \in E} 1_{\{X_k=y\}} \right) = E_x \left( \sum_{k=0}^{\tau_x-1} 1 \right) = E_x(\tau_x)$$

Donc soit il existe un mesure invariante infinie et alors  $\mu_x(E) = E_x(\tau_x) = +\infty$  soit il existe une mesure finie et l'unique proba  $\pi$  est obtenue en normalisant  $\pi = \mu_x/E_x(\tau_x)$  et donc  $E_x(\tau_x)\pi(\{x\}) = 1$  comme annoncé.

Il reste à voir les convergences des moyennes en appliquant la LGN. Donnons juste l'idée de preuve. On voit que  $\tau_x^{(k)}$  de la section (3.2) vérifie que  $T_k = \tau_x^{(k)} - \tau_x^{(k-1)}$  sont i.i.d sous  $P_x$  de loi celle de  $\tau_x$ . Par la LGN on a

$$\lim_{k \rightarrow \infty} \frac{\tau_x^{(k)}}{k} = E_x(\tau_x).$$

Soit  $N_n(x) = \sum_{k=0}^n 1_{\{X_k=x\}}$  alors on voit par définition que  $\tau_x^{N_n(x)} \leq n < \tau_x^{N_n(x)+1}$ , par récurrence, on a  $N_n(x) \rightarrow \infty$ , donc par comparaison :

$$\lim_{n \rightarrow \infty} \frac{\tau_x^{N_n(x)}}{N_n(x)} = \lim_{n \rightarrow \infty} \frac{n}{N_n(x)} = \lim_{n \rightarrow \infty} \frac{\tau_x^{N_n(x)+1}}{N_n(x)+1} = E_x(\tau_x).$$

Vu la relation entre  $E_x(\tau_x) = \mu_x(E)$  cela conclut dans les deux cas. Le cas  $y \neq x$  ne change que  $\tau^{(1)}$  et est donc similaire.

## 5 Le cas des espaces d'états finis en détail.

Tout se simplifie sur les espaces d'états finis, on sait décrire toutes les mesures invariantes et la classification ne contient pas d'états transitoires fuyant à l'infini ni de classe récurrente nulle. Le théorème suivant résume la situation. On fait ensuite un exemple complet :

**Théorème 52.** *Soit  $E$  un espace d'état fini et  $(X_n) \sim CM(E, \mu, P)$ .*

1. *L'ensemble  $R$  des états récurrents est non-vide et toute classe irréductible récurrente est récurrente positive. De plus  $T_f = \emptyset$  et  $P(T_R < \infty) = 1$  ou  $T_R = \inf\{n \geq 1 : X_n \in R\}$ .*
2. *Une classe irréductible est close si et seulement si elle est récurrente. Autrement dit, les classes irréductibles dont aucune flèche ne sort est récurrente, sinon, les autres sont transitoires. Un état  $x$  est transitoire si et seulement si il existe un état  $y$  tel que  $x \rightsquigarrow y$  et  $y \not\rightsquigarrow x$ .*
3. *Les mesures invariantes ne chargent pas les états transitoires. Si  $R_1, \dots, R_r$  sont les classes de récurrences, chaque  $R_i$  porte une unique mesure de probabilité invariante  $\mu_i$  qui est l'unique vecteur propre à coordonnées positives et valeur propre 1 de la matrice transposée  $P^T$  restreinte à  $R_i$ . Toute mesure invariante est de la forme  $\sum_{i=1}^r \alpha_i \mu_i$ ,  $\alpha_i \geq 0$ .*

Le second point montre que la classification des états ne dépend que de la connectivité du graphe et pas des poids.

*Démonstration.* 1/ Pour  $y \in T$  transitoire,  $U(x, y) = \sum_{k=0}^{\infty} P_{x,y}^k \leq U(y, y) < \infty$  donc  $\lim_{k \rightarrow \infty} P_{x,y}^k = 0$   
 Or  $\sum_{y \in R} P_{x,y}^k + \sum_{y \in T} P_{x,y}^k = 1$  et la somme étant finie on obtient

$$\mathbf{P}_x(X_k \in R) = \sum_{y \in R} P_{x,y}^k \xrightarrow{k \rightarrow \infty} 1.$$

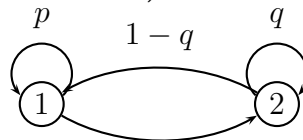
D'où vu  $\{X_k \in R\} \subset \{X_{k+1} \in R\}$ , on a  $\mathbf{P}_x(T_R < \infty) = \mathbf{P}_x(\exists k, X_k \in R) = \lim_{k \rightarrow \infty} \mathbf{P}_x(X_k \in R) = 1$  donc  $R$  est non vide. D'après le théorème 51, les classes irréductibles récurrentes possèdent une mesure invariante, nécessairement finie (vu le nombre fini d'états), donc ces classes sont récurrentes positives. Le fait que  $\mathbf{P}_x(T_R < \infty)$  pour tout  $x$  implique  $T_f = \emptyset$ . Par combinaison linéaire finie, on déduit  $\mathbf{P}_\mu(T_R < \infty) = 1$ .

2/ Pour tout  $y$  comme  $\mathbf{P}_y(T_R < \infty)$ , il existe  $x \in R$  avec  $y \rightsquigarrow x$ . Si  $y$  récurrent on a vu au théorème 50 1., que  $x \rightsquigarrow y$  et réciproquement si  $x \rightsquigarrow y$  le même théorème donne  $y$  récurrent. Donc par contraposé,  $y$  est transitoire si et seulement si il existe  $x$  (en fait récurrent) tel que  $y \rightsquigarrow x$  et  $x \not\rightsquigarrow y$ , comme voulu. Les deux premiers énoncés sont des conséquences évidentes : on a vu qu'une classe irréductible récurrente est toujours close et réciproquement, une classe irréductible close ne contient pas des  $y$  comme ci-dessus, donc ces membres ne sont pas transitoires. Être clos revient à dire qu'aucune flèche ne sort de la classe.

3/ Enfin, si  $y$  transitoire, alors pour tout  $x$ , on a vu en 1.  $\lim_{k \rightarrow \infty} P_{x,y}^k = 0$  donc pour une mesure invariante  $\mu(\{y\}) = \sum_{x \in E} \mu(\{y\}) P_{x,y}^n \rightarrow_{n \rightarrow \infty} 0$  donc  $\mu(\{y\}) = 0$  donc une mesure invariante ne charge pas les états transitoires. En conséquence, toute mesure invariante est combinaison linéaire de la normalisation de sa restriction à chaque classe irréductible récurrente. L'unicité de la mesure invariante de probabilité restreinte à chaque classe  $\mu_i$  vient du théorème 51. Enfin l'identification avec les vecteurs propres est directe par la définition.  $\square$

**Exemple 38.** On reprend l'exemple 25. Toutes les chaînes à deux états ont pour  $p, q \in [0, 1]$  une matrice de transition de la forme :  $P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$ .

Son graphe est :



Si  $p \neq 1 \neq q$ , les 2 états se connectent l'un à l'autre il n'y a qu'une classe irréductible récurrente positive. Pour trouver l'unique mesure de proba invariante, on résout :

$$\mu_1 + \mu_2 = 1, (\mu_1, \mu_2)P = (\mu_1, \mu_2)$$

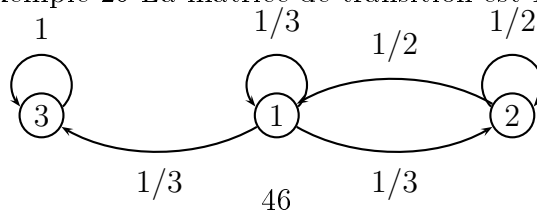
soit  $\mu_1 p + \mu_2(1-q) = \mu_1, \mu_1(1-p) + \mu_2 q = \mu_2$  qui donne  $\mu_2 = \mu_1(1-p)/(1-q)$  donc  $\mu_1 = 1/(1 + (1-p)/(1-q)) = (1-q)/(2-p-q), \mu_2 = (1-p)/(2-p-q)$ .

Si  $p = 1 = q$ , les deux états sont absorbants, n a 2 classes irréductibles récurrentes  $\{1\}, \{2\}$  et toute mesure est invariante.

Si  $p = 1 \neq q$  l'état 1 est absorbant donc récurrent (le dirac en  $\delta_1$  est invariant), 2 est connecté à 1 qui n'est pas connecté à 2 donc 2 est transitoire.  $\delta_1$  est donc l'unique mesure invariante.

**Exemple 39.** On reprend l'exemple 26 La matrice de transition est  $P = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ .

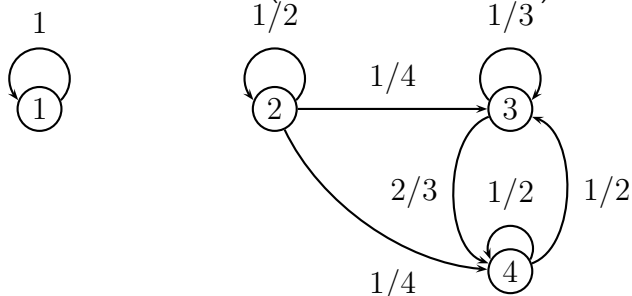
Son graphe est :



3 est absorbant donc récurrent.  $\{1, 2\}$  est une classe irréductible. Comme 1 est connecté à 3 mais pas l'inverse, 1 est transitoire et donc 2 aussi. L'unique mesure invariante est donc  $\delta_3$ .

**Exemple 40.** La matrice de transition est  $P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/4 & 1/4 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$ .

Son graphe est :



1 est absorbant donc récurrent.  $\{2\}, \{3, 4\}$  sont les autres classes irréductibles. Comme 2 est connecté à 3 mais pas l'inverse, 2 est transitoire. Comme  $\{2, 3, 4\}$  est clos donc forme une chaîne de Markov finie par elle même et qu'elle contient forcément une classe récurrente,  $R_2 = \{3, 4\}$  est une classe récurrente (on peut aussi dire qu'il est clos irréductible fini).

$\delta_1$  est une mesure invariante. Trouvons l'unique mesure de probabilité invariante associée à  $R_2 : \mu_3\delta_3 + \mu_4\delta_4$ . On doit vérifier  $\mu_3/3 + \mu_4/2 = \mu_3, 2\mu_3/3 + \mu_4/2 = \mu_4, \mu_3 + \mu_4 = 1$  soit  $\mu_3 = 3\mu_4/4 = 3/7$ . On déduit que  $\mathbf{E}_3(\tau_3) = 1/\mu_3 = 7/3$  et de même  $\mathbf{E}_4(\tau_4) = 7/4$ . On a aussi  $\mathbf{E}_1(\tau_1) = 1$ .

On applique enfin Markov faible au temps 1 après avoir remarqué que 3, 4 ne reviennent jamais en 2 :  $\mathbf{E}_2(\tau_2) = \mathbf{E}_2(\tau_2(1_{X_1=2} + 1_{X_1=3} + 1_{X_1=4})) = \mathbf{E}_2(\tau_2(1_{X_1=2})) = \mathbf{E}_2(1(1_{X_1=2})) + \infty = \infty$  vu que  $\mathbf{E}_3(\tau_2) = \infty$  puisqu'il n'y a pas de retour possible de 3 vers 2.

*Exercice 3.* Calculer les autres  $\mathbf{E}_i(\tau_j)$  dans les exemples précédents.

## 6 Exemple détaillé en finance : Modèle de Cox, Ross et Rubinstein

Nous vous référons au livre *Cours de Calcul stochastique appliqué à la finance* de Lambertson et Lapeyre pour plus de détails sur ce modèle (version discrète du modèle de Black et Scholes du dernier chapitre) et surtout pour un traitement beaucoup plus complet des notions utilisées en finance et que nous n'introduisons que sur le cas du modèle.

### 6.1 Un modèle général

En finance, on considère souvent un placement sans risque à intérêt fixe dont la valeur à l'instant  $n$  est  $Z_n = Y_0(1+r)^n$  si il vaut  $Y_0$  à l'instant 0 et gagne un taux d'intérêt  $r$  à chaque pas de temps. On considère également un placement risqué de valeur  $Y_n$  à l'instant  $n$ .

On suppose que

$$Y_{n+1} = T_n Y_n$$

avec  $T_n : \Omega \rightarrow \{1+a, 1+b\}$  une variable de Bernoulli telle que  $0 < 1+a < 1+b$ .  $a, b$  sont donc les taux d'intérêt gagnés possibles (dont l'un peut être négatif ou inférieur à  $r$ ) et  $T_n - 1$  le taux d'intérêt aléatoire donc risqué. On considère la filtration engendrée  $\mathcal{F}_n = \sigma(T_1, \dots, T_{n-1})$ .



1. Comme le placement à taux d'intérêt  $r$  est toujours disponible, on regarde  $X_n = Y_n/Z_n$  que l'on appelle *valeur actualisée* à l'instant  $n$ . On cherche à savoir quand  $\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n$  (on va expliquer pourquoi c'est une condition de viabilité du marché). Or  $X_{n+1} = T_n Y_n / Z_n (1+r) = X_n T_n / (1+r)$  et  $X_n$  est  $\mathcal{F}_n$  mesurable donc

$$\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n \frac{\mathbf{E}(T_n|\mathcal{F}_n)}{1+r}.$$

Donc  $\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n$  si et seulement si  $\mathbf{E}(T_n|\mathcal{F}_n) = 1+r$  pour tout  $n$ . On va voir que cela force  $X_n, Y_n$  à être des suites récurrentes aléatoires de la forme de l'exemple 28, donc des chaînes de Markov.

2. On parle d'*arbitrage* si il y a possibilité d'avoir un gain strictement positif avec une mise nulle en empruntant/achetant éventuellement l'actif sans risque. Par exemple, si  $r \geq b$  on vend un actif risqué (à découvert) pour une valeur  $Y_0$  et on achète avec le gain de l'actif sans risque. A la date  $N$ , on récupère l'intérêt  $Y_0(1+r)^N$  sur le marché sans risque et on paie la valeur  $Y_N$  de l'actif risqué soit un gain  $Y_0(1+r)^N - Y_N \geq Y_0[(1+r)^N - (1+b)^N]$ . On a donc gagné un montant strictement positif sans mise, donc sans risque de perdre de l'argent. De même, on montre que si  $r \leq a$ , il y a possibilité d'arbitrage (en empruntant sur le marché sans risque de quoi acheter l'actif risqué). Un marché financier est dit *vable* si il n'y a pas de possibilité d'arbitrage. Ici, cela impose  $r \in ]a, b[$ .
3. On suppose maintenant  $r \in ]a, b[$ . En finance, si l'actif actualisé  $X_n$  vérifie  $\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n$  (on dit  $X_n$  est une martingale) pour une certaine loi de probabilité, il est facile de voir qu'il n'y a pas de stratégie d'arbitrage et que le marché est (donc dit) viable (c'est pour nous la définition que l'on prendra d'un marché viable). Si en plus, comme ci-dessous, il y a une unique loi (ici pour les  $T_n$ ) qui permet d'obtenir ce résultat, on dit que le marché est *complet*, et on utilise cette (unique) loi pour calculer des prix et des stratégies (de couverture pour les marchés dérivés).

**Lemme 53.** *Le marché est viable si et seulement si les  $T_n$  sont i.i.d de loi*

$$P_{T_n} = \frac{b-r}{b-a} \delta_{1+a} + \frac{r-a}{b-a} \delta_{1+b}.$$

*Démonstration.* Si les  $T_n$  sont i.i.d comme indiqué, il suffit de calculer vu que  $T_n$  indépendant de  $\mathcal{F}_n$ ,

$$\mathbf{E}(T_n|\mathcal{F}_n) = \mathbf{E}(T_n) = \frac{b-r}{b-a}(1+a) + \frac{r-a}{b-a}(1+b) = 1 + \frac{a(b-r) + b(r-a)}{b-a} = 1+r,$$

comme il fallait d'après le point 1. Réciproquement, si pour tout  $n$   $\mathbf{E}(T_n|\mathcal{F}_n) = 1+r$ , vu que  $T_n = 1 + a1_{\{T_n=1+a\}} + b1_{\{T_n \neq 1+a\}} = 1 + b + (a-b)1_{\{T_n=1+a\}}$  est de Bernoulli, on voit

$$r-b = \mathbf{E}((a-b)1_{\{T_n=1+a\}}|\mathcal{F}_n).$$

Donc pour tout  $A \in \mathcal{F}_n$ , par (PC) pour l'espérance conditionnelle et le calcul ci-dessus

$$P(T_n = 1+a|A) = \frac{\mathbf{E}(1_A 1_{\{T_n=1+a\}})}{P(A)} = \frac{\mathbf{E}(1_A \mathbf{E}(1_{\{T_n=1+a\}}|\mathcal{F}_n))}{P(A)} = \frac{(r-b)\mathbf{E}(1_A)}{(a-b)P(A)} = \frac{(r-b)}{(a-b)},$$

donc  $\{T_n = 1+a\}, \{T_n = 1+b\}$  sont indépendants de  $A$  donc de  $\mathcal{F}_n$  comme voulu, soit  $T_n$  indépendant de  $\mathcal{F}_n$  et est de loi souhaitée. □

## 6.2 Etude du point de vue des chaînes de Markov dans un cas simple

**Exemple 41** (Modèle simple de portefeuille d'actions). Il est utile de considérer le cas symétrique  $r = \frac{a+b}{2}$ . Dans ce cas, on appelle  $r$  taux d'actualisation et  $\sigma = \frac{b-a}{2}$  le coefficient de volatilité. On pose  $\epsilon_n = \frac{T_n - (1+r)}{\sigma} \in \{-1, 1\}$ . On a donc

$$Y_n = (1 + r + \sigma\epsilon_n)Y_{n-1}.$$

On a donc vu que  $X_n = (1+r)^{-n}Y_n$  est une martingale si et seulement si les  $\epsilon_n$  sont i.i.d avec  $P_{\epsilon_n} = \frac{1}{2}(\delta_1 + \delta_{-1})$ . En supposant ce cadre, on voit que  $Y_n$  est une martingale si et seulement si  $r = 0$ , une sous-martingale (resp. surmartingale) si et seulement si  $r > 0$  (resp.  $r < 0$ ). Noter que la condition  $1+a > 0$  revient à  $|\sigma| < 1+r$ .

On prend  $Y_0 = 1$

Comme c'est une suite récurrente aléatoire comme à l'exemple 28, c'est une chaîne de Markov. Les valeurs qu'elle prend sont dans l'espace d'état  $E = \{(1+r+\sigma)^k(1+r-\sigma)^l, k, l \in \mathbb{N}\}$  Bien sûr, l'espace d'état change beaucoup si il existe ou non  $k, l$  tel que  $(1+r+\sigma)^k(1+r-\sigma)^l = 1$ .

**1/ Cas**  $\forall k, l(1+r+\sigma)^k(1+r-\sigma)^l \neq 1$

Dans ce cas, les états sont décrits par les couples  $(k, l)$  et la somme  $k+l$  augmente de 1 à chaque pas, donc chaque état est une classe irréductible et tous les états sont transitoires.

**2/ Cas**  $\exists k, l(1+r+\sigma)^k(1+r-\sigma)^l = 1, (1+r+\sigma)(1+r-\sigma) \neq 1$

Sans spécifier l'espace d'état exactement, on peut tout de même classifier partiellement la chaîne. Quel que soit l'état  $(1+r+\sigma)^K(1+r-\sigma)^L$  on a une proba  $(1/2)^{k+l-1}$  d'arriver en :  $(1+r+\sigma)^K(1+r-\sigma)^L(1+r+\sigma)^{k-1}(1+r-\sigma)^l = (1+r+\sigma)^{K-1}(1+r-\sigma)^L$  et de même à  $(1+r+\sigma)^K(1+r-\sigma)^{L-1}$  donc tous les états sont connectés à tous les états, il n'y a qu'une classe irréductible.

Voyons que tous les états sont transitoires. En prenant le logarithme  $\log(Y_n) = \sum_{k=1}^n \log(1+r+\sigma\epsilon_k)$  est une marche aléatoire et

$$E(\log(1+r+\sigma\epsilon_n)) = [\log(1+r+\sigma) + \log(1+r-\sigma)]/2 = \log((1+r)^2 - \sigma^2)/2 \neq 0$$

par l'hypothèse. Donc on est dans le cas de l'exemple 36, la loi des grands nombres donne  $p.s. \log(Y_n) \rightarrow \infty$  si  $(1+r)^2 - \sigma^2 > 1$  soit  $Y_n \rightarrow +\infty p.s.$  soit  $p.s. \log(Y_n) \rightarrow -\infty$  si  $(1+r)^2 - \sigma^2 < 1$  soit  $Y_n \rightarrow 0 p.s.$  Dans les deux cas, cela implique que  $p.s.$  la chaîne ne revient plus en aucun état fixé, qui est donc transitoire. On a donc une seule classe irréductible transitoire.

**3. Cas**  $(1+r+\sigma)(1+r-\sigma) = 1$

Alors l'espace d'état  $E = \{(1+r+\sigma)^n, n \in \mathbb{Z}\}$  et on a passage de  $(1+r+\sigma)^n \rightarrow (1+r+\sigma)^{n+1}$  avec proba  $1/2$  et  $(1+r+\sigma)^n \rightarrow (1+r+\sigma)^{n-1}$  avec proba  $1/2$ . Comme avant la chaîne est irréductible. Voyons qu'elle est récurrente.

On calcule  $U((1+r+\sigma)^n, (1+r+\sigma)^n) = \sum_{k=0}^{\infty} P_{(1+r+\sigma)^n, (1+r+\sigma)^n}^k = \sum_{k=0}^{\infty} P_{1,1}^k$  vu que la loi de saut est invariante par translation multiplicative. Noter que

$$P_{1,1}^k = P\left(\sum_{n=1}^k \log(1+r+\sigma\epsilon_n) = 0\right) = P\left(\log(1+r+\sigma) \sum_{n=1}^k \epsilon_n = 0\right) = P\left(\sum_{n=1}^k (\epsilon_n + 1)/2 = k/2\right)$$

vu que  $\log(1+r \pm \sigma) = \pm \log(1+r+\sigma)$  par l'hypothèse. Or  $\sum_{n=1}^k (\epsilon_n + 1)/2$  est la somme de  $k$  bernoulli indépendante donc une loi binomiale  $(k, 1/2)$ , donc la valeur est nulle si  $k$  impair, et si  $k = 2l$  on a  $P(\sum_{n=1}^k (\epsilon_n + 1)/2 = l) = \frac{(2l)!}{(l!)^2 2^l}$  donc :

$$U((1+r+\sigma)^n, (1+r+\sigma)^n) = \sum_{l=1}^{\infty} \frac{(2l)!}{(l!)^2 4^l} = +\infty.$$

En effet par la formule de Stirling  $n! \sim e^{-n} n^n \sqrt{2\pi n}$  donc  $\frac{(2l)!}{(l!)^2 4^l} \sim \frac{(2l)^{2l} e^{-2l} \sqrt{4\pi l}}{l^{2l} e^{-2l} (2\pi l)^{4^l}} = \frac{1}{\sqrt{\pi l}}$  terme général d'une série divergente (de Riemann) donc le critère des équivalents conclut à la divergence annoncée. On obtient donc la récurrence annoncée.

Montrons enfin que la chaîne est récurrente nulle. En effet,  $\mu(\{(1+r+\sigma)^n\}) = 1$  définit une mesure invariante car  $\mu(\{(1+r+\sigma)^n\}) = 1/2\mu(\{(1+r+\sigma)^{n-1}\}) + 1/2\mu(\{(1+r+\sigma)^{n+1}\})$  est l'équation d'une mesure invariante de notre chaîne. C'est une mesure de masse infinie, donc la chaîne est récurrente nulle.

### 6.3 Introduction aux produits dérivés

Ceci prépare au chapitre sur le mouvement brownien car le modèle ci-dessus converge en loi (des marginales fini-dimensionnelles) vers un mouvement brownien géométrique (une fois correctement normalisé).

1. Les fameux "produits dérivés" intervenant en finance, parmi les plus simples, sont les options européennes d'achat et de vente construites sur l'actif risqué  $Y_n$ . Une option d'achat (ou en anglais call) est le droit (mais non l'obligation) d'acheter l'actif à l'instant  $N$  au prix  $K$  (l'adjectif européen désigne les options d'achat-vente à terme, ici l'instant final  $N$ ). Elle rapporte donc la différence  $C_N = (Y_N - K)_+ = \max(0, Y_N - K)$  (puisque si  $Y_N - K \neq 0$  on n'a pas intérêt à acheter l'actif comme on a le droit avec l'option). De même, une option de vente (ou put) au prix  $K$  rapporte  $P_N = (K - Y_N)_+ = \max(0, K - Y_N)$ . On appelle valeur de l'option à l'instant  $n$ , l'actif dont la valeur actualisée est une martingale fermée de valeur à l'instant  $N$ , la valeur (actualisée) de l'option ( $C_N$  ou  $P_N$ ), soit :

$$C_n = Z_n \mathbf{E}\left(\frac{(Y_N - K)_+}{Z_N} \middle| \mathcal{F}_n\right) = (1+r)^{n-N} \mathbf{E}((Y_N - K)_+ | \mathcal{F}_n), \quad P_n = (1+r)^{n-N} \mathbf{E}((K - Y_N)_+ | \mathcal{F}_n).$$

La probabilité pour laquelle est calculée ces valeurs (dans un marché viable et complet) est celle pour laquelle le prix de l'actif actualisé est une martingale.

2. Calculons la "relation de parité call-put" :

$$\begin{aligned} C_n - P_n &= (1+r)^{n-N} \mathbf{E}((S_N - K)_+ - (K - S_N)_+ | \mathcal{F}_n) = (1+r)^{n-N} \mathbf{E}(S_N - K | \mathcal{F}_n) \\ &= (1+r)^n \mathbf{E}(X_N | \mathcal{F}_n) - (1+r)^{n-N} K = (1+r)^n X_n - (1+r)^{n-N} K \\ &= S_n - K(1+r)^{n-N}. \end{aligned}$$

On a utilisé que  $X_n = S_n(1+r)^{-n}$  est une martingale par hypothèse et la relation élémentaire  $(S_N - K)_+ - (K - S_N)_+ = S_N - K$ .

3. Calculons finalement la valeur explicite de  $C_n$  dans ce modèle c'est à dire

$$C_n = (1+r)^{n-N} \mathbf{E}((Y_n T_n \dots T_{N-1} - K)_+ | \mathcal{F}_n).$$

Pour appliquer la proposition 39 puisque  $Y_n$  et  $T_n \dots T_{N-1}$  sont indépendants et  $\ln(T_n \dots T_{N-1})$  est une somme de v.a. de Bernoulli indépendantes de loi  $p\delta_{\ln(1+a)} + (1-p)\delta_{\ln(1+b)}$  avec  $p = \frac{b-r}{b-a}$

de sorte que  $\frac{1}{\ln(\frac{1+b}{1+a})}[\ln(T_n \dots T_{N-1}) - (N-n) \ln(1+a)]$  est de loi binomiale  $\mathcal{B}(N-n, 1-p)$  (et vaut  $j$  quand  $T_n \dots T_{N-1}$  vaut  $(1+b)^j(1+a)^{N-n-j}$ ,  $N \geq n+1$ ) et donc on calcule par transfert

$$c(n, s) := \mathbf{E}((sT_n \dots T_{N-1} - K)_+) = \sum_{k=0}^{N-n} C_{N-n}^k p^k (1-p)^{N-n-k} (s(1+a)^k (1+b)^{N-n-k} - K)_+.$$

On obtient finalement la formule explicite  $C_n = (1+r)^{n-N} c(n, Y_n)$ .

# Chapitre 5

## Processus de Poisson

Nous continuons en étudiant un processus à temps continue mais à espace d'état discret (à valeur entière), le processus de Poisson. Les temps de saut du processus modélisent les temps de désintégration d'atomes radioactifs, les arrivées des appels à un central téléphonique, etc. Nous commençons par passer en revue les différents points de vue en terme de nombre de désintégration ayant eu lieu à l'instant  $t$ , temps de la  $n$ -ième désintégration, ou intervalle de temps entre deux désintégrations.

### 1 Définition en terme de processus de comptage

**Définition 27.** Un processus  $(N_t)_{t \in \mathbb{R}^+}, N_t : (\Omega, \mathcal{T}, P) \rightarrow \mathbb{R}$  est un processus de comptage si  $N : \mathbb{R}^+ \times \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ , est mesurable  $N(0) = 0$  et si il existe un  $A \in \mathcal{T}$  vérifiant  $P(A) = 1$  et pour tout  $\omega \in A$ ,  $t \mapsto N_t(\omega)$  est croissant (soit  $N$  p.s. croissant).

**Définition 28.** Un processus de comptage  $(N_t)_{t \in \mathbb{R}^+}$  est dit :

1. à *accroissements indépendants* si pour tout  $t_1 < \dots < t_n$   $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$  sont indépendants
2. à *accroissements stationnaires* si pour tout  $s < t$   $N_t - N_s$  a même loi que  $N_{t-s}$ .
3. *localement continu à droite en probabilité* si pour tout  $t \geq 0$   $\lim_{h \rightarrow 0^+} P(N_{t+h} - N_t \geq 1) = 0$ .
4. *p.s. à sauts simples* si il existe  $A$  mesurable tel que  $P(A) = 1$  et pour tout  $\omega \in A$ , pour tout  $t \geq 0$   $\lim_{h \rightarrow 0} N_{t+h}(\omega) - N_t(\omega) \leq 1$ . (noter que la limite existe p.s. vu que  $N$  est p.s. croissant).

**Définition 29.** Un processus de comptage  $(N_t)_{t \in \mathbb{R}^+}$  est un *processus de Poisson d'intensité  $\lambda > 0$*  si le processus est à accroissements indépendants et pour tout  $t, h > 0$ ,  $N_{t+h} - N_t$  est de loi de Poisson de paramètre  $\lambda h$ , soit :

$$P(N_{t+h} - N_t = n) = e^{-\lambda h} \frac{(\lambda h)^n}{n!}, \quad n \geq 0.$$

On voit immédiatement qu'un processus de Poisson est à accroissements stationnaires.

On a aussi :

**Lemme 54.** *Un processus de Poisson est localement continu à droite en probabilité.*

*Démonstration.* On a  $P(N_{t+h} - N_t \geq 1) = 1 - P(N_{t+h} - N_t = 0) = 1 - e^{-\lambda h} \rightarrow_{h \rightarrow 0^+} 0$ . □

**Lemme 55.** *Le processus de Poisson est p.s. à sauts simples. De plus pour tout  $t$  il n'a p.s. qu'un nombre fini de saut sur  $[0, t]$ .*

*Démonstration.* Par intersection dénombrable d'évènement de proba 1 obtenu par stationnarité, il suffit de voir que la restriction en temps à  $[0,1]$  vérifie la même propriété. On pose

$$A^c = \bigcap_n B_n$$

avec  $B_n = \{\exists j \in [1, 2^n] : N_{j/2^n} - N_{(j-1)/2^n} \geq 2\}$  Alors la suite des  $B_n$  est décroissante (par croissance de  $N$ ) et par stationnarité :

$$P(B_n) \leq \sum_{j=1}^{2^n} P(N_{1/2^n} \geq 2) = 2^n \sum_{k=2}^{\infty} \frac{(\lambda/2^n)^k}{k!} e^{-\lambda/2^n} \leq 2^n \sum_{k=0}^{\infty} \frac{(\lambda/2^n)^{k+2}}{k!} e^{-\lambda/2^n} = 2^n (\lambda/2^n)^2 \rightarrow_{n \rightarrow \infty} 0$$

d'où  $P(A^c) = \lim P(B_n) = 0$  de plus si  $\omega \in A \exists n, \forall j \in [1, 2^n] : N_{j/2^n} - N_{(j-1)/2^n} \leq 1$  ce qui empêche aucun saut plus grand que 1 **et un nombre infini de saut sur  $[0, 1]$  (il y a un  $n$  avec au plus  $2^n$  sauts)**.  $\square$

**Théorème 56.** *Un processus de comptage  $(N_t)_{t \in \mathbb{R}^+}$  (c'est-à-dire un processus p.s. croissant à valeur entière partant de  $N_0 = 0$ ) est un processus de Poisson si et seulement si il vérifie les 4 conditions :*

1. *il est à accroissements stationnaires et indépendants,*
2. *il est non trivial (à savoir, non p.s. nul soit  $\exists t > 0 P(N_t > 0) > 0$ ),*
3. *il est localement continu à droite en probabilité : pour tout  $t \geq 0 \lim_{h \rightarrow 0^+} P(N_{t+h} - N_t \geq 1) = 0$ .*
4. *il est p.s. à sauts simples : il existe  $A$  mesurable tel que  $P(A) = 1$  et pour tout  $\omega \in A$ , pour tout  $t \geq 0 \lim_{h \rightarrow 0} N_{t+h}(\omega) - N_t(\omega) \leq 1$ .*

Alors, l'intensité du processus est donnée par la limite suivante qui existe :

$$\lim_{h \rightarrow 0^+} \frac{1}{h} P(N_h = 1) = \lambda > 0.$$

*Démonstration.* On a déjà vu qu'un processus de Poisson vérifiait les 4 conditions et  $\frac{1}{h} P(N_h = 1) = \lambda e^{-\lambda h}$  tend bien vers  $\lambda$ .

Considérons donc  $N$  un processus vérifiant les 4 conditions. Soit  $p(h) = P(N_h \geq 1)$  On a par stationnarité  $p(h) = P(N_{t+h} - N_t \geq 1)$ . On sait qu'il existe  $h$  avec  $p(h) > 0$  par non trivialité et la condition de continuité dit que  $p(h) \rightarrow_{h \rightarrow 0^+} 0$ .

1. Montrons d'abord que  $p(t) = 1 - e^{-\lambda t}$  avec  $\lambda > 0$ .

En effet, par croissance des incréments on a  $f(t+s) := 1 - p(t+s) = P(N_{t+h} = 0) = P(N_{t+h} - N_h = 0 \text{ et } N_h = 0)$  et par indépendance puis stationnarité :

$$f(t+s) = P(N_{t+h} = 0) = P(N_{t+h} - N_h = 0) P(N_h = 0) = f(t) f(s).$$

Maintenant, obtenir la formule pour  $f$  est un exercice d'analyse standard on écrit  $f(h) = e^{-\lambda h} < 1$  donné par  $P(N_h) > 0$ , ce qui implique  $\lambda > 0$ . Par multiplicativité pour  $p$  entier  $f(ph) = e^{-p\lambda h}$  puis pour  $p, q$  entier  $f(ph/q)^q = f(ph)$  donc  $f(th) = e^{-t\lambda h}$  pour  $t$  rationnel. Enfin, par multiplicativité, et continuité à droite de  $f$  en 0,  $f$  est continue à droite en tout point, et en prenant  $t_n \in \mathbb{Q}$  décroissante avec  $t_n \rightarrow t$  on obtient à la limite  $f(th) = e^{-t\lambda h}$ . Ceci donne la formule pour  $p(t)$ .

2. Pour tout intervalle  $]a, b]$ , l'évènement,  $I_{a,b} = \{N_a < \infty \text{ et } N_b = +\infty\}$ , disant que  $N_t$  a une infinité de sauts dans  $]a, b]$  est de probabilité nulle.

Sinon en prenant  $c = (a + b)/2$  on voit que  $I_{a,b} = I_{a,c} \cup I_{c,b}$  et par l'hypothèse de stationnarité  $P(I_{a,c}) = P(I_{c,b})$  et comme les deux dépendent des accroissements ils sont indépendants mais de plus ils sont disjoints vu la condition  $N_b = +\infty$  incompatible avec  $N_b < \infty$  soit  $1 - P(I_{a,b}) = (1 - P(I_{a,c}))^2 = (1 - P(I_{a,b})/2)^2 = 1 + P(I_{a,b})^2/4 - P(I_{a,b})$  soit  $P(I_{a,b}) = 0$  comme voulu.

3. Soit  $t > 0$  fixé.  $A^c = \bigcap_n B_n$  avec  $B_n = \{\exists j \in [1, 2^n] : N_{jt/2^n} - N_{(j-1)t/2^n} \geq 2\}$  Montrons que  $P(A) = 1$ . En effet, pour  $\omega$  fixé (dans un ensemble de proba 1), il y a au plus  $m$  sauts dans  $[0, t]$  par le point précédent et aucun saut double, donc on a une distance minimale entre 2 sauts et donc il existe  $n$  minimal avec  $1/2^n$  entre 2 sauts donc  $\omega \in B_n^c \subset A$ .

4. Montrons que  $N_t$  suit la loi de Poisson de paramètre  $\lambda t$ . Soit  $X_k = N_{tk/n} - N_{t(k-1)/n}$ ,  $1 \leq k \leq n$  sont i.i.d à valeurs entières par hypothèse. Soit  $Y_k = 1_{\{X_k \geq 1\}}$  et  $S_n = Y_1 + \dots + Y_n$  de sorte que  $S_n \leq N_t$ . Soit  $A$  l'ensemble obtenu à la question précédente, comme  $\{S_n \neq N_t\} = \{\exists j X_j \geq 2\}$  (car l'égalité  $S_n = N_t$  équivaut à l'égalité de tous les termes des sommes qui les définissent  $X_k = Y_k$  vu que la différence est positive et qu'une somme de termes positifs est nulle si et seulement si ils sont tous nuls). On en déduit que sur  $A$  pour  $n$  assez grand on a  $S_{2^n} = N_t$ . Donc  $S_{2^n} \rightarrow N_t$  p.s. donc en loi.

Or  $Y_k$  est de loi de Bernoulli  $p(t/n)\delta_1 + (1 - p(t/n))\delta_0$ , donc de fonction caractéristique  $p(t/n)(e^{it} - 1) + 1$ . Par indépendance la fonction caractéristique de  $S_n$  est

$$[(1 - e^{-\lambda t/n})(e^{it} - 1) + 1]^n = \exp(n \ln((1 - e^{-\lambda t/n})(e^{it} - 1) + 1)) \xrightarrow{n \rightarrow \infty} \exp(\lambda t(e^{it} - 1))$$

C'est bien la fonction caractéristique d'une loi de Poisson, donc par le théorème de Paul Lévy, la limite en loi  $N_t$  est bien de loi de Poisson de paramètre  $\lambda t$ . □

L'application suivante est plus simple à vérifier en pratique :

**Corollaire 57.** *Un processus de comptage  $(N_t)_{t \in \mathbb{R}^+}$  (c'est-à-dire un processus p.s. croissant à valeur entière partant de  $N_0 = 0$ ) est un processus de Poisson si et seulement si il vérifie les 3 conditions :*

1. *il est à accroissements stationnaires et indépendants,*
2. *il est non trivial (à savoir, non p.s. nul soit  $\exists t > 0 P(N_t > 0) > 0$ ),*
3. *On a  $\lim_{h \rightarrow 0^+} \frac{P(N_h \geq 2)}{h} = 0$ .*

*Démonstration.* exo : cf TD □

**Lemme 58.** *Un processus de Poisson  $(N_t)_{t \in \mathbb{R}^+}$  admet une modification continue à droite qui est un processus de Poisson de même intensité  $Y_t$  : i.e. pour tout  $t$   $P(Y_t = N_t) = 1$ .*

*Plus précisément  $Y_t = N_{t+} := \inf_{h > 0} N_{t+h}$  est cette modification.*

**On ne considérera donc à partir de maintenant que des processus de Poisson p.s. continus à droite (On peut l'ajouter si on veut à la définition).**

*Démonstration.* D'abord,  $N_t \leq Y_t \leq N_{t+h}$ , donc  $\{Y_t \neq N_t\} \subset \{N_{t+h} \neq N_t\} = \{N_{t+h} - N_t \geq 1\}$  donc  $P(Y_t \neq N_t) \leq P(N_{t+h} - N_t \geq 1) \xrightarrow{h \rightarrow 0} 0$  donc  $Y_t = N_t$  p.s. comme voulu. (Attention l'égalité p.s. est à  $t$  fixé et pas pour tout  $t$  à la fois). Voyons que  $Y_t$  est un processus de Poisson. Les

incrémentes  $Y_{t_1}, Y_{t_2} - Y_{t_1}, \dots, Y_{t_n} - Y_{t_{n-1}}$  sont p.s. la limite de  $N_{t_1+h}, N_{t_2+h} - N_{t_1+h}, \dots, N_{t_n+h} - N_{t_{n-1}+h}$  donc la limite en loi de ces variables qui sont indépendantes, en utilisant le théorème de Paul Lévy et la caractérisation de l'indépendance, on a  $Y_{t_1}, Y_{t_2} - Y_{t_1}, \dots, Y_{t_n} - Y_{t_{n-1}}$  indépendant, soit  $Y_t$  à accroissements indépendants et ce sont aussi des variables de Poisson.  $\square$

## 2 Définition en terme d'incrémentes, Construction et Simulation

Comme le processus de Poisson est à sauts simples en nombre fini (et même p.s. continu à droite) dans chaque intervalle, on peut le décrire de deux façons alternatives. D'abord par la donnée des instants de sauts. L'instant du  $n$ -ième saut est donné par  $S_n = \inf\{t : N_t = n\}$ ,  $\{N_t \geq n\} = \{S_n \leq t\}$ .

Réciproquement, étant donné une suite croissante  $S_n$  avec  $\lim_{n \rightarrow \infty} S_n = +\infty$ , on obtient le processus

$$N_t = \sup\{n \geq 0 : S_n \leq t\}$$

Alternativement, on peut aussi considérer la suite  $(T_n)$  des temps entre 2 sauts ("suite des inter-arrivées") de sorte que  $T_n = S_n - S_{n-1}$  ( $S_0 = 0$ ) soit  $S_n = T_1 + \dots + T_n$ .

Le premier résultat pour décrire le processus de Poisson à partir des suites des temps de sauts et des inter-arrivées est le suivant :

**Proposition 59.** Soit  $N_t$  un processus de Poisson d'intensité  $\lambda > 0$ . Soit la suite des inter-arrivées  $T_n = S_n - S_{n-1}$  ( $S_n = \inf\{t \geq 0 : N_t = n\}$ )

1.  $(T_n)$  est formée de variables i.i.d. de loi exponentielle  $\mathcal{E}(\lambda)$  soit  $P_{T_n}(dt) = \lambda e^{-\lambda t} 1_{\{t > 0\}} dt$ .
2.  $S_n$  suit la loi Gamma de paramètre  $(n, \lambda)$  soit  $P_{S_n}(s) = \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s} 1_{\{s > 0\}} ds$
3.  $\mathcal{L}((S_1, \dots, S_n) | N_t = n) = STU(n, [0, t])$  (statistique d'ordre uniforme sur  $[0, t]$ ) de densité :

$$\frac{n!}{t^n} 1_{\{0 < s_1 < \dots < s_n < t\}}(s_1, \dots, s_n).$$

*Exercice 4.* Montrer que  $\mathcal{L}((S_1, \dots, S_n) | S_{n+1} = t) = STU(n, [0, t])$ .

*Démonstration.* On calcule la loi de  $(S_1, \dots, S_n)$ . Soit  $s_i, h_i > 0$  avec  $0 < s_1 < s_1 + h_1 < s_2 < \dots < s_n < s_n + h_n$ . Calculons la probabilité de  $A_n = \{S_1 \in [s_1, s_1 + h_1[, S_2 \in [s_2, s_2 + h_2[, \dots, S_n \in [s_n, s_n + h_n[ \}$ .

On peut réécrire  $A_n$  en terme de  $N_t$ .

$$A_n = \{N_{s_1} = 0, N_{s_1+h_1} = 1 = N_{s_2}, N_{s_2+h_2} = 2 = N_{s_2}, \dots, N_{s_n+h_n} \geq n\}$$

puis en terme des incrémentes indépendants :

$$A_n = \{N_{s_1} = 0, N_{s_1+h_1} - N_{s_1} = 1, N_{s_2} - N_{s_1+h_1} = 0, N_{s_2+h_2} - N_{s_2} = 1, N_{s_3} - N_{s_2+h_2} = 0, \dots, N_{s_n+h_n} - N_{s_n} \geq 1\}.$$

Donc on obtient en utilisant l'indépendance et la loi de Poisson des incrémentes :

$$P(A_n) = e^{-\lambda s_1} e^{-\lambda h_1} \lambda h_1 e^{-\lambda(s_2 - s_1 - h_1)} e^{-\lambda h_2} \lambda h_2 \dots (1 - e^{-\lambda h_n}) = e^{-\lambda s_n} \lambda^{n-1} h_1 \dots h_{n-1} (1 - e^{-\lambda h_n})$$

Si on a une densité  $f$  continue, on a  $\frac{1}{h_1 \dots h_n} P(A_n) = \frac{1}{h_1 \dots h_n} \int_{s_1}^{s_1+h_1} \dots \int_{s_n}^{s_n+h_n} f(x_1, \dots, x_n) \rightarrow f(s_1, \dots, s_n)$ .



En prenant la limite  $(h_1, \dots, h_n) \rightarrow 0$  cela suggère le candidat de loi  $P_{(s_1, \dots, s_n)}(ds_1, \dots, ds_n) = e^{-\lambda s_n} \lambda^n \mathbf{1}_{\{0 < s_1 < \dots < s_n\}} ds_1 \dots ds_n$ . Pour vérifier que c'est bien la loi, on applique le lemme de classe monotone 1, comme les produits d'intervalles  $[s_1, s_1 + h_1] \times [s_2, s_2 + h_2] \dots \times [s_n, s_n + h_n]$ , avec des inégalités comme avant, sont stables par intersections finies dans le simplexe  $\{0 < s_1 < s_2, \dots < s_n\}$  et engendrent la tribu borélienne, il suffit que cette loi donne la bonne valeur sur ces produits, et un calcul d'intégrale montre que c'est le cas. Le changement de variables linéaire  $t_1 + \dots + t_k = s_k$  (de jacobien 1) donne la loi  $P_{(T_1, \dots, T_n)}(dt_1, \dots, dt_n) = e^{-\lambda(t_1 + \dots + t_n)} \lambda^n \mathbf{1}_{\{0 < t_1\}} \dots \mathbf{1}_{\{0 < t_n\}} dt_1 \dots dt_n$  ce qui est la loi de  $n$  variables exponentielles indépendantes.

En intégrant les premières variables, on obtient la densité de  $S_n$  :

$$\int_{\mathbb{R}^{n-1}} ds_1 \dots ds_{n-1} e^{-\lambda s_n} \lambda^n \mathbf{1}_{\{0 < s_1 < \dots < s_n\}} = e^{-\lambda s_n} \frac{\lambda^n s_n^{n-1}}{(n-1)!}$$

(on peut par exemple calculer l'air du pavé et diviser par le nombre de permutation permettant de les ordonner).

Si  $s_n + h_n < t$ , on peut calculer  $P(A_n \cap N_t = n)$  la dernière condition  $N_{s_n + h_n} - N_{s_n} \geq 1$  est remplacée par  $N_{s_n + h_n} - N_{s_n} = 1, N_t - N_{s_n + h_n} = 0$  d'où  $P(A_n \cap N_t = n) = e^{-\lambda t} \lambda^n h_1 \dots h_n$ . Donc vu  $P(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$  on conclut  $P(A_n | N_t = n) = \frac{n!}{t^n} h_1 \dots h_n$ . Le même raisonnement avec le lemme de classe monotone identifie la densité voulue par un calcul d'intégrale évident (pavé pour mesure uniforme).  $\square$

Réciproquement, on a le théorème de construction suivant, qui montre l'existence du processus de Poisson :

**Théorème 60.** (de construction du processus de Poisson) Soit  $T_n$  une suite de variables indépendantes et identiquement distribuées de loi  $\mathcal{E}(\lambda)$  soit  $P_{T_n}(dt) = \lambda e^{-\lambda t} \mathbf{1}_{\{t > 0\}} dt$ . Si on pose

$$N_t = \sup\{n \in \mathbb{N} : T_1 + \dots + T_n \leq t\}$$

alors  $N_t$  est un processus de Poisson d'intensité  $\lambda$ .

*Remarque 11.* Noter que cette construction du processus de Poisson donne un processus p.s. continu à droite car  $\sup\{n \in \mathbb{N} : T_1 + \dots + T_n \leq t\} = \inf_{h > 0} \sup\{n \in \mathbb{N} : T_1 + \dots + T_n \leq t + h\}$ . En effet,  $\leq$  est évident car  $\sup\{n \in \mathbb{N} : T_1 + \dots + T_n \leq t\} \leq \sup\{n \in \mathbb{N} : T_1 + \dots + T_n \leq t + h\}$  par inclusion des ensembles donc on peut passer à l'inf.  $n = \inf_{h > 0} \sup\{n \in \mathbb{N} : T_1 + \dots + T_n \leq t + h\}$ ,  $T_1 + \dots + T_n \leq t + h$  et en prenant l'inf sur  $h$   $T_1 + \dots + T_n \leq t$ , donc  $n \leq \sup\{n \in \mathbb{N} : T_1 + \dots + T_n \leq t\}$ .

*Démonstration.* Par construction le processus  $N_t$  est un processus de comptage. Il est p.s. à sauts simples car un saut double en  $t$  est l'évènement  $\cup_i \{T_1 + \dots + T_i = t, T_{i+1} = 0\}$  mais comme la densité de chaque  $T_i$  est continue  $P(T_{i+1} = 0) = 0$ .

On calcule la loi de  $N_t$  en utilisant  $\{N_t \geq n\} = \{S_n = T_1 + \dots + T_n \leq t\}$  donc en utilisant la distribution  $\Gamma$  obtenue pour la somme de variables exponentielles indépendantes à la précédente proposition. On obtient la valeur souhaitée pour une loi de Poisson :

$$P(N_t = n) = P(S_n \leq t) - P(S_{n+1} \leq t) = \int_0^t \left[ \frac{\lambda^n s^{n-1}}{(n-1)!} - \frac{\lambda^{n+1} s^n}{n!} \right] e^{-\lambda s} ds = \int_0^t d \left[ \frac{\lambda^n s^n}{n!} e^{-\lambda s} \right] = \frac{\lambda^n t^n}{n!} e^{-\lambda t}.$$

On considère ensuite des incréments pour les temps  $0 = s_0 < s_1 < \dots < s_n$  et des entiers  $j_k \in \mathbb{N}$  dont on nomme les sommes  $J_k = j_1 + \dots + j_k$ .

On doit décrire la loi des incréments de  $N$  à partir des  $S$ , on décrit l'évènement consistant à avoir  $j_k$  sauts entre  $t_{k-1}$  et  $t_k$  de deux façons :

$$\bigcap_{k=1}^n \{N_{s_k} - N_{s_{k-1}} = j_k\} = \{S_{1+J_n} > s_n\} \cap \bigcap_{k=1}^n \{s_{k-1} < S_{J_{k-1}+1} < \dots < S_{J_k} \leq s_k\}$$

En utilisant le théorème de transfert et en posant  $D = \prod_{k \in [1, n], j_k \neq 0} \{s_{k-1} < u_{J_{k-1}+1} < \dots < u_{J_k} \leq s_k\} \times \{u_{1+J_n} > s_n\}$ , on obtient :

$$P\left(\bigcap_{k=1}^n \{N_{s_k} - N_{s_{k-1}} = j_k\}\right) = \int dP_{(S_1, \dots, S_{J_n+1})}(u_1, \dots, u_{1+J_n}) 1_D = \lambda^{J_n+1} \int_D e^{-\lambda u_{1+J_n}} du_1 \dots du_{1+J_n},$$

en utilisant que la loi des  $S_i$  est obtenue comme à la proposition précédente. Par le théorème de Fubini-Tonelli, il s'agit d'un produit :

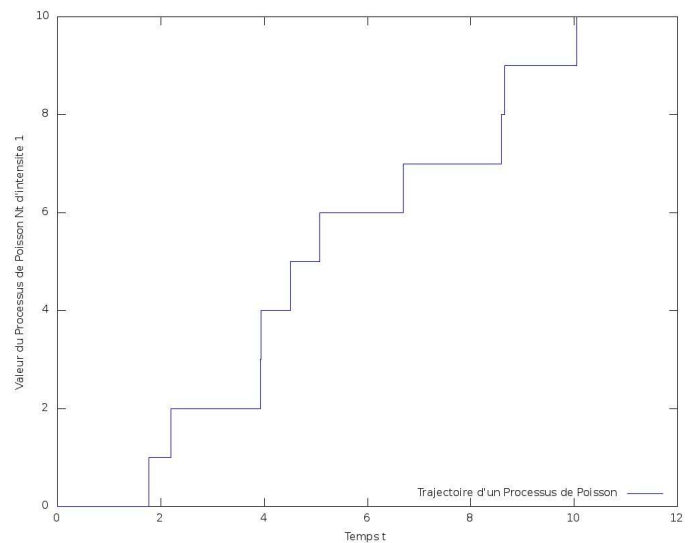
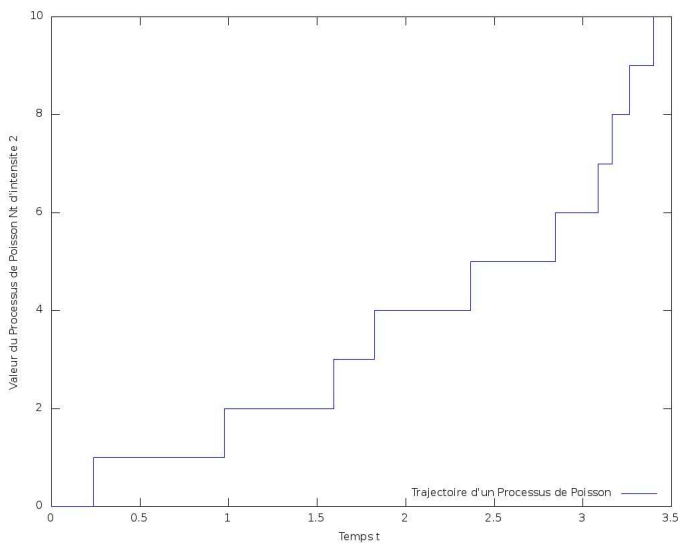
$$P\left(\bigcap_{k=1}^n \{N_{s_k} - N_{s_{k-1}} = j_k\}\right) = \prod_{k \in [1, n], j_k \neq 0} \lambda^{j_k} \int_{\{s_{k-1} < z_1 < \dots < z_{j_k} \leq s_k\}} dz_1 \dots dz_{j_k} \times \lambda \int_{z > s_n} e^{-\lambda z} dz.$$

Donc en calculant les intégrales de simplexes on obtient la valeur voulue :

$$P\left(\bigcap_{k=1}^n \{N_{s_k} - N_{s_{k-1}} = j_k\}\right) = \prod_{k \in [1, n], j_k \neq 0} \lambda^{j_k} \frac{(s_k - s_{k-1})^{j_k}}{j_k!} \times e^{-\lambda s_n}.$$

□

A partir de cette construction, la simulation est évidente, il suffit de simuler les temps inter-arrivées indépendants de lois exponentielles  $T_n = -\log(U_n)/\lambda$  avec  $U_n$  obtenu par un générateur de nombres aléatoires uniformes sur  $[0, 1]$ . ( à gauche intensité 2, à droite intensité 1, on voit que les temps entre sauts d'espérance  $1/2$ , dans le premier cas, sont plus courts) :



### 3 Propriété de Markov

La propriété de Markov simple est une conséquence simple de la définition :

**Proposition 61** (Markov simple). *Soit  $(N_t)_{t \geq 0}$  un processus de Poisson d'intensité  $\lambda$ , alors pour tout  $u \geq 0$ ,  $(N_{t+u} - N_u)_{t \geq 0}$  est un processus de Poisson d'intensité  $\lambda$ , indépendant de  $(N_s)_{0 \leq s \leq u}$ .*

*Démonstration.* Les incréments du nouveau processus  $N_t^u = N_{t+u} - N_u$  sont encore des incréments de l'ancien, donc ils suivent la loi de Poisson et sont indépendants (si correspondant à des intervalles disjoints). Comme c'est bien sûr un processus de comptage, on obtient que c'est un processus de Poisson, l'indépendance vient de celle des incréments puisque la tribu engendrée par  $N_t^u$  est engendrée par des incréments comme celle de  $(N_s)_{0 \leq s \leq u}$ . Par regroupement par paquet, on obtient l'indépendance voulue.  $\square$

**Définition 30.** Un temps d'arrêt par rapport à la tribu du processus de Poisson  $\mathcal{F}_t = \sigma(N_s, s \leq t)$  est une variable aléatoire  $T : \Omega \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  vérifiant  $\{T \leq t\} \in \mathcal{F}_t$  pour tout  $t \in \mathbb{R}_+$ .

**Proposition 62** (Propriétés de Markov forte). *Soit  $(N_t)_{t \geq 0}$  un processus de Poisson d'intensité  $\lambda$  (p.s. continue à droite), alors pour tout temps d'arrêt  $T$  conditionnellement sur  $\{T < +\infty\}$ , le processus  $(N_{T+h} - N_T)_{h \geq 0}$  est un processus de Poisson d'intensité  $\lambda$  indépendant de  $T$  et  $(N_{T \wedge s})_{s \geq 0}$ .*

*Démonstration.* On note comme avant  $T_n$  les temps inter-arrivées de  $N$ .

**1/ Réduction au cas où  $T$  prend un nombre dénombrable de valeurs finies**

On doit montrer que pour  $h$ -  $\sigma(T, (N_{T \wedge s})_{s \geq 0})$ -mesurable,  $f$  mesurable bornée,  $h_1, \dots, h_m$  :

$$E(f(N_{T+h_1} - N_T, \dots, N_{T+h_m} - N_T)h) = E(f(N_{h_1}, \dots, N_{h_m}))E(h).$$

Il suffit de montrer cela pour  $h$   $\sigma(T, N_{T \wedge s_1}, \dots, N_{T \wedge s_k})$ -mesurable bornée. En effet si c'est le cas, on a le résultat pour tout  $h = 1_A$  dans cette tribu pour tout  $s_1, \dots, s_k$  et c'est une classe stable par intersection finie qui engendre  $\sigma(T, (N_{T \wedge s})_{s \geq 0})$ , donc le lemme de classe monotone 1 conclura à l'égalité des deux côtés pour tout  $h = 1_A$  et ensuite par linéarité et en approchant par des fonctions étagées, pour tout  $h$  (cf demo du théorème de transfert ou du lemme de Doob-Dynkin). On fixe donc  $h$   $\sigma(T, N_{T \wedge s_1}, \dots, N_{T \wedge s_k})$ -mesurable bornée, par le lemme de Doob-Dynkin 4, il s'écrit  $h = g(T, N_{T \wedge s_1}, \dots, N_{T \wedge s_k})$  avec  $g$  mesurable bornée.

On pose

$$T^{(n)} = \sum_{k \in \mathbb{N}} \frac{k}{2^n} 1_{\{2^n T \in ]k-1, k]\}} \geq T^{(n+1)} \geq T.$$

C'est une suite décroissante de temps d'arrêts qui converge vers  $T$ , p.s.. En effet, la limite  $T$  est par construction et théorème des gendarmes vu  $T^{(n)} \leq T + \frac{1}{2^n}$ . De plus,  $T^{(n)}$  est bien un temps d'arrêt car  $\{T^{(n)} \leq t\} = \cup_k \{T^{(n)} \leq t, 2^n T \in ]k-1, k]\} = \cup_k \{k \leq t2^n, 2^n T \in ]k-1, k]\} = \cup_{k \leq t2^n} \{2^n T \in ]k-1, k]\}$  et chaque terme est dans  $\mathcal{F}_{k/2^n} \subset \mathcal{F}_t$ .

Comme  $N$  est p.s. continu à droite, on déduit :  $N_{T^{(n)} \wedge s_1} \rightarrow N_{T \wedge s_1}$  p.s. et  $N_{T^{(n)}+h_m} - N_{T^{(n)}} \rightarrow N_{T+h_m} - N_T$  p.s. Comme les fonctions ci dessous sont mesurables bornées, on déduit du TCD (dominé par des constantes) que :

$$\begin{aligned} & E(f(N_{T^{(n)}+h_1} - N_{T^{(n)}}), \dots, N_{T^{(n)}+h_m} - N_{T^{(n)}})g(T^{(n)}, N_{T^{(n)} \wedge s_1}, \dots, N_{T^{(n)} \wedge s_k})) \\ &= E(f(N_{h_1}, \dots, N_{h_m}))E(g(T^{(n)}, N_{T^{(n)} \wedge s_1}, \dots, N_{T^{(n)} \wedge s_k})) \\ &\rightarrow_{n \rightarrow \infty} E(f(N_{T+h_1} - N_T, \dots, N_{T+h_m} - N_T)g(T, N_{T \wedge s_1}, \dots, N_{T \wedge s_k})) \\ &= E(f(N_{h_1}, \dots, N_{h_m}))E(g(T, N_{T \wedge s_1}, \dots, N_{T \wedge s_k})). \end{aligned}$$

**2/ Cas où  $T$  prend un nombre dénombrable de valeurs** Soit  $Y_h = N_{T+h} - N_T$ .  $Y_h$  est un processus de comptage p.s. à sauts simples avec un nombre infini de sauts, il est donc décrit par une suite de temps de sauts  $R_n = \inf\{t : Y_t = n\} - \inf\{t : Y_t = n-1\}$ ,  $n \geq 1$ ,  $\{Y_t \geq n\} = \{R_1 + \dots + R_n \leq t\}$ . La loi de la suite détermine donc la loi de  $Y$ . On doit montrer que pour  $h = f(T, N_{T \wedge s_1}, \dots, N_{T \wedge s_k})$  avec  $f$  mesurable bornée.

Comme  $\{T < +\infty\} = \cup_{n \geq 0} \{T_n \leq T < T_{n+1}\}$  et donc en décomposant aussi  $\Omega = \cup_{a \in T(\Omega)} \{T = a\}$ , par  $\sigma$ -additivité des deux côtés, il suffit de voir (pour  $r_i > 0$ ),  $a \in T(\Omega)$  :

$$E(1_{\{R_1 \geq r_1, \dots, R_m \geq r_m, T_n \leq T = a < T_{n+1}\}} h) = P(T_1 \geq r_1, \dots, T_m \geq r_m) E(1_{\{T_n \leq T = a < T_{n+1}\}} h).$$

Or si  $T_n \leq T < T_{n+1}$ , on a  $R_1 = T_{n+1} - T$ ,  $T_{n+m} = R_m$ ,  $m \geq 2$ , :

$$E(1_{\{R_1 \geq r_1, \dots, R_m \geq r_m, T_n \leq T = a < T_{n+1}\}} h) = E(1_{\{T_{n+1} \geq a+r_1, T_{n+2} \geq r_2, \dots, T_{n+m} \geq r_m, T_n \leq T = a\}} f(a, N_{a \wedge s_1}, \dots, N_{a \wedge s_k}))$$

Or les  $T_n$  sont indépendants et i.i.d. de loi exponentielle et si  $T_n \leq a < T_{n+1}$ ,  $N_a = n$ , donc  $N_{a \wedge s_i} = \sup\{k \leq n, T_1 + \dots + T_k \leq a \wedge s_i\}$  est  $\sigma(T_1, \dots, T_n)$  mesurable. De même  $\{T = a\}$  est  $\mathcal{F}_a = \sigma(N_{a \wedge s}, s \geq 0)$ -mesurable donc  $\sigma(T_1, \dots, T_n)$  mesurable. En utilisant l'indépendance et la distribution, on obtient :

$$E(1_{\{R_1 \geq r_1, \dots, R_m \geq r_m, T_n \leq T = a < T_{n+1}\}} h) = e^{-\lambda(a+r_1)} P(T_2 \geq r_2, \dots, T_m \geq r_m) E(1_{\{T_n \leq T = a\}} f(a, N_{a \wedge s_1}, \dots, N_{a \wedge s_k}))$$

ce qui vaut aussi en utilisant les lois exponentielles et la même indépendance :

$$\begin{aligned} & E(1_{\{R_1 \geq r_1, \dots, R_m \geq r_m, T_n \leq T = a < T_{n+1}\}} h) \\ &= P(T_{n+1} \geq a) P(T_1 \geq r_1, T_2 \geq r_2, \dots, T_m \geq r_m) E(1_{\{T_n \leq T = a\}} f(a, N_{a \wedge s_1}, \dots, N_{a \wedge s_k})) \\ &= P(T_1 \geq r_1, T_2 \geq r_2, \dots, T_m \geq r_m) E(1_{\{T_n \leq T = a \leq T_{n+1}\}} f(a, N_{a \wedge s_1}, \dots, N_{a \wedge s_k})) \end{aligned}$$

ce qui est le résultat voulu. □

## 4 Exemples de généralisations

Si on veut utiliser le processus de Poisson comme modèle, les hypothèses caractéristiques peuvent sembler restrictives.

1. La stationnarité peut être remplacée en remplaçant dans la définition les variables de Poisson de paramètre  $\lambda t$  constant par ceux de paramètre  $N_t - N_s \sim \mathcal{P}(\int_s^t \lambda(u) du)$  pour une fonction intensité instantanée  $\lambda(u) > 0$ . C'est le *processus de Poisson inhomogène*.
2. On peut voir le processus de Poisson comme un processus ponctuel, c'est à dire une variable aléatoire à valeur mesure :  $\omega \mapsto \Lambda(\omega) = \sum_{i=1}^{\infty} \delta_{S_i(\omega)}$  de sorte que  $N_t = \Lambda([0, t])$  est la fonction de répartition aléatoire de cette mesure. C'est le moyen de généraliser le processus de Poisson en dimension supérieure (comme mesure aléatoire ou fonction de répartition aléatoire).
3. Si on remplace les sauts exponentiels par des temps d'attente ayant une autre loi, on obtient un *processus de renouvellement* qui peut modéliser des durées de vie aléatoires (par exemple pour des machines) etc.
4. Si on remplace les sauts toujours de +1 dans le processus de Poisson, par des sauts quelconques sur l'espace d'états  $\mathbb{N}$  selon une chaîne de Markov, on obtient un processus de Markov à temps continu (et espace d'état discret).

En général, les généralisations précédentes n'ont pas des incréments indépendants. Nous nous concentrons sur une généralisation à incréments indépendants. C'est un cas particulier de Processus de Lévy, dont le mouvement brownien sera un autre exemple continu.

## 4.1 Processus de Poisson composé

On fixe  $N_t$  un processus de Poisson d'intensité  $\lambda > 0$  et  $Y_n$  une suite de variables aléatoires i.i.d, indépendantes de  $N$ . Le processus :

$$Y(t) = \sum_{k=1}^{N_t} Y_k$$

est appelé *processus de Poisson composé*.

On calcule sa fonction caractéristique :

**Proposition 63.** *Le processus  $Y$  est un processus stationnaire à accroissements indépendants de loi donnée par la fonction caractéristique :*

$$\Phi_{Y(t)}(x) = e^{\lambda t(\Phi_{Y_1}(x)-1)}.$$

*Démonstration.* On note juste  $\Phi_Y = \Phi_{Y_1}$ . Soit  $n_0 = t_0 = 0 \leq t_1 \leq \dots \leq t_n$

$$E\left(e^{i \sum_{k=1}^m x_k (Y(t_k) - Y(t_{k-1}))}\right) = \sum_{n_1 \leq \dots \leq n_k} \mathbf{E}\left(e^{i \sum_{k=1}^m x_k (\sum_{l=n_{k-1}+1}^{n_k} Y_l)} \mid N_{t_1} = n_1, \dots, N_{t_k} = n_m\right) P(\forall i \leq m, N_{t_i} = n_i)$$

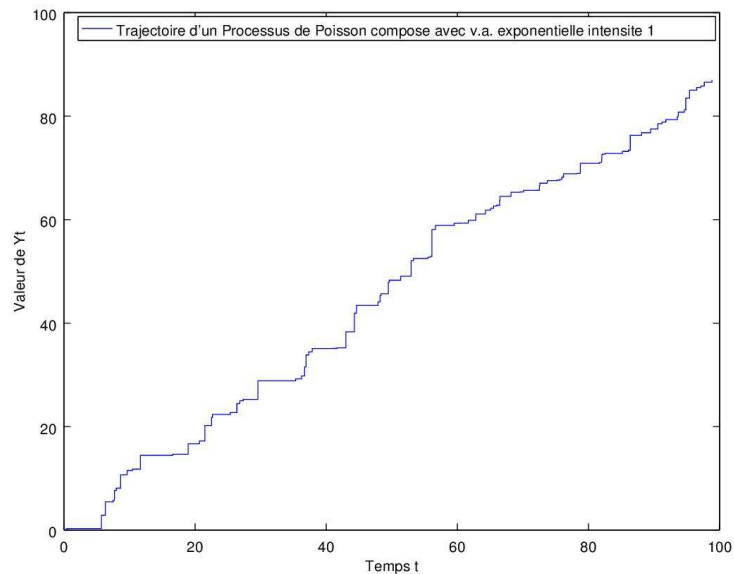
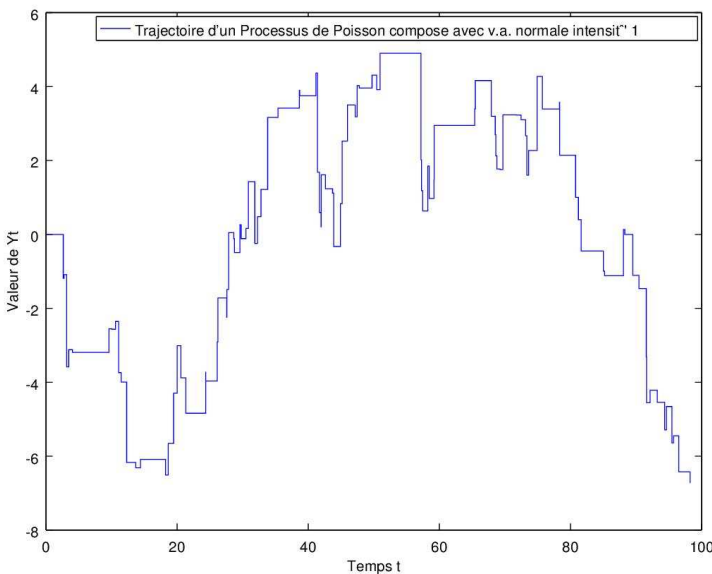
et par indépendance des  $Y_l$  et de  $N$ , c'est égal à :

$$\sum_{n_1 \leq \dots \leq n_k} \mathbf{E}\left(e^{i \sum_{k=1}^m x_k (\sum_{l=n_{k-1}+1}^{n_k} Y_l)}\right) P(\forall i \leq m, N_{t_i} = n_i) = \sum_{n_1 \leq \dots \leq n_k} \prod_{k=1}^m (\Phi_Y(x_k))^{n_k - n_{k-1}} e^{-\lambda t_m} \prod_{i \leq m} \frac{(\lambda(t_i - t_{i-1}))^{n_i - n_{i-1}}}{(n_i - n_{i-1})!}$$

ce qui donne en sommant les séries de l'exponentielle :

$$E\left(e^{i \sum_{k=1}^m x_k (Y(t_k) - Y(t_{k-1}))}\right) = \prod_{k \leq m} e^{(\Phi_Y(x_k) - 1) \lambda (t_i - t_{i-1})}.$$

Ceci donne l'indépendance (forme produit) la stationnarité et la loi. □



## 5 Processus de Poisson et limites

### 5.1 Processus de Poisson comme Limite du Processus de Bernoulli

On rappelle le processus de Bernoulli de l'exemple 27 de paramètre  $p$ . C'est la chaîne de Markov sur  $\mathbb{N}$  de matrice de transition  $P_{n,n} = (1 - p), P_{n,n+1} = p$ .

**Proposition 64.** *Soit  $\lambda > 0$  et  $p_k \in ]0, 1[$  une suite de nombres telles que  $kp_k \rightarrow_{k \rightarrow \infty} \lambda$ . Soit  $B_n^k$  un processus de Bernoulli de paramètre  $p_k$ , et  $N_t$  un processus de Poisson d'intensité  $\lambda$ . Alors pour tout  $0 \leq t_1 \leq \dots \leq t_n$ ,  $(B_{[t_1 k]}^k, \dots, B_{[t_n k]}^k)$  converge en loi pour  $k \rightarrow \infty$  vers  $(N_{t_1}, \dots, N_{t_n})$*

*Démonstration.* D'abord, par changement de variable, il suffit de montrer que les incréments  $(B_{[t_1 k]}^k, B_{[t_2 k]}^k - B_{[t_1 k]}^k, \dots, B_{[t_n k]}^k - B_{[t_{n-1} k]}^k)$  convergent en loi vers  $(N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}})$ .

Par la construction de la proposition 44,  $B_t^k = \sum_{n=1}^l X_n$  avec  $X_l$  i.i.d Bernoulli  $\mathcal{B}(p_k)$ . Les incréments sont donc indépendants comme à la limite, et il suffit, par le théorème de Paul Lévy, de voir que  $\Phi_{B_{[t_n k]}^k - B_{[t_{n-1} k]}^k}(\xi) \rightarrow \Phi_{N_{t_n} - N_{t_{n-1}}}(\xi)$ .

Mais par indépendance

$$\begin{aligned} \Phi_{B_{[t_n k]}^k - B_{[t_{n-1} k]}^k}(\xi) &= \Phi_{X_1}(\xi)^{[t_n k] - [t_{n-1} k]} \\ &= \exp(\ln(1 + (e^{i\xi} - 1)p_k)([t_n k] - [t_{n-1} k])) \rightarrow \exp((e^{i\xi} - 1)(t_n - t_{n-1})\lambda), \end{aligned}$$

car par encadrement  $p_k([t_n k] - [t_{n-1} k]) \in [p_k(t_n - t_{n-1})k - 2p_k, p_k(t_n - t_{n-1}) + 2p_k]$  et les deux bornes tendent vers  $(t_n - t_{n-1})\lambda$ .  $\square$

### 5.2 Cas limite du processus de Poisson

On peut aussi appliquer la loi des grands nombres et le TC au processus de Poisson pour retrouver  $\lambda$  et construire un mouvement brownien. Comme on va le voir, ce sera un processus à incréments indépendants, donc il suffira de voir la limite des lois unidimensionnelles :

**Proposition 65.** *Soit  $N_t$  un processus de Poisson d'intensité  $\lambda$  alors  $N_t/t \rightarrow \lambda$  p.s. et  $\sqrt{t}(N_t/t - \lambda) \rightarrow^{\mathcal{L}} \mathcal{N}(0, \lambda)$ .*

*Démonstration.* Il suffit d'appliquer LGN et TCL aux incréments indépendants en remarquant que espérance et variance sont  $\lambda$  pour un loi de Poisson de paramètre  $\lambda$ .  $\square$

# Chapitre 6

## Introduction au mouvement Brownien : Application au modèle de Black et Scholes

### 1 Définition et propriétés fondamentales du mouvement Brownien

**Définition 31.** Un processus gaussien centré  $(B_t)_{t \in \mathbb{R}^+}$ ,  $B_t : (\Omega, \mathcal{T}, P) \rightarrow \mathbb{R}$  est un mouvement brownien (standard) si il existe  $A \in \mathcal{T}$  avec  $P(A) = 1$  tel que pour tout  $\omega \in A$ ,  $t \mapsto B_t(\omega)$  est continue et si il est de covariance  $\mathbf{E}(B_s B_t) = \min(s, t)$ .

*Remarque 12.* On rappelle que, pour vérifier qu'un processus est gaussien, il faut vérifier que  $(B_{t_1}, \dots, B_{t_n})$  est un vecteur gaussien pour tout  $(t_1, \dots, t_n)$ . Pour vérifier qu'il est centré, on doit vérifier  $\mathbf{E}(B_t) = 0$ . Il n'est pas évident que le mouvement brownien existe (c'est-à-dire que l'on peut construire le processus gaussien ci-dessus avec toutes ses trajectoires continues. On l'admettra.

Par ailleurs, certains appellent  $x + \mu t + \sigma B_t$  aussi mouvement brownien (non standard), partant de  $x$ , avec dérive  $\mu$  et variance  $\sigma t$ . Nous n'utiliserons que des mouvements browniens standards.

**Théorème 66.** *Un processus  $(B_t)_{t \in \mathbb{R}^+}$  est un mouvement brownien si et seulement si il vérifie*

- (continuité p.s.) *Il existe  $A \in \mathcal{T}$  avec  $P(A) = 1$  tel que pour tout  $\omega \in A$ ,  $t \mapsto B_t(\omega)$  est continue*
- (accroissements indépendants) *Pour tout  $n$  et tout  $t_1 < \dots < t_n$   $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$  sont indépendants*
- (Accroissements gaussiens stationnaires) *Pour tout  $s < t$   $B_t - B_s \sim \mathcal{N}(0, t - s)$ .*

PREUVE : Si  $B_t$  vérifie les hypothèses,  $(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}})$  est un vecteur gaussien (suite de variables indépendantes gaussiennes), donc par transformation linéaire  $(B_{t_1}, B_{t_2}, \dots, B_{t_n})$  est un vecteur gaussien, donc  $(B_t)$  est un processus gaussien (centré car chaque  $B_t$  est supposé centré). Comme en plus, il est continu, c'est un mouvement brownien si il a la bonne covariance, et en effet, si  $s < t$

$$\mathbf{E}(B_t B_s) = \mathbf{E}((B_t - B_s)B_s) + \mathbf{E}(B_s^2) = \mathbf{E}(B_t - B_s)\mathbf{E}(B_s) + s = s = \min(s, t)$$

en utilisant la variance et l'indépendance. Par échange de  $s$  et  $t$ , on a le cas général.

Réciproquement, si  $(B_t)$  est un mouvement brownien, il est supposé p.s. continu et chaque  $(B_{t_1}, B_{t_2}, \dots, B_{t_n})$  est un vecteur gaussien donc par combinaison linéaire  $(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}})$  est un vecteur gaussien, et il suffit de calculer les covariances nulles pour obtenir l'indépendance.

Or pour  $i < n$ , on a

$$\begin{aligned}\mathbf{E}((B_{t_i} - B_{t_{i-1}})(B_{t_n} - B_{t_{n-1}})) &= \mathbf{E}(B_{t_i}B_{t_n} - B_{t_{i-1}}B_{t_n} - B_{t_i}B_{t_{n-1}} + B_{t_{i-1}}B_{t_{n-1}}) \\ &= \min(t_i, t_n) - \min(t_{i-1}, t_n) - \min(t_i, t_{n-1}) + \min(t_{i-1}, t_{n-1}) \\ &= t_i - t_{i-1} - t_i + t_{i-1} = 0\end{aligned}$$

De même  $B_t - B_s$  est une variable gaussienne et il ne reste qu'à calculer la variance pour  $t > s$ .

$$\begin{aligned}\mathbf{E}((B_t - B_s)^2) &= \mathbf{E}(B_t^2 - 2B_sB_t + B_s^2) \\ &= t - 2s + s = (t - s).\end{aligned}$$

On a donc obtenu toutes les propriétés. ■

**Proposition 67.** (*Invariance*) Si  $B_t$  est un mouvement brownien, alors pour  $\lambda > 0$ ,  $(\lambda^{-1}B_{\lambda^2 t})_{t \geq 0}$  et  $(B_t - B_\lambda)_{t \geq \lambda}$  sont des mouvements browniens.

PREUVE : La continuité est évidente et par combinaison linéaire on a bien des processus gaussiens centrés, il suffit donc de vérifier la covariance pour le premier  $\mathbf{E}(\lambda^{-1}B_{\lambda^2 t}\lambda^{-1}B_{\lambda^2 s}) = \lambda^{-2} \min(\lambda^2 t, \lambda^2 s) = \min(t, s)$ . Pour le deuxième on utilise le théorème, comme les incréments  $(B_t - B_\lambda)_{t \geq \lambda}$  sont des incréments de  $B_t$  on trouve l'indépendance et la stationnarité, comme voulu. ■

## 1.1 Limite de marches aléatoires

Soit  $S_n = X_1 + \dots + X_n$  avec  $X_i$  i.i.d,  $\mathbf{E}(X_1) = 0$ ,  $\mathbf{E}(X_1^2) = 1$ . On s'intéresse au chemin  $k \rightarrow S_k$  et on introduit pour cela une normalisation similaire au TCL pour se ramener à  $[0, 1]$ .

$$S_t^n = \frac{1}{\sqrt{n}} S_{[nt]}, t \in [0, 1].$$

$S_1^n$  est la variable du TCL qui tend vers une variable gaussienne. Le processus tend vers un processus gaussien qui n'est rien d'autre que le mouvement brownien (en loi fini-dimensionnelle) :

**Théorème 68.** (*Donsker*) Pour tout  $(t_1, \dots, t_m)$ ,  $(S_{t_1}^n, \dots, S_{t_m}^n)$  converge en loi vers  $(B_{t_1}, \dots, B_{t_m})$  avec  $(B_t)_{t > 0}$  un mouvement brownien.

En fait le meilleur théorème de Donsker établit qu'une variante continue linéaire par morceau converge en un sens plus fort (en loi sur l'espace des fonctions continues).

PREUVE : Il est équivalent de considérer la convergence en loi des incréments (changement de variables linéaire) soit pour  $t_1 < \dots < t_m$   $(S_{t_1}^n, S_{t_2}^n - S_{t_1}^n, \dots, S_{t_m}^n - S_{t_{m-1}}^n)$ . Or par regroupement par paquets ces variables sont indépendantes car  $S_{t_m}^n - S_{t_{m-1}}^n \in \sigma(X_i, i \in [[t_{m-1}n] + 1, [t_m n]])$  qui donnent des familles disjointes d'indices donc indépendance. De plus  $S_{t_m}^n - S_{t_{m-1}}^n$  a même loi que  $\frac{\sqrt{[t_m n] - [t_{m-1} n]}}{\sqrt{n}} S_1^{([t_m n] - [t_{m-1} n])}$  qui converge donc en loi par le TCL vers une  $\mathcal{N}(0, t_m - t_{m-1})$ . En utilisant l'indépendance (et par exemple les fonctions caractéristiques), on obtient que  $(S_{t_1}^n, S_{t_2}^n - S_{t_1}^n, \dots, S_{t_m}^n - S_{t_{m-1}}^n)$  converge en loi vers des variables indépendantes ayant même loi que le mouvement brownien. ■



## 1.2 Quelques faits géométriques [facultatifs]

Le théorème suivant admis donne une idée de l'allure des trajectoires du mouvement brownien

**Théorème 69.** Soit  $(B_t)_{t>0}$  un mouvement brownien, alors

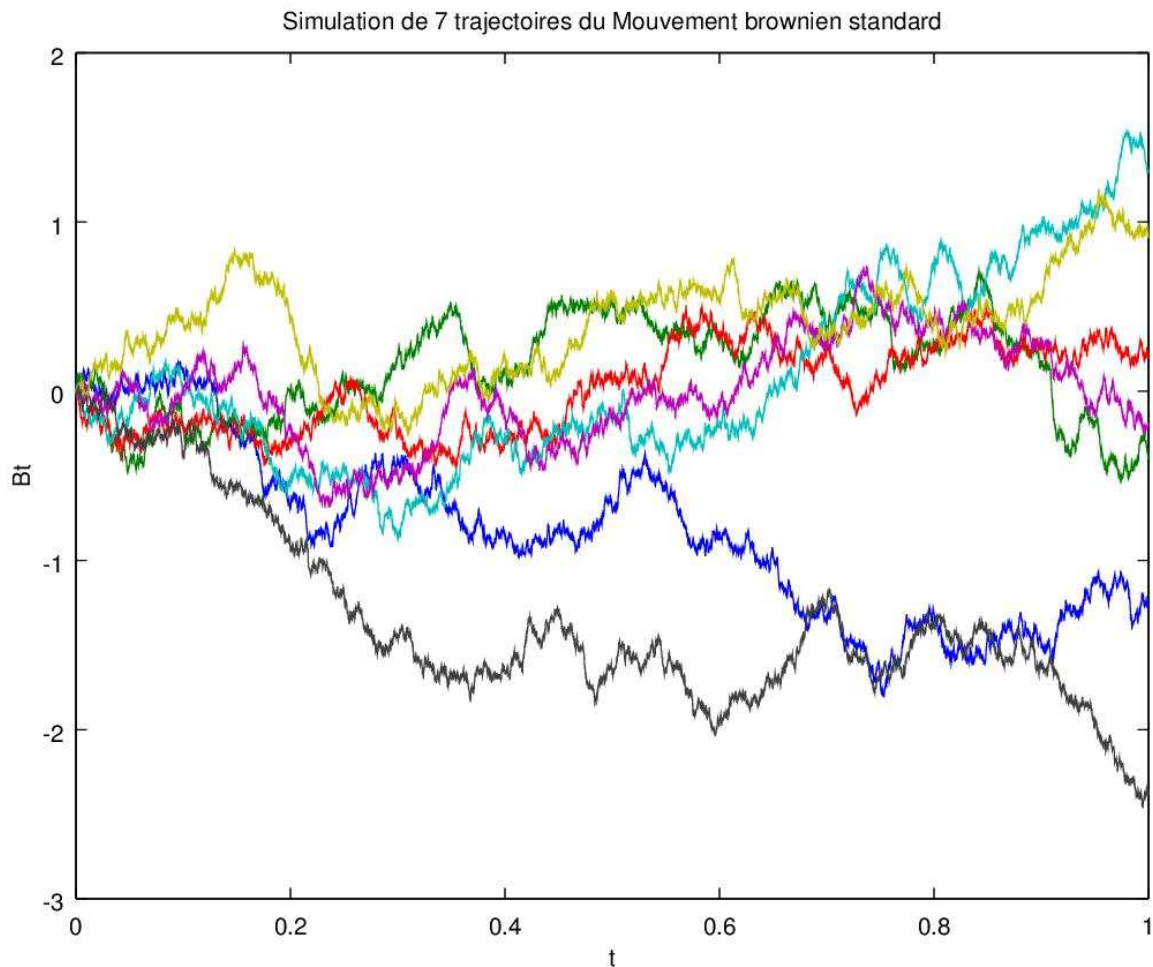
1. Pour tout  $\epsilon > 0$ , p.s.  $\sup_{s \in ]0, \epsilon[} B_s > 0$  et  $\inf_{s \in ]0, \epsilon[} B_s < 0$ ,
2.  $\sup_{s \in ]0, t[} B_s$  a même loi que  $|B_t|$ ,
3. On a p.s.  $\limsup_{s \rightarrow \infty} B_s = +\infty$  et  $\liminf_{s \rightarrow \infty} B_s = -\infty$
4. L'ensemble des trajectoires nulle part dérivables contient un évènement de probabilité 1.

## 1.3 Simulation du mouvement brownien

Une première méthode pour simuler le mouvement brownien, soit une discrétisation de celui-ci de la forme  $(B_{t/n}, B_{2t/n}, \dots, B_t)$  est d'utiliser la méthode de simulation des vecteurs gaussiens de la section 2 du chapitre 2. Mais avec le calcul de la racine carrée matricielle, ce n'est pas très efficace. L'avantage est que cela fonctionne pour tout processus gaussien.

Une méthode plus efficace vient du théorème 66. Il suffit de simuler  $(Z_1, \dots, Z_n)$   $n$  variables gaussiennes standard i.i.d et de considérer

$$\sqrt{t/n}(Z_1, Z_1 + Z_2, \dots, Z_1 + \dots + Z_n) \sim (B_{t/n}, B_{2t/n}, \dots, B_t).$$



## 2 Modèle de Black et Scholes

**Définition 32.** Un *mouvement Brownien géométrique* de taux d'intérêt  $\mu$  et de volatilité  $\sigma$  est le processus  $(S_t)_{t>0}$  donné par la formule :

$$S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right),$$

avec  $(B_t)_{t>0}$  un mouvement brownien.

C'est le modèle d'action proposé par Black et Scholes.

*Remarque 13.* Le modèle de Black et Scholes n'est qu'un modèle financier parmi d'autres, particulièrement simple. On pourrait facilement intégrer un taux d'intérêt et une volatilité dépendant du temps pour améliorer le modèle. Plus fondamentalement, on lui reproche souvent que les trajectoires du mouvement brownien géométrique sont continues alors que les marchés financiers ont des sauts (nécessité de processus de Poisson/Lévy). On reproche aussi dans la même ligne que les trajectoires du mouvement brownien sont trop régulières pour représenter qualitativement les marchés financiers.

La caractérisation du mouvement brownien permet de trouver une propriété caractéristique plus parlante probabilistiquement :

**Proposition 70.** *Un processus positif  $(S_t)_{t \in \mathbb{R}^+}$  est un mouvement brownien géométrique si et seulement si il vérifie*

- (continuité p.s.) Il existe  $A \in \mathcal{T}$  avec  $P(A) = 1$  tel que pour tout  $\omega \in A$ ,  $t \mapsto S_t(\omega)$  est continue
- (accroissements relatifs indépendants) Pour tout  $u < t$ ,  $S_t/S_u$  est indépendant de la tribu  $\sigma(S_v; v \leq u)$ .
- (Accroissements stationnaires log-normaux) Pour tout  $u < t$ ,  $S_t/S_u$  est de loi log normale  $\mathcal{LN}\left(\left(\mu - \frac{\sigma^2}{2}\right)(t-u), \sigma^2(t-u)\right)$ , c'est-à-dire de loi :

$$P_{S_t/S_u}(ds) = \frac{1}{s\sigma\sqrt{2\pi(t-u)}} \exp\left(-\frac{\left(\ln(s) - \left(\mu - \frac{\sigma^2}{2}\right)(t-u)\right)^2}{2\sigma^2(t-u)}\right) 1_{]0, \infty[}(s) ds.$$

PREUVE : Il est facile de voir que les hypothèses se traduisent pour  $\frac{1}{\sigma} \left(\ln(S_t/S_0) - \left(\mu - \frac{\sigma^2}{2}\right)t\right)$  sur les hypothèses du théorème 66 et donc qu'il s'agit d'un mouvement brownien (standard)  $B_t$ . ■

Le théorème suivant va être utilisé pour calculer le prix des options associées à ce modèle. Le modèle de Black et Scholes suppose donné un actif sans risque de prix à l'instant  $t$  de taux d'intérêt (instantané)  $r$  :

$$S_t^0 = \exp(rt).$$

On a le cas particulier suivant du théorème de Girsanov (de changement de mesure des processus stochastiques) :

**Théorème 71** (Girsanov avec dérive constante). *Soit  $(B_t)_{t \in [0, T]}$  un mouvement brownien (standard) dans  $(\Omega, \mathcal{T}, P)$ . Soit  $\mathcal{F}_T = \sigma(B_t, t \leq T)$ . Alors la formule suivante définit une mesure de probabilité sur  $(\Omega, \mathcal{F}_T)$*

$$\tilde{P}(A) = \mathbf{E}\left(1_A \exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_T - \frac{T}{2} \left(\frac{\mu-r}{\sigma}\right)^2\right)\right).$$

Dans  $(\Omega, \mathcal{F}_T, \tilde{P})$ ,  $W_t = B_t + \left(\frac{\mu-r}{\sigma}\right) t$  est un mouvement brownien (standard) sur  $[0, T]$ .

PREUVE : On commence par montrer que l'on a bien une probabilité. Comme

$$L_T := \exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_T - \frac{T}{2} \left(\frac{\mu-r}{\sigma}\right)^2\right)$$

est positive, on a bien la positivité, la sigma-additivité (par convergence monotone), il reste à noter que  $\tilde{P}(\Omega) = 1$  ce qui revient à

$$\mathbf{E}(\exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_T\right)) = \exp\left(\frac{T}{2} \left(\frac{\mu-r}{\sigma}\right)^2\right)$$

C'est le calcul fait dans la preuve du lemme 5 pour la transformée de Laplace d'une variable gaussienne.

Pour montrer que  $W_t$  est un mouvement brownien, on va vérifier les hypothèses du théorème 66. On va utiliser souvent la propriété suivante (intégration par partie gaussienne) d'une variable  $Y$  gaussienne  $\mathcal{N}(0, \sigma^2)$ , pour tout  $f$   $C^1$  telle que  $Yf(Y), f'(Y) \in L^1(\Omega)$ , on a (facile à vérifier après transfert par IPP) :

$$\mathbf{E}(Yf(Y)) = \sigma^2 \mathbf{E}(f'(Y)). \quad (6.1)$$

On remarque que la continuité p.s. est évidente avec le même  $A$ .

On remarque d'abord pour  $f$  mesurable bornée ou  $f(x) = x^k$   $k = 1, 2$  (d'abord  $k = 2$  pour obtenir intégrabilité), par indépendance des accroissements du mouvement brownien :

$$\begin{aligned} \mathbf{E}(f(B_t) \exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_T\right)) &= \mathbf{E}(f(B_t) \exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_t\right) \exp\left(-\left(\frac{\mu-r}{\sigma}\right) (B_T - B_t)\right)) \\ &= \mathbf{E}(f(B_t) \exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_t\right)) \mathbf{E}(\exp\left(-\left(\frac{\mu-r}{\sigma}\right) (B_T - B_t)\right)) \end{aligned}$$

donc pour  $t \leq T$ ,  $\mathbf{E}(f(B_t)L_T) = \mathbf{E}(f(B_t)L_t)$ . (En fait on peut voir que  $L_t$  est une martingale à temps continu, du type des martingales exponentielles de l'exemple ??). Il est alors facile de voir que  $W_t \in L^2(\Omega, \tilde{P})$ , car  $B_t^2 \exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_t\right) \in L^1(\Omega)$  par une borne par une fonction  $\exp(cB_t) + \exp(-cB_t) \in L^1(\Omega)$ .

**-Calcul de l'espérance.**

On utilise (6.1) pour obtenir :

$$\mathbf{E}(B_t \exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_t\right)) = -\left(\frac{\mu-r}{\sigma}\right) t \mathbf{E}(\exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_t\right))$$

On déduit alors du calcul utilisant l'indépendance ci-dessus que

$$\tilde{\mathbf{E}}(W_t) = \mathbf{E}((B_t + \left(\frac{\mu-r}{\sigma}\right) t)L_T) = \mathbf{E}((B_t + \left(\frac{\mu-r}{\sigma}\right) t)L_t) = \mathbf{E}(B_t L_t) + \left(\frac{\mu-r}{\sigma}\right) t = 0.$$

**-Calcul de la variance.**

On utilise (6.1) (puis le calcul précédent) pour obtenir :

$$\begin{aligned} \mathbf{E}(B_t^2 \exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_t\right)) &= t \mathbf{E}(\exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_t\right)) - \left(\frac{\mu-r}{\sigma}\right) t \mathbf{E}(B_t \exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_t\right)) \\ &= [t + \left(\frac{\mu-r}{\sigma}\right)^2 t^2] \mathbf{E}(\exp\left(-\left(\frac{\mu-r}{\sigma}\right) B_t\right)). \end{aligned}$$

On déduit alors du calcul utilisant l'indépendance ci-dessus et du calcul de l'espérance que

$$\tilde{\mathbf{E}}(W_t^2) = \mathbf{E}(B_t^2 L_t) + 2\mathbf{E}(B_t L_t) \left( \frac{\mu - r}{\sigma} \right) t + \left( \frac{\mu - r}{\sigma} \right)^2 t^2 = \mathbf{E}(B_t^2 L_t) - \left( \frac{\mu - r}{\sigma} \right)^2 t^2 = t.$$

**-Calcul de transformées de Fourier pour obtenir les incréments gaussiens.** En fait, cette étape suffit car on va recalculer la variance et la moyenne avec la transformée de Fourier. On utilise juste la preuve du lemme 5 :

$$\begin{aligned} \mathbf{E}(\exp(isB_t)\exp\left(-\left(\frac{\mu - r}{\sigma}\right)B_t\right)) &= \exp\left(\frac{t}{2}\left(is - \frac{\mu - r}{\sigma}\right)^2\right) \\ &= \exp\left(-\frac{ts^2}{2} + \frac{t}{2}\left(\frac{\mu - r}{\sigma}\right)^2 - ist\left(\frac{\mu - r}{\sigma}\right)\right) \end{aligned}$$

d'où la valeur attendue :

$$\tilde{\mathbf{E}}(\exp(isW_t)) = \mathbf{E}(\exp(isB_t + ist\left(\frac{\mu - r}{\sigma}\right))L_t) = \exp\left(-\frac{ts^2}{2}\right).$$

### -Indépendance et stationnarité des incréments

Pour tout  $n$  et tout  $t_1 < \dots < t_n$ , montrons que  $W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$  sont indépendants. Prenons donc aussi  $f_1, \dots, f_n$  mesurables bornées (par exemple des exponentielles donnant des fonctions caractéristiques). On note  $\theta = \frac{\mu - r}{\sigma}$  et on obtient en utilisant l'indépendance des incréments de  $B_t$  sous  $P$  :

$$\begin{aligned} \tilde{\mathbf{E}}(f_1(W_{t_1})f_2(W_{t_2} - W_{t_1})\dots f_n(W_{t_n} - W_{t_{n-1}})) &= \mathbf{E}(f_1(W_{t_1})f_2(W_{t_2} - W_{t_1})\dots f_n(W_{t_n} - W_{t_{n-1}})L_{t_n}) \\ &= \mathbf{E}\left(f_1(W_{t_1})f_2(W_{t_2} - W_{t_1})\dots f_n(W_{t_n} - W_{t_{n-1}})L_{t_1} \prod_{k=2}^n \exp\left(-\theta(B_{t_k} - B_{t_{k-1}}) - \frac{t_k - t_{k-1}}{2}\theta^2\right)\right) \\ &= \mathbf{E}(f_1(W_{t_1})L_{t_1}) \prod_{k=2}^n \mathbf{E}\left(f_k(W_{t_k - t_{k-1}})\exp\left(-\theta B_{t_k - t_{k-1}} - \frac{t_k - t_{k-1}}{2}\theta^2\right)\right) \\ &= \tilde{\mathbf{E}}(f_1(W_{t_1})) \tilde{\mathbf{E}}(f_2(W_{t_2} - W_{t_1})) \dots \tilde{\mathbf{E}}(f_n(W_{t_n} - W_{t_{n-1}})) \end{aligned}$$

la dernière ligne étant obtenue en prenant tous les  $f_i$  égaux à 1 sauf un pour identifier les termes. La stationnarité des incréments se déduit du même calcul en utilisant celle de  $B_t$ .  $\blacksquare$

En finance, un produit dérivé (ou option) est un contrat sur l'actif financier  $S_t$  qui produit un paiement positif  $H$ ,  $\mathcal{F}_T$  mesurable au temps d'échéance  $T$ . Nous ne considérerons que les options dites européennes de la forme  $H = \Phi(S_T)$  avec  $\Phi$  mesurable positive. La définition suivante (motivée dans la section suivante) permet de donner une valeur à l'option à l'instant  $t$  (même si l'option n'est pas cotée, et permettant de détecter des "anomalies" de marché (par exemple possibilités d'arbitrages) si la cotation ne correspond pas à la valeur théorique).

**Définition 33.** La valeur au temps  $t \leq T$  de l'option, définie par la variable aléatoire  $\Phi(S_T)$  d'échéance  $T$ , est donnée par la formule utilisant la probabilité  $\tilde{P}$  du théorème 71 par :

$$V_t = \tilde{\mathbf{E}}(e^{-r(T-t)}\Phi(S_T)|\mathcal{F}_t).$$

Le prix à l'instant  $t = 0$  est la valeur de vente conseillée (par la théorie) dans ce modèle pour l'option. On va obtenir le calcul explicite dans le cas des options de type call  $\Phi(x) = \max(x - K, 0) =: (x - K)_+$  et put  $\Phi(x) = (K - x)_+$ .

On note la fonction de répartition d'une gaussienne standard

$$N(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{x^2}{2}} dx$$

et on pose

$$\rho_{\pm}(t, x) = \frac{\ln(\frac{x}{K}) + (r \pm \frac{\sigma^2}{2})(T - t)}{\sigma\sqrt{T - t}}.$$

On remarquera que  $\rho_+(t, x) = \rho_-(t, x) + \sigma\sqrt{T - t}$ .

**Théorème 72.** (valeur des put et call de Black-Scholes) La valeur au temps  $t \leq T$  de l'option de type call  $(S_T - K)_+$  d'échéance  $T$  est donnée par  $V_t = F_c(t, S_t)$  avec :

$$F_c(t, x) = xN(\rho_+(t, x)) - Ke^{-r(T-t)}N(\rho_-(t, x)).$$

La valeur au temps  $t \leq T$  de l'option de type put  $(K - S_T)_+$  d'échéance  $T$  est donnée par  $V_t = F_p(t, S_t)$  avec :

$$F_p(t, x) = Ke^{-r(T-t)}N(-\rho_-(t, x)) - xN(-\rho_+(t, x)).$$

On remarquera que ces valeurs sont indépendantes de  $\mu$  (grâce au changement de probabilité) et ne dépendent que de la seule volatilité  $\sigma$ .

PREUVE : On note d'abord qu'avec les notations du Théorème de Girsanov, la valeur actualisée de l'actif est par calcul élémentaire  $\tilde{S}_t := S_t e^{-rt} = S_0 \exp(\sigma W_t - \sigma^2 t/2)$  et donc

$$S_T = S_t e^{r(T-t)} \exp(\sigma(W_T - W_t) - \sigma^2(T - t)/2).$$

Donc  $\tilde{\mathbf{E}}(S_T - K | \mathcal{F}_t) = S_t e^{r(T-t)} \mathbf{E}(\exp(\sigma(W_T - W_t) - \sigma^2(T - t)/2)) - K = S_t e^{r(T-t)} - K = F(t, S_t)$  avec  $F(t, x) = x e^{r(T-t)} - K$ . Puisque  $(S_T - K)_+ - (K - S_T)_+ = S_T - K$  on déduit donc, après multiplication par  $e^{-r(T-t)}$  la relation que nos formules vérifient :

$$F_c(t, x) - F_p(t, x) = x - Ke^{-r(T-t)}.$$

On considère donc le cas du call. Donc avec  $Y = \frac{W_T - W_t}{\sqrt{T-t}}$  on obtient par les calculs précédents avec  $\theta := \sigma\sqrt{T-t}$  :

$$e^{-r(T-t)}(S_T - K)_+ = (S_t \exp(\theta Y - \theta^2/2) - Ke^{-r(T-t)})_+ = (S_t \exp(\theta Y - \theta^2/2) - Ke^{-r(T-t)}) 1_{\{Y + \rho_-(t, S_t) \geq 0\}}.$$

En effet, on a noté  $S_t \exp(\theta Y - \theta^2/2) - Ke^{-r(T-t)}$  est positif si et seulement si  $\ln(S_t/K) \geq -r(T-t) - (\theta Y - \theta^2/2)$  soit  $\ln(S_t/K) + r(T-t) - \theta^2/2 + \theta Y \geq 0$  ou encore  $Y + \rho_-(t, S_t) \geq 0$ .

En prenant l'espérance conditionnelle sur  $\mathcal{F}_t$  en utilisant l'indépendance des accroissements de  $W_t$  sous  $\tilde{P}$  c'est à dire celle de  $Y$  avec  $\mathcal{F}_t$  et la proposition 39, on obtient  $V_t = F_c(t, S_t)$  avec

$$F_c(t, x) = x \tilde{\mathbf{E}}(\exp(\theta Y - \theta^2/2) 1_{\{Y + \rho_-(t, x) \geq 0\}}) - Ke^{-r(T-t)} \tilde{\mathbf{E}}(1_{\{Y + \rho_-(t, x) \geq 0\}}).$$

Comme  $Y \sim \mathcal{N}(0, 1)$  sous  $\tilde{P}$  on a évidemment

$$\tilde{\mathbf{E}}(1_{\{Y + \rho_-(t, x) \geq 0\}}) = \tilde{P}(Y \geq -\rho_-(t, x)) = \tilde{P}(Y \leq \rho_-(t, x)) = N(\rho_-(t, x)).$$

De même, on a la relation conclusive (obtenue par 2 transferts et changement de variable  $y - \theta$ ) :

$$\tilde{\mathbf{E}}(\exp(\theta Y - \theta^2/2) 1_{\{Y + \rho_-(t, x) \geq 0\}}) = \int_{-\rho_-(t, x)}^{\infty} \exp(-(y - \theta)^2/2) \frac{dy}{\sqrt{2\pi}} = \tilde{\mathbf{E}}(1_{\{Y + \theta + \rho_-(t, x) \geq 0\}}) = N(\rho_+(t, x)).$$

■

### 3 Motivation de l'intégrale stochastique pour l'étude du modèle de Black et Scholes

Le but de cette section sans preuve est de motiver l'utilisation d'une notion d'intégrale stochastique pour l'interprétation du prix défini ci-dessus dans le modèle de Black et Scholes.

Il existe une notion d'intégrale stochastique qui permet d'écrire une formule intégrale pour  $f(t, B_t)$  si  $f : (t, x) \mapsto f(t, x)$  est  $C^2$  (ou même seulement  $C^1$  en  $t$ ). On obtient alors la formule d'Ito :

$$f(t, B_t) = f(0, 0) + \int_0^t \left( \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \right)(s, B_s) ds + \int_0^t \frac{\partial f}{\partial x}(s, B_s) dB_s$$

Le changement notable par rapport au cas dérivable en temps (on rappelle que  $B_t$  est p.s. nulle part dérivable), c'est le terme faisant intervenir la dérivée seconde de  $f$ . Par exemple, on obtient

$$S_t = S_0 + \int_0^t \mu S_s ds + \int_0^t \sigma S_s dB_s.$$

En écriture infinitésimale  $dS_t = S_t(\mu dt + \sigma dB_t)$  qui est très semblable au modèle discret de l'exemple 41  $Y_n - Y_{n-1} = Y_{n-1}(r + \sigma \epsilon_n)$ . On peut manipuler ces expressions en utilisant les "règles"

$$dt^2 = 0, dt dB_t = 0, dB_t^2 = dt.$$

C'est un théorème qu'une intégrale stochastique  $\int_0^t f_s dB_s$  est définie pour  $f_s$  adaptée dans  $L^2([0, T], L^2(\Omega, P))$   $L^2([0, T] \times \Omega, dt \otimes P)$  par l'isométrie d'Ito :

$$\mathbf{E} \left[ \left| \int_0^t f_s dB_s \right|^2 \right] = \int_0^t \mathbf{E} [ |f(s)|^2 ] dt.$$

C'est souvent une "martingale" sur  $[0, T]$  si  $f$  est suffisamment bornée (sinon c'est une "martingale locale").

Une stratégie (ou portefeuille) pour le couple  $(S_t^0, S_t)$  est une paire de valeurs  $\phi_t = (\alpha_t, \beta_t)$  représentant les valeurs achetées de l'actif sans risque  $S_t^0$  et de l'actif risqué  $S_t$  au temps  $t$ . La valeur du portefeuille est donc :

$$V_t(\phi) = \alpha_t S_t^0 + \beta_t S_t$$

Celui-ci est auto-financé si (avec des hypothèses appropriées sur  $\phi_t$  pour que les intégrales aient un sens)

$$V_t(\phi) = V_0(\phi) + \int_0^t \alpha_s dS_s^0 + \int_0^t \beta_s dS_s$$

qui revient à demander, dans l'intervalle  $[t, t + dt]$ , que la variation du portefeuille  $dV_t(\phi) = \alpha_s dS_t^0 + \beta_s dS_t$  ne viennent que des variations des valeurs des actifs et pas d'apports au portefeuille ( $d\alpha_s, d\beta_s$ ). On suppose en particulier  $(\alpha_t), (\beta_t)$  adaptés à la filtration  $\mathcal{F}_t$ . Si on regarde les valeurs actualisées  $\tilde{S}_t = (S_t^0)^{-1} S_t, \tilde{V}_t(\phi) = (S_t^0)^{-1} V_t(\phi) = \alpha_t + \beta_t \tilde{S}_t$ , on a les relations (pour un portefeuille autofinancé)  $d\tilde{S}_t = \tilde{S}_t[(\mu - r)dt + \sigma dB_t]$ , et

$$d\tilde{V}_t(\phi) = -r e^{-rt} dt V_t(\phi) + e^{-rt} dV_t(\phi) = -r(\alpha_t + \beta_t \tilde{S}_t) dt + r \alpha_t dt + \beta_t e^{-rt} dS_t = \beta_t d\tilde{S}_t$$

si bien que la valeur de  $\alpha_t$  est déterminée par

$$\alpha_t = V_t(\phi) - \beta_t \tilde{S}_t = V_0(\phi) + \int_0^t \beta_s d\tilde{S}_s - \beta_t \tilde{S}_t$$

On retrouve la condition  $\mu = r$  pour que  $\tilde{S}_t$  et  $\tilde{V}_t(\phi)$  soient des martingales (locales) sous  $P$  (similaires à l'exemple 41, ce sont toujours des martingales locales sous  $\tilde{P}$ ).

On peut montrer que si  $V_t(\phi) \geq 0$  (on dit admissible) alors il n'y a pas de possibilité d'arbitrage, c'est à dire de stratégies auto-financées avec  $V_0 = 0, V_T \geq 0$  et  $P(V_T > 0) > 0$ .

Une option  $\Phi(S_T)$  est dite répliquée si il existe un portefeuille autofinancé tel que  $V_T(\phi) = \Phi(S_T)$ . La valeur de l'option (au sens du paragraphe précédent).  $V_t = V_t(\phi)$  est la valeur d'un tel portefeuille autofinancé au temps  $t$

On peut montrer l'existence du portefeuille sous l'hypothèse  $\mathbf{E}(L_T^2 \Phi(S_T)^2) < \infty$ . Elle nécessite des outils de calculs stochastiques dépassant la portée de ce cours (et la condition  $\mathbf{E}(L_T^2 H^2) = \tilde{\mathbf{E}}(H^2) < \infty$  fonctionne pour des options non-européennes  $H \neq \Phi(S_T)$  mais la solution est alors moins explicite).

Le calcul du portefeuille simulant est utile en finance pour la "couverture de l'option" (par le vendeur vendant au prix  $V_0(\phi)$  qui utilisera cette stratégie pour produire une richesse  $V_T(\phi)$  au temps  $T$  et qu'il doit donner à l'acheteur). Donc le cas du put et du call, on a les formules

$$\alpha_t = e^{-rt}(F_{p/c}(t, S_t) - \beta_t S_t), \quad \beta_t = \frac{\partial F_{p/c}}{\partial x}(t, S_t).$$

Pour le call on a  $\beta_t = N(\rho_+(t, S_t))$  et pour le put  $\beta_t = -N(-\rho_+(t, S_t))$ .