

# Analyse spectrale des processus

Philippe Jaming

`Philippe.Jaming@math.u-bordeaux1.fr`



## Contents

Chapître 1. Rappels de probabilité	5
1. Généralités	5
2. Variables aléatoires gaussiennes	10
3. Convergence de variables aléatoires	14
4. Espérance conditionnelle	17
5. Solution de certains exercices	21
Chapître 2. Chaînes de Markov et processus gaussien	27
1. Introduction	27
2. Chaîne de Markov homogène	28
3. Processus Gaussien	30
4. Opérateur de covariance	31
5. Filtrage des processus stationnaires	32
6. Processus ARMA	35
Appendice A. Matrices stochastiques	39
1. Rappels d’algèbre linéaire	39
2. Matrices stochastiques	40
Appendice B. Travaux pratiques 1: Simulation de variables aléatoires (densités, répartition, histogrammes et corrélation)	43
1. Introduction	43
2. Génération de réalisations d’une variable aléatoire	43
3. Densité de probabilité et histogramme	44
4. Couple, transformation linéaire et corrélation	45
5. Algorithme de Box-Muller	46
6. Simulation d’un vecteur gaussien	46
7. Méthode d’inversion	46
8. Méthode du rejet	47
9. Solution: simulation de v.a. discrète	47
10. Théorème centrale limite	47
Appendice C. Travaux pratiques 2: chaînes de Markov	49
1. Chaîne de Markov	49
2. Solution	49
Appendice D. Travaux pratique: débruitage	51
1. Du bruit dans des signaux et images	51
2. Débruitage par filtrage linéaire	52



## Rappels de probabilité

### 1. Généralités

**1.1. Tribus et mesures.** Soit  $\Omega$  un ensemble. Une *tribu*  $\mathcal{A}$  sur  $\Omega$  est un ensemble de parties de  $\Omega$ ,  $\mathcal{A} \subset \mathcal{P}(\Omega)$  tel que

- (i)  $\emptyset \in \mathcal{A}$  et  $\Omega \in \mathcal{A}$ ;
- (ii) si  $A \in \mathcal{A}$  alors  $A^c = \Omega \setminus A \in \mathcal{A}$ ;
- (iii) si pour tout  $k \in \mathbb{N}$ ,  $A_k \in \mathcal{A}$ , alors  $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$  et  $\bigcap_{i \in \mathbb{N}} A_i \in \mathcal{A}$ .

Notez que si  $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$  alors il suffit de vérifier que  $\mathcal{A}$  est stable par intesection (resp. réunion) dénombrable.

Une *mesure* (positive) est une application  $\mu : \mathcal{A} \rightarrow [0, +\infty)$  telle que  $\mu(\emptyset) = 0$ , si  $A$  et  $B$  sont disjoints, alors  $\mu(A \cup B) = \mu(A) + \mu(B)$ . On suppose de plus que  $\mu$  est continue au sens suivant: si  $(A_n)$  est une suite croissante d'ensembles *i.e.*  $A_n \subset A_{n+1}$ , alors  $\mu(\bigcup A_n) = \lim \mu(A_n)$  et si  $(A_n)$  est une suite décroissante d'ensembles *i.e.*  $A_{n+1} \subset A_n$ , alors  $\mu(\bigcap A_n) = \lim \mu(A_n)$ .

Une mesure est dite *finie* si  $\mu(\Omega) < +\infty$  et de *probabilité* si  $\mu(\Omega) = 1$ .

EXEMPLE 1.1. Les mesures que nous rencontrerons le plus souvent sont de deux types:

— *discrètes*:  $\mu(x) = \sum \lambda_j \delta_{x_j}$  avec  $\lambda_j \geq 0$ ,  $\sum \lambda_j = 1$  et il existe  $\delta > 0$  tel que, si  $j \neq k$ , alors  $\|x_j - x_k\| \geq \delta$ . Le plus souvent, les  $x_j$  sont des entiers.

— *continues*:  $\mu(x) = \int f(x) dx$  avec  $dx$  mesure de Lebesgue. Alors  $\mu$  est une mesure de probabilité si  $f \geq 0$  et  $\int_{\Omega} f(x) dx = 1$ . Par exemple  $f(x) = (2\pi)^{-1/2} e^{-x^2/2}$  (mesure gaussienne),  $f(x) = \frac{1}{|\Omega|} \mathbf{1}_{\Omega}$

(mesure uniforme sur  $\Omega$ ) ou  $f(x) = \frac{1}{\pi(1+x^2)}$  (mesure de Poisson).

### 1.2. Variable et vecteur aléatoire.

DÉFINITION 1.2. Une *variable aléatoire*  $X$  est une application de  $(\Omega, \mathcal{A}, \mathbb{P})$  dans  $\mathbb{R}$  (ou  $\mathbb{C}$ ) qui est mesurable *i.e.* pour tout ouvert  $O$  de  $\mathbb{R}$ ,  $X^{-1}(O) = \{\omega \in \Omega : X(\omega) \in O\}$  est un ensemble mesurable.

Plus généralement, un *vecteur aléatoire*  $X$  est une application de  $(\Omega, \mathcal{A}, \mathbb{P})$  dans  $\mathbb{R}^n$  (resp. dans  $\mathbb{C}^n$ ) qui est mesurable, *i.e.* l'image réciproque d'un ouvert de  $\mathbb{R}^n$  (resp. d'un ouvert de  $\mathbb{C}^n$ ) est dans  $\mathcal{A}$ .

Ainsi une *vecteur* aléatoire  $X$  est en fait une *application*  $\forall \omega \in \Omega$ ,  $X(\omega) \in \mathbb{R}^n$ . Les composantes  $X_j$  du vecteur  $X = (X_1, \dots, X_n)$  sont des variables aléatoires. Réciproquement la données de  $n$  variables aléatoires constitue un vecteur aléatoire.

On parlera de variable *discrète* si les valeurs de  $X$  sont dans  $\mathbb{N}$  ou  $\mathbb{Z}$  (ou un autre ensemble discret). On dira que  $X$  est de Bernouilli si  $X$  est à valeurs dans  $\{0, 1\}$ .

DÉFINITION 1.3. L'*espérance* ou *moyenne* d'une variable aléatoire  $X$  à valeurs dans  $\mathbb{R}$  (ou  $\mathbb{C}$ ) est

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

Si  $X = (X_1, \dots, X_n)$  est un vecteur aléatoire, alors l'espérance de  $X$  est le vecteur des espérances des coordonnées de  $X$ :  $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$ .

La *variance* d'une variable aléatoire  $X$  à valeurs dans  $\mathbb{R}$  (ou  $\mathbb{C}$ ) est  $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$  donc

$$\text{Var}(X) = \int_{\Omega} (X(\omega) - \mathbb{E}[X])^2 d\mathbb{P}(\omega).$$

Nous verrons plus loin l'extension de la définition de la variance aux vecteurs aléatoires.

REMARQUE 1.4. L'espérance vérifie les propriétés de l'intégration: elle est linéaire, préserve la positivité. De plus  $\mathbb{E}[1] = \mathbb{P}(\Omega) = 1$  puisqu'on a une mesure de probabilité.

Pour que  $\mathbb{E}[X]$  soit défini, il faut que  $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$  et pour que  $\text{Var}(X)$  soit définie, il faut de plus que  $X \in L^2(\Omega)$ . Notons que, comme  $\mathbb{E}[1] = 1$ , l'inégalité de Cauchy Schwarz donne  $\mathbb{E}[|X|] \leq \mathbb{E}[X^2]^{1/2}$  donc si  $\mathbb{E}[X^2]$  est définie,  $\mathbb{E}[X]$  aussi et de plus, par linéarité de l'espérance

$$\text{Var}(X) = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2\mathbb{E}[1] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

DÉFINITION 1.5. On appelle *loi* du vecteur aléatoire  $X = (X_1, \dots, X_n)$  la probabilité  $\mathbb{P}_X$  sur  $\mathbb{R}^n$  telle que, pour tous  $a_1, b_1, a_2, b_2, \dots, a_n, b_n$ ,

$$\mathbb{P}_X([a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]) := \mathbb{P}(X^{-1}([a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n])).$$

EXEMPLE 1.6. Soit  $X$  le nombre de 6 qu'on obtient en lançant 2 dés. Ainsi  $X$  est une variable aléatoire de  $\Omega := \{1, \dots, 6\} \times \{1, \dots, 6\} \rightarrow \{0, 1, 2\}$  donnée par  $X(6, 6) = 2$ ,  $X(1, 6) = X(6, 1) = 1$  et  $X(j, k) = 0$  dans les autres cas. La mesure de probabilité sur  $\Omega$  que nous considérons est la mesure de probabilité uniforme (on suppose les dés non pipés)  $\mathbb{P}(F) = \frac{|F|}{|\Omega|} = \frac{|F|}{36}$ . Mais alors la loi de  $X$  est une mesure de probabilité sur  $\{0, 1, 2\}$  donnée par

$$\mathbb{P}_X(\{2\}) = \mathbb{P}(\{X = 2\}) = \frac{1}{36}, \quad \mathbb{P}_X(\{1\}) = \mathbb{P}(\{X = 1\}) = \frac{2}{36} = \frac{1}{18}$$

et

$$\mathbb{P}_X(\{0\}) = \mathbb{P}(\{X = 0\}) = \frac{33}{36} = \frac{11}{12}.$$

Une caractérisation de la loi de  $X$  est encore donnée par le lemme suivant

LEMME 1.7. Soit  $X$  un vecteur aléatoire sur un espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^n$ . Alors la loi de  $X$  est l'unique mesure de probabilité  $\mathbb{P}_X$  sur  $\mathbb{R}^n$  telle que

$$\forall h \in L^\infty(\mathbb{R}^n), \quad \mathbb{E}[h(X)] := \int_{\Omega} h(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathbb{R}^n} h(x) d\mathbb{P}_X(x).$$

Ainsi l'application  $X$  devient la variable  $x$  quand on passe à la loi. Par exemple, même si cela ne rentre pas directement dans le cadre du lemme, si  $X$  est une variable aléatoire.

$$\mathbb{E}[X] = \int_{\mathbb{R}} x d\mathbb{P}_X(x) \quad \text{et} \quad \text{Var}(X) = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 d\mathbb{P}_X(x) = \int_{\mathbb{R}} x^2 d\mathbb{P}_X(x) - \left( \int_{\mathbb{R}} x d\mathbb{P}_X(x) \right)^2.$$

Si on connaît la loi du vecteur  $X$ , on connaît la loi de chacune de ses composantes, comme *loi marginale*: Fixons  $h_1 \in L^\infty(\mathbb{R})$  et définissons  $h \in L^\infty(\mathbb{R}^n)$  par  $h(x_1, \dots, x_n) = h_1(x_1)$ . Alors d'une part

$$\mathbb{E}[h(X)] = \int_{\Omega} h(X_1(\omega), \dots, X_n(\omega)) d\omega = \int_{\Omega} h_1(X_1(\omega)) d\omega = \mathbb{E}[h_1(X_1)] = \int_{\mathbb{R}} h_1(x_1) d\mathbb{P}_{X_1}(x_1)$$

par définition de la loi de  $X_1$ . D'autre part, par définition de la loi de  $X$ ,

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^n} h(x_1, \dots, x_n) d\mathbb{P}_X(x_1, \dots, x_n) = \int_{\mathbb{R}^n} h_1(x_1) d\mathbb{P}_X(x_1, \dots, x_n).$$

Ainsi, pour tout  $h_1 \in L^\infty(\mathbb{R})$ ,

$$\int_{\mathbb{R}} h_1(x_1) d\mathbb{P}_{X_1}(x_1) = \int_{\mathbb{R}^n} h_1(x_1) d\mathbb{P}_X(x_1, \dots, x_n).$$

Dans le cas particulier où  $h_1(x) = \mathbf{1}_A$  avec  $A \in \mathcal{A}$ , on obtient

$$\mathbb{P}_{X_1}(A) = \mathbb{P}_X(A \times \mathbb{R} \times \cdots \times \mathbb{R}).$$

Enfin, notons que si  $\mathbb{P}_X$  a une densité (par rapport à la mesure de Lebesgue),  $\mathbb{P}_X = f_X(x) dx$ , alors  $X_1$  aussi,  $\mathbb{P}_{X_1} = f_{X_1}(x_1) dx_1$  et

$$f_{X_1}(x_1) = \int_{\mathbb{R}^{n-1}} f(x_1, t_2, \dots, t_n) dt_2 \cdots dt_n.$$

En particulier, si deux vecteurs  $X = (X_1, \dots, X_n)$  et  $Y = (Y_1, \dots, Y_n)$  ont même loi, alors leurs composantes aussi:  $X_j$  et  $Y_j$  ont même loi pour  $j = 1, \dots, n$ . Notons que la réciproque est fautive en générale, il existe des vecteurs dont les composantes ont même loi mais qui n'ont pas eux-même la même loi (cf. le chapitre sur les vecteurs gaussiens).

### EXEMPLE 1.8. Principales lois discrètes

Une variable aléatoire  $X$  à valeurs dans  $\mathbb{Z}$  sera dite

- (i) Bernouilli( $p$ ) avec  $p \in [0, 1]$  si  $\mathbb{P}_X(0) = \mathbb{P}(X = 0) = 1 - p$ ,  $\mathbb{P}_X(1) = \mathbb{P}(X = 1) = p$ . Dans ce cas  $\mathbb{E}[X] = p$  et  $\text{Var}(X) = p(1 - p)$ .
- (ii) Binomiale( $n, p$ ) avec  $n > 0$  et  $p \in [0, 1]$  si  $\mathbb{P}_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$  si  $k = 0, \dots, n$ . Dans ce cas  $\mathbb{E}[X] = np$  et  $\text{Var}(X) = np(1 - p)$ .
- (iii) Géométrique( $p$ ) avec  $p \in [0, 1]$  si  $\mathbb{P}_X(k) = \mathbb{P}(X = k) = p(1 - p)^{k-1}$  pour  $k = 1, 2, \dots$ . Dans ce cas  $\mathbb{E}[X] = \frac{1}{p}$  et  $\text{Var}(X) = \frac{1-p}{p^2}$ .
- (iv) Poisson( $\lambda$ ) avec  $\lambda > 0$  si  $\mathbb{P}_X(k) = \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$  pour  $k = 0, 1, \dots$ . Dans ce cas  $\mathbb{E}[X] = \lambda$  et  $\text{Var}(X) = \lambda$ .

Par convention,  $\mathbb{P}_X(k) = 0$  si elle n'est pas définie par une des formules ci-dessus.

exo:espvardisc

EXERCICE 1.1. Dans les cas ci-dessus, vérifier que  $\mathbb{P}_X$  est bien une mesure de probabilité sur  $\mathbb{Z}$ . Calculer les espérances et les variances des variables correspondantes.

### EXEMPLE 1.9. Principales lois continues

Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{R}$ . On dit que  $X$  suit une loi

- (i) uniforme  $\mathcal{U}(a, b)$  avec  $a < b \in \mathbb{R}$  si  $\mathbb{P}_X(t) = \frac{1}{b-a} \mathbf{1}_{[a, b]}(x) dx$ . Dans ce cas  $\mathbb{E}[X] = \frac{a+b}{2}$ ,  $\text{Var}(X) = \frac{(a-b)^2}{12}$ .
- (ii) normale  $\mathcal{N}(m, \sigma^2)$   $m \in \mathbb{R}$ ,  $\sigma > 0$  si  $\mathbb{P}_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx$ . Dans ce cas  $\mathbb{E}[X] = m$ ,  $\text{Var}(X) = \sigma^2$ .
- (iii) exponentielle  $\mathcal{E}(\lambda)$ ,  $\lambda > 0$  si  $\mathbb{P}_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x) dx$ . Dans ce cas  $\mathbb{E}[X] = \frac{1}{\lambda}$ ,  $\text{Var}(X) = \frac{1}{\lambda^2}$ .
- (iv) de Cauchy de paramètre  $a > 0$  si  $\mathbb{P}_X(x) = \frac{a}{\pi(a^2 + x^2)} dx$ . Dans ce cas  $X$  n'a ni espérance ni variance.

ex:loicont

EXERCICE 1.2. Dans les cas ci-dessus, vérifier que  $\mathbb{P}_X$  est bien une mesure de probabilité sur  $\mathbb{R}$ . Calculer les espérances et les variances (si elles existent) des variables aléatoires correspondantes.

### 1.3. Indépendance et fonction caractéristique.

DÉFINITION 1.10. Soit  $X$  et  $Y$  deux variables aléatoires  $\Omega \rightarrow \mathbb{R}$ . La variance de  $X$  est définie par  $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$  et la covariance de  $X$  et  $Y$  est définie par  $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ .

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire  $\Omega \rightarrow \mathbb{R}^n$ . La covariance de  $X$  est la matrice symétrique  $\text{Cov}(X) = [\text{Cov}(X_j, X_k)]_{1 \leq j, k \leq n}$ .

REMARQUE 1.11. Si on note  $\tilde{X} = X - \mathbb{E}[X]$  alors  $\mathbb{E}[\tilde{X}] = \mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[\mathbb{E}[X]] = 0$ . On dit que  $\tilde{X}$  est *centrée*.

Ainsi  $\text{Cov}(X, Y) = \text{Cov}(\tilde{X}, \tilde{Y}) = \langle \tilde{X}, \tilde{Y} \rangle$  où  $\langle \cdot, \cdot \rangle$  est le produit scalaire (réel) de  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ .

On dit que  $X, Y$  sont *décorrélées* si  $\text{Cov}(X, Y) = 0$ , i.e. si  $\tilde{X}$  et  $\tilde{Y}$  sont orthogonales.

Notons enfin que  $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ . Des variables sont donc décorréllées si  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ .

**lem:cov**

LEMME 1.12. Soit  $X$  un vecteur aléatoire dans  $\mathbb{R}^d$ ,  $m \in \mathbb{R}^n$  et  $A \in \mathcal{M}_{n,d}$  une matrice  $(n, d)$  réelle. Alors

$$\mathbb{E}[AX + m] = A\mathbb{E}[X] + m \quad \text{et} \quad \text{Cov}(AX + m) = A\text{Cov}(X)A^t.$$

DÉMONSTRATION.  $(AX + m)_j = \sum_k a_{j,k}X_k + m_j$  donc, en utilisant la linéarité de  $\mathbb{E}$  et le fait que  $\mathbb{E}[1] = 1$ ,

$$\mathbb{E}[(AX + m)_j] = \sum_k a_{j,k}\mathbb{E}[X_k] + m_j = (A\mathbb{E}[X] + m)_j.$$

Mais alors  $AX + m - \mathbb{E}[AX + m] = A(X - \mathbb{E}[X])$ . Pour calculer la covariance, on peut donc supposer que  $\mathbb{E}[X] = 0$  et que  $m = 0$ .

Mais alors  $(AX)_j = \sum_\ell a_{j,\ell}X_\ell$  et, en notant  $a_{j,k}^t = a_{k,j}$  les entrées de la matrice  $A^t$ ,

$$(AX)_j(AX)_k = \sum_\ell \sum_{\ell'} a_{j,\ell}a_{k,\ell'}X_\ell X_{\ell'} = \sum_\ell \sum_{\ell'} a_{j,\ell}a_{\ell',k}^t X_\ell X_{\ell'}.$$

Ainsi

$$\begin{aligned} \mathbb{E}[(AX)_j(AX)_k] &= \sum_\ell \sum_{\ell'} a_{j,\ell}a_{k,\ell'}\mathbb{E}[X_\ell X_{\ell'}]a_{\ell',k}^t = \sum_\ell \sum_{\ell'} a_{j,\ell}\text{Cov}(X)_{\ell,\ell'}a_{\ell',k}^t \\ &= \sum_\ell a_{j,\ell}(\text{Cov}(X)A^t)_{\ell,k} = (A\text{Cov}(X)A^t)_{j,k} \end{aligned}$$

qui est bien le résultat annoncé.  $\square$

DÉFINITION 1.13. On dit que  $X_1, \dots, X_n$  sont indépendantes si  $\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ .

REMARQUE 1.14. Tout d'abord, notons que  $\mathbb{P}_{X+m}(E) = \mathbb{P}_X(E - m)$  et, si  $A$  est inversible,  $\mathbb{P}_{AX}(E) = \mathbb{P}_X(A^{-1}E)$ . En particulier, la translation ne modifie pas l'indépendance.

Si  $X_1, \dots, X_n$  sont indépendantes, alors elles sont deux à deux décorréllées puisque

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[\tilde{X}\tilde{Y}] = \int_{\mathbb{R}^2} xy \, dP_{(\tilde{X}, \tilde{Y})}(x, y) = \int_{\mathbb{R}} \int_{\mathbb{R}} xy \, dP_{\tilde{X}}(x) \, dP_{\tilde{Y}}(y) \\ &= \int_{\mathbb{R}} xy \, dP_{\tilde{X}}(x) \int_{\mathbb{R}} y \, dP_{\tilde{Y}}(y) = \mathbb{E}[\tilde{X}]\mathbb{E}[\tilde{Y}] = 0. \end{aligned}$$

Des variables indépendantes sont donc décorréllées.

Enfin, si  $X_1, \dots, X_n$  sont décorréllées, alors en utilisant l'expression en terme de produits scalaire, on voit immédiatement que

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$$

puisque

$$\begin{aligned} \text{Var}(X_1 + \dots + X_n) &= \left\| \tilde{X}_1 + \dots + \tilde{X}_n \right\|_{L^2}^2 = \left\langle \tilde{X}_1 + \dots + \tilde{X}_n, \tilde{X}_1 + \dots + \tilde{X}_n \right\rangle \\ &= \sum_{j,k} \left\langle \tilde{X}_j, \tilde{X}_k \right\rangle = \sum_j \left\langle \tilde{X}_j, \tilde{X}_j \right\rangle \\ &= \text{Var}(X_1) + \dots + \text{Var}(X_n). \end{aligned}$$

Notons que cette interprétation de la corrélation comme un produit scalaire a une autre conséquence. La matrice de corrélation  $\text{Cov}(X) = [\langle \tilde{X}_j, \tilde{X}_k \rangle]_{j,k}$ . Une matrice de la forme  $G = G(a_1, \dots, a_n) =$



$[\langle a_j, a_k \rangle]_{j,k=1,\dots,n}$  avec  $a_1, \dots, a_n \in \mathcal{H}$  un espace de Hilbert, est appelé *matrice de Gram* de  $a_1, \dots, a_n$  et est toujours positive au sens où  $\langle Gx, x \rangle_{\mathbb{R}^n} \geq 0$  pour tout  $x \in \mathbb{R}^n$ . En effet

$$\langle Gx, x \rangle = \sum_{j=1}^n (Gx)_j x_j = \sum_{j=1}^n \sum_{k=1}^n G_{j,k} x_k x_j = \sum_{j=1}^n \sum_{k=1}^n \langle a_j, a_k \rangle x_j x_k = \left\| \sum_{j=1}^n x_j a_j \right\|^2.$$

Notons que  $G$  est symétrique réelle, donc diagonalisable. Comme  $G$  est positive, ses valeurs propres sont positives ( $x$  vecteur propre  $Gx = \lambda x$  donc  $\lambda \|x\|^2 = \langle Gx, x \rangle \geq 0$ ). Ainsi, on peut écrire  $G = P\Delta P^t$  avec  $P$  une matrice orthogonale et

$$\Delta = \text{diag}(\delta_1, \dots, \delta_n) := \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_n \end{pmatrix} \quad \delta_1, \dots, \delta_n \geq 0.$$

Ainsi  $\Delta = \text{diag}(\delta_1^{1/2}, \dots, \delta_n^{1/2})^2$  et  $G = \Sigma^2 = \Sigma \Sigma^t$  avec  $\Sigma = P \text{diag}(\delta_1^{1/2}, \dots, \delta_n^{1/2}) P^t$ .

DÉFINITION 1.15. Soit  $X$  un vecteur aléatoire dans  $\mathbb{R}^n$ . Sa *fonction caractéristique* est la fonction définie sur  $\mathbb{R}^n$  par

$$\varphi_X(t) = \mathbb{E}[e^{i\langle t, X \rangle}] = \int_{\mathbb{R}^n} e^{i\langle t, x \rangle} d\mathbb{P}_X(x).$$

Plus généralement, si  $\mathbb{P}$  est une probabilité sur  $\mathbb{R}$ , sa fonction caractéristique est donnée par  $\varphi_{\mathbb{P}}(t) = \mathbb{E}[e^{it\cdot}] = \int e^{it\cdot} d\mathbb{P}(x)$ .

C'est donc la transformée de Fourier (inverse) de la mesure de probabilité  $\mathbb{P}_X$ . Elle caractérise donc la loi de  $X$ . Le théorème de Lebesgue montre immédiatement que  $\varphi_X$  est continue. De plus, la fonction caractéristique caractérise l'indépendance.

THÉORÈME 1.16. Soient  $X_1, \dots, X_n$  des variables aléatoires réelles et  $X = (X_1, \dots, X_n)$ . Alors  $X_1, \dots, X_n$  sont indépendants si et seulement si  $\varphi_X(t_1, \dots, t_n) = \varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n)$ .

DÉMONSTRATION. Supposons d'abord que les  $X_i$  sont indépendants. Alors

$$\begin{aligned} \varphi_X(t) &= \int_{\mathbb{R}^n} e^{i\langle t, x \rangle} d\mathbb{P}_X(t) = \int_{\mathbb{R}^n} e^{i(t_1 x_1 + \cdots + t_n x_n)} d\mathbb{P}_{(X_1, \dots, X_n)}(t_1, \dots, t_n) \\ &= \int_{\mathbb{R}^n} e^{it_1 x_1} \cdots e^{it_n x_n} d\mathbb{P}_{X_1}(t_1) \cdots d\mathbb{P}_{X_n}(t_n) = \int_{\mathbb{R}} e^{it_1 x_1} d\mathbb{P}_{X_1}(t_1) \cdots \int_{\mathbb{R}} e^{it_n x_n} d\mathbb{P}_{X_n}(t_n) \\ &= \varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n). \end{aligned}$$

Réciproquement, on vient de voir que la probabilité  $\mu = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$  a pour fonction caractéristique  $\varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n)$ . Si on suppose que

$$\varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n) = \varphi_X(t_1, \dots, t_n),$$

alors  $\mu = \mathbb{P}_X$  puisque la fonction caractéristique caractérise la mesure, donc  $X_1, \dots, X_n$  sont bien indépendants.  $\square$

#### 1.4. Exercices.

exo:repart

EXERCICE 1.3. Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{N}$  sur un espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$ . On note  $F_X(\lambda) = \mathbb{P}(X \leq \lambda)$  sa *fonction de répartition*. Vérifier que

- (i)  $F_X$  est croissante,  $\lim_{\lambda \rightarrow -\infty} F_X(\lambda) = 0$  et  $\lim_{\lambda \rightarrow +\infty} F_X(\lambda) = 1$ ;
- (ii)  $F_X$  est continue à droite et plus précisément

$$\lim_{\lambda \rightarrow \lambda_0^-} F_X(\lambda) = \mathbb{P}(X < \lambda_0)$$

- (iii)  $F_X$  détermine  $\mathbb{P}_X$ .

**exo:gener**

EXERCICE 1.4. Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{N}$  sur un espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$ . On note  $G_X(s) = \mathbb{E}(s^X) = \sum_{k \in \mathbb{N}} \mathbb{P}(X = k) s^k$  sa *fonction génératrice*.

- (i) Montrer que  $G_X$  est bien définie sur  $[-1, 1]$ .
- (ii) Montrer que  $G_X$  caractérise la loi de  $X$ .
- (iii) On suppose que  $X$  et  $X^2$  sont intégrables. On note  $G'_X$  et  $G''_X$  les dérivées de  $G_X$ . Montrer que

$$\mathbb{E}[X] = G'_X(1) \quad \text{et} \quad \mathbb{E}[X^2] = G'_X(1) + G''_X(1).$$

- (iv) en déduire l'expression de  $\text{Var}(X)$  en fonction de  $G_X$ .
- (v) Pour chaque une des principales lois discrètes, calculer la fonction génératrice correspondante et en déduire la moyenne et la variance.

**ex:lois**

EXERCICE 1.5. On considère une suite d'expériences indépendantes dont l'issue est un succès avec probabilité  $p$  et un échec avec probabilité  $1 - p$ .

- (1) Montrer que le nombre de succès parmi les  $n$  premières expériences suit une loi binomiale de paramètres  $n$  et  $p$ .
- (2) Montrer que l'instant où a lieu le premier succès suit une loi géométrique de paramètre  $p$ .

**ex:bool**

EXERCICE 1.6. Une variable aléatoire booléenne supposée vraie passe par  $n$  intermédiaires. Chaque intermédiaire transmet correctement l'information avec une probabilité  $p$  et se trompe avec une probabilité  $1 - p$ . Quelle est la probabilité  $p_n$  pour qu'on reçoive la bonne information. En déduire la limite de  $p_n$ .

**exo:unif**

EXERCICE 1.7. Soit  $X$  une variable aléatoire de loi uniforme sur  $]-\pi/2, \pi/2[$ . Quelle est la loi de  $Y = \tan X$ .

**exo:ind1**

EXERCICE 1.8. Soit  $(X, Y)$  un vecteur aléatoire de loi  $\mathbb{P}_{(X,Y)}$  ayant pour densité  $f(x, y) = \frac{2}{e-1} x e^y \mathbf{1}_{[0,1]^2}$ . Déterminer les densités de  $X$  et  $Y$ . Sont-elles indépendantes.

## 2. Variables aléatoires gaussiennes

**sec:normal**

DÉFINITION 1.17. Une variable aléatoire est dite gaussienne, s'il existe  $\sigma > 0$  et  $m \in \mathbb{R}$  tels que  $\mathbb{P}_X$  soit de densité

$$\gamma_{\sigma^2, m}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{\sigma^2}\right)$$

et on note  $X \sim \mathcal{N}(m, \sigma^2)$ .

Dans le cas  $\sigma = 0$ ,  $X = m$  presque partout, on parle de variable gaussienne dégénérée et on note  $X \sim \mathcal{N}(m, 0)$ .

**prop:gauss**

PROPOSITION 1.18. Soit  $X$  une variable aléatoire gaussienne  $X \sim \mathcal{N}(m, \sigma^2)$ . Alors

- (i)  $\mathbb{E}[X] = m$  et  $\text{Var}(X) = \sigma^2$ .
- (ii) La fonction caractéristique  $\varphi_X$  de  $X$  est donnée par  $\varphi_X(\omega) = e^{im\omega - \sigma^2\omega^2/2}$ .

DÉMONSTRATION. Rappelons que  $I := \int_{\mathbb{R}} e^{-t^2/2} dt = \sqrt{2\pi}$  ce qui peut se démontrer de la façon suivante

$$\begin{aligned} I^2 &= \int_{\mathbb{R}} e^{-s^2/2} ds \int_{\mathbb{R}} e^{-t^2/2} dt \\ &= \int_{\mathbb{R}^2} e^{-(t^2+s^2)/2} ds dt \quad \text{Fubini} \\ &= \int_0^{2\pi} \int_0^{+\infty} e^{-r^2/2} r dr d\theta \quad \text{passage en polaire} \\ &= 2\pi. \end{aligned}$$

Mais alors

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{\sigma^2}\right) dx = 1$$

avec le changement de variable  $t = (x-m)/\sigma$ .

Ensuite, le même changement de variable donne

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathbb{R}} x d\mathbb{P}_X(x) = \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-m)^2}{\sigma^2}\right) dx \\ &= \int_{\mathbb{R}} (\sigma t + m) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = m \end{aligned}$$

en utilisant le fait que  $te^{-t^2/2}$  est impaire donc  $\int_{\mathbb{R}} te^{-t^2/2} dt = 0$ .

Enfin,

$$\begin{aligned} \text{Var}(X) &= \int_{\mathbb{R}} (x-m)^2 d\mathbb{P}_X(x) = \int_{\mathbb{R}} (x-m)^2 \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-m)^2}{\sigma^2}\right) dx \\ &= \sigma^2 \int_{\mathbb{R}} t^2 e^{-t^2/2} dt = \sigma^2 \int_{\mathbb{R}} t (e^{-t^2/2})' dt \\ &= \sigma^2 \left( [-te^{-t^2/2}]_{-\infty}^{+\infty} + \int_{\mathbb{R}} e^{-t^2/2} dt \right) = \sigma^2. \end{aligned}$$

Finalement,

$$\begin{aligned} \varphi_X(\omega) &= \int_{\mathbb{R}} e^{ix\omega} d\mathbb{P}_X(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \int_{\mathbb{R}} \exp\left(ix\omega - \frac{(x-m)^2}{\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i(\sigma t + m)\omega - t^2/2} dt = \frac{e^{im\omega}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(t-i\sigma\omega)^2/2 - \sigma^2\omega^2} dt \\ &= e^{im\omega - \sigma^2\omega^2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(t-i\sigma\omega)^2/2} dt. \end{aligned}$$

Enfin, en utilisant la méthode des résidus, on voit aisément que

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(t-i\sigma\omega)^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2/2} dt = 1$$

d'où le résultat.  $\square$

**DÉFINITION 1.19.** Un vecteur aléatoire  $X = (X_1, \dots, X_n)$  est dit *gaussien* si, pour tout  $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ ,  $\langle a, X \rangle$  est gaussien *i.e.* si  $a_1 X_1 + \dots + a_n X_n$  est gaussien *i.e.* toute combinaison linéaire des composantes de  $X$  est gaussienne.

**REMARQUE 1.20.** Si  $X$  est gaussien, en prenant pour  $a$  les vecteurs de la base canonique  $a = e_j := (\delta_{j,k})_{k=1, \dots, n}$  on voit que si  $X$  est gaussien alors ses coordonnées sont gaussiennes. La réciproque est fautive.

Comme une variable gaussienne est de carré intégrable,  $\mathbb{E}[X_i^2] < +\infty$ , on peut définir les covariances  $\text{Cov}(X_j, X_k) = \mathbb{E}[(X_j - \mathbb{E}[X_j])(X_k - \mathbb{E}[X_k])]$  et donc la matrice de covariance de  $X$ .

Soit  $X$  un vecteur gaussien,  $m = \mathbb{E}[X] \in \mathbb{R}^n$  et  $\Sigma^2 = \text{Cov}(X)$ . Fixons  $a \in \mathbb{R}^n$ , identifions  $a$  et  $X$  à des matrices colonne  $(n, 1)$ . Alors, avec le lemme [1.12](#) appliqué à  $\langle a, X \rangle = a^t X$ ,  $\mathbb{E}[\langle a, X \rangle] = a^t \mathbb{E}[X] = \langle a, m \rangle$  et  $\text{Var}(\langle a, X \rangle) = a^t \Sigma^2 a$ . Comme la variable aléatoire  $X_a = \langle a, X \rangle$  est gaussienne, sa fonction caractéristique est donnée par la proposition [1.18](#):

$$\mathbb{E}[e^{i\omega \langle X, a \rangle}] = \varphi_{X_a}(\omega) = e^{i\omega \langle a, m \rangle} e^{-a^t \Sigma^2 a \omega^2 / 2}.$$

Mais, en prenant  $\omega = 1$ , on trouve

$$\varphi_X(a) = \mathbb{E}[e^{i \langle X, a \rangle}] = e^{i \langle a, m \rangle} e^{-a^t \Sigma^2 a / 2}.$$

Nous venons donc de démontrer la proposition:

**prop:vecgauss**

PROPOSITION 1.21. Si  $X$  est un vecteur gaussien, de moyenne  $m$  et de matrice de covariance  $\Sigma^2$ , sa fonction caractéristique est donnée par

$$\varphi_X(a) = \mathbb{E}[e^{i\langle X, a \rangle}] = e^{i\langle a, m \rangle} e^{-a^t \Sigma^2 a / 2}.$$

Ainsi, (la loi d')un vecteur gaussien est entièrement caractérisé par sa moyenne  $m$  et sa matrice de covariance  $\Sigma^2$  puisque sa fonction caractéristique l'est. Nous noterons  $X \sim \mathcal{N}(m, \Sigma^2)$ .

COROLLAIRE 1.22. Soient  $X_1, \dots, X_n$  des variables gaussiennes indépendantes,  $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$ , alors  $X = (X_1, \dots, X_n)$  est un vecteur gaussien

$$X \sim \mathcal{N}((m_1, \dots, m_n), \text{diag}(\sigma_1^2, \dots, \sigma_n^2)).$$

En particulier, toute combinaison linéaire de variables gaussiennes indépendantes est encore une variable gaussienne. Plus précisément  $x_1 X_1 + \dots + x_n X_n$  est gaussienne et

**eq:gausscomb**

$$(1) \quad x_1 X_1 + \dots + x_n X_n \sim \mathcal{N}\left(\sum_{i=1}^n x_i m_i, \sum_{i=1}^n x_i^2 \sigma_i^2\right)$$

Inversément, si  $X = (X_1, \dots, X_n)$  est un vecteur gaussien, de moyenne  $m = (m_1, \dots, m_n)$  et de matrice de covariance diagonale  $\text{Cov}(X) = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ , alors  $X_1, \dots, X_n$  sont gaussiennes indépendantes avec  $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$ .

REMARQUE 1.23. Il résulte de ce corollaire que l'indépendance de variables aléatoires gaussiennes est équivalente à leur décorrélation.

DÉMONSTRATION. En effet, si  $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$ , alors  $\varphi_{X_i}(\omega) = e^{im_i \omega} e^{-\sigma_i^2 \omega^2 / 2}$ . Mais, comme les  $X_i$  sont indépendants

$$\begin{aligned} \varphi_X(\omega_1, \dots, \omega_n) &= \prod_{i=1}^n \varphi_{X_i}(\omega_i) = \prod_{i=1}^n e^{im_i \omega_i} e^{-\sigma_i^2 \omega_i^2 / 2} \\ &= e^{i\langle m, \omega \rangle} e^{-\omega^t \Sigma^2 \omega} \end{aligned}$$

avec  $m = (m_1, \dots, m_n)$ ,  $\omega = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_n \end{pmatrix}$  et  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ . Comme la fonction caractéristique est celle d'une gaussienne,  $X$  est bien un vecteur gaussien.

La formule (1) résulte directement du lemme 1.12 avec  $A = (x_1 \ \dots \ x_n)$ .

Pour la réciproque, le calcul ci-dessus montre que si  $\text{Cov}(X)$  est diagonale, alors la fonction caractéristique de  $X$  est le produit des fonctions caractéristiques de ses coordonnées, celles-ci sont donc indépendantes.  $\square$

COROLLAIRE 1.24. Soit  $X$  un vecteur gaussien  $X \sim \mathcal{N}(m, \Sigma)$ ,  $m' \in \mathbb{R}^p$ ,  $A \in \mathcal{M}_{p,n}$ , alors  $AX + m'$  est un vecteur gaussien avec  $AX + m' \sim \mathcal{N}(Am + m', A\Sigma A^t)$ .

En particulier, si  $m \in \mathbb{R}^n$  et  $\Sigma^2$  une matrice symétrique positive, alors il existe un vecteur gaussien  $X \sim \mathcal{N}(m, \Sigma^2)$ .

DÉMONSTRATION. Tout d'abord,  $AX$  est bien gaussien car ses coordonnées sont des combinaisons linéaires des  $X_i$  donc les combinaisons linéaires de ces coordonnées sont des combinaisons linéaires des  $X_i$ . Elles sont donc gaussiennes. Comme les constantes sont gaussiennes et indépendantes de  $AX$  (de toute variable aléatoire),  $AX + m'$  est encore une gaussienne.

La moyenne et la variance de  $AX + m'$  est donnée par le lemme 1.12.

Enfin, une matrice symétrique positive s'écrit  $\Sigma^2 = AA^t$  avec  $A$  une matrice  $(n, n)$ . Il suffit donc de prendre  $X_1, \dots, X_n$  des variables  $\mathcal{N}(0, 1)$  indépendantes et  $Y = A(X_1, \dots, X_n) + m$ .  $\square$

DÉFINITION 1.25. Un espace gaussien est un sous-espace fermé de  $L^2(\Omega, \mathcal{A}, \mathbb{P})$  constitué de variables aléatoires gaussiennes.

Un vecteur est donc gaussien si l'espace engendré par ses coordonnées est gaussien. Ainsi, si  $X_1, \dots, X_n$  sont gaussiennes centrées et indépendantes, l'espace qu'elles engendrent est un espace gaussien et  $X_1, \dots, X_n$  est une base orthogonale de cet espace. Nous allons maintenant montrer une réciproque. Pour cela, nous aurons besoin de la définition suivante:

**DÉFINITION 1.26.** Soit  $\{X_n\}_{n \in \mathbb{Z}}$  une suite de variables aléatoires dans l'espace  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ . On dira que cette suite est *stochastiquement indépendante* si, pour tout entier  $n \in \mathbb{N}$  et tous  $(k_1, \dots, k_n) \in \mathbb{Z}^n$ , les variables  $X_{k_1}, \dots, X_{k_n}$  sont indépendantes.

On a alors une réciproque à l'exemple précédent.

**THÉORÈME 1.27.** Soit  $G \subset L^2(\Omega, \mathcal{A}, \mathbb{P})$  un espace gaussien formé de variables centrées. Alors il existe une base orthonormale  $\{X_j\}_{j \in \mathbb{N}}$  de  $G$  avec  $X_j \sim \mathcal{N}(0, 1)$  et  $\{X_j\}_{j \in \mathbb{N}}$  stochastiquement indépendante.

**DÉMONSTRATION.**  $G$  étant un sous-espace fermé d'un espace de Hilbert,  $G$  admet une base orthonormée. Nous allons montrer que toute base orthonormée  $\{X_j\}_{j \in \mathbb{N}}$  répond à la question. Pour cela, il suffit de remarquer que les  $X_j$  sont centrées puisque  $G$  ne contient que des variables centrées, donc le produit scalaire de  $L^2(\Omega, \mathcal{A}, \mathbb{P})$  coïncide avec la covariance:  $\delta_{j,k} = \langle X_j, X_k \rangle = \text{Cov}(X_j, X_k)$ . Ainsi, si on extrait une sous-suite finie  $X_{k_1}, \dots, X_{k_n}$  de  $\{X_n\}$ , le vecteur  $X = (X_{k_1}, \dots, X_{k_n})$  est gaussien puisque  $G$  est gaussien et sa matrice de covariance est l'identité, donc  $X_{k_1}, \dots, X_{k_n}$  sont gaussiennes indépendantes  $\mathcal{N}(0, 1)$ .  $\square$

**PROPOSITION 1.28.** Soit  $X_1, \dots, X_n$  des variables aléatoires indépendantes de loi  $\mathcal{N}(m, \sigma^2)$  avec  $\sigma > 0$ . Les variables aléatoires

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k \quad \text{et} \quad S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

sont indépendantes. De plus  $\bar{X}$  suit la loi  $\mathcal{N}(m, \sigma^2/n)$  et  $(n-1)s^2/\sigma^2$  suit la loi du  $\chi^2$  à  $n-1$  degrés de liberté, notée  $\chi^2(n-1)$  et dont la densité est

$$\frac{1}{2^{n/2} \Gamma(n/2)} y^{n/2-1} e^{-y/2} \mathbf{1}_{y>0}.$$

**REMARQUE 1.29.**  $(n-1)s^2/\sigma^2$  a même loi que  $\sum_{k=1}^n Z_k$  où les  $Z_k$  sont des variables aléatoires indépendantes de loi  $\mathcal{N}(0, 1)$ .

exo:gauss1

**EXERCICE 1.9.** Soit  $X$  une variable aléatoire de loi  $\mathcal{N}(0, 1)$ . On pose

$$Y = \begin{cases} X & \text{si } |X| \leq 1 \\ -X & \text{si } |X| \geq 1 \end{cases}.$$

Montrer que  $Y \sim \mathcal{N}(0, 1)$  mais le vecteur  $\begin{pmatrix} X \\ Y \end{pmatrix}$  n'est pas gaussien.

exo:gauss2

**EXERCICE 1.10.** Soient  $X$  et  $Y$  deux variables aléatoires gaussiennes indépendantes. Montrer que  $X + Y$  et  $X - Y$  sont indépendantes si et seulement si  $\text{Var}(X) = \text{Var}(Y)$ .

exo:gauss3

**EXERCICE 1.11.** Soit  $X \sim \mathcal{N}(0, 1)$  et  $\varepsilon$  une variable aléatoire indépendante de  $X$  à valeurs dans  $\{-1, 1\}$  telle que  $\mathbb{E}[\varepsilon] = 0$ . Montrer que  $X$  et  $\varepsilon X$  sont gaussiennes, orthogonales (décorrélées), mais pas indépendantes.

**EXERCICE 1.12.** Montrer que, pour tous  $\beta, \alpha > 0$ , la fonction

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}_{[0, +\infty)}(x)$$

est une densité de probabilité. La loi qui lui est associée sera notée  $\Gamma(\alpha, \beta)$ .

Montrer par récurrence sur  $n$  que la loi  $\chi_n^2$  coïncide avec la loi  $\Gamma(n/2, 1/2)$ .

### 3. Convergence de variables aléatoires

#### 3.1. Définitions.

DÉFINITION 1.30. Soient  $X_n$  une suite de v. a. r. et  $X$  une v. a. r.

(i) Soient  $\varphi_n(t) = \mathbb{E}[e^{itX_n}]$  et  $\varphi(t) = \mathbb{E}[e^{itX}]$  les fonctions caractéristiques des  $X_n$  et de  $X$ . On dit que  $X_n$  tend en loi vers  $X$  et on note  $X_n \xrightarrow{\mathcal{L}} X$  si  $\varphi_n$  tend simplement vers  $\varphi$  i.e. si pour tout  $t \in \mathbb{R}$ ,  $\varphi_n(t) \rightarrow \varphi(t)$ .

(ii) On dit que  $X_n$  tend en probabilités vers  $X$  et on note  $X_n \xrightarrow{\mathcal{P}} X$  si, pour tout  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\{|X_n - X| \geq \varepsilon\}) = 0.$$

(iii) On dit que  $X_n$  tend presque sûrement vers  $X$  et on note  $X_n \xrightarrow{p.s.} X$  si,

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\}) = 1.$$

(iv) On dit que  $X_n$  tend en moyenne d'ordre  $p \geq 1$  vers  $X$  et on note  $X_n \xrightarrow{L^p} X$  si,

$$\mathbb{E}[|X_n - X|^p] \rightarrow 0 \quad \text{quand } n \rightarrow +\infty.$$

C'est la convergence dans  $L^p(\Omega, \mathcal{A}, \mathbb{P})$ .

PROPOSITION 1.31. On a les implications suivantes (les réciproques sont fausses)

$$\left. \begin{array}{l} \text{Conv. p.s.} \\ \text{Conv. } L^1 \end{array} \right\} \Rightarrow \text{Conv. } \mathcal{P} \Rightarrow \text{Conv. } \mathcal{L}.$$

#### 3.2. Loi des grands nombres.

THÉORÈME 1.32. Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires et notons  $S_n = \frac{1}{n} \sum_{j=1}^n X_j$ . Supposons que  $\mathbb{E}[X_n]$  et  $\text{Var}(X_n)$  soient constantes,  $\mathbb{E}[X_n] = \lambda$ ,  $\text{Var}(X_n) = \sigma^2$ . Alors  $S_n$  converge en probabilités (loi faible) et presque sûrement (loi forte) vers  $\lambda$  quand  $n$  tend vers l'infini.

Bien sûr, la loi forte implique la loi faible. Toutefois, nous ne démontrerons que la loi faible qui est bien plus simple.

Tout d'abord, rappelons l'inégalité de Bienaymé-Tchebichev:

LEMME 1.33 (Bienaymé-Tchebichev). Soit  $X$  une variable aléatoire intégrable, i.e. telle que  $\mathbb{E}[|X|] < +\infty$ . Alors, pour tout  $a > 0$ ,

eq:markov

$$(2) \quad \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[|X|]}{a}.$$

Si elle est de carré intégrable, i.e.  $\mathbb{E}[X^2] < +\infty$  alors, pour tout  $a > 0$ ,

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

L'inégalité <sup>eq:markov</sup>(8) est aussi appelée *inégalité de Markov*.

DÉMONSTRATION DE LEMME. Pour la première partie, on a

$$\begin{aligned} \mathbb{P}(X \geq a) &= \mathbb{P}(\mathbf{1}_{\{|X| \geq a\}}) = \int_{\Omega} \mathbf{1}_{\{|X| \geq a\}}(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} \mathbf{1}_{\{|\omega| \geq a\}} d\mathbb{P}_X(\omega) \\ &\leq \int_{\mathbb{R}} \frac{|\omega|}{a} \mathbf{1}_{\{|\omega| \geq a\}} d\mathbb{P}_X(\omega) \leq \int_{\mathbb{R}} \frac{|\omega|}{a} d\mathbb{P}_X(\omega) \\ &= \frac{1}{a} \mathbb{E}[|X|]. \end{aligned}$$

Pour la seconde partie du lemme, remarquons que Cauchy-Schwarz implique  $\mathbb{E}[|X|] \leq \mathbb{E}[X^2]^{1/2}$  et que

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 < +\infty$$

donc  $Y = (X - \mathbb{E}[X])^2$  vérifie les hypothèses de la première partie du lemme et  $\mathbb{E}[|Y|] = \text{Var}(X)$  est bien définie. Ainsi

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) = \mathbb{P}(Y \geq a^2) \leq \frac{\mathbb{E}[|Y|]}{a^2} = \frac{\text{Var}(X)}{a^2}$$

comme annoncé. □

DÉMONSTRATION DE LA LOI FAIBLE DES GRANDS NOMBRES. Tout d'abord

$$\mathbb{E}\left[\frac{1}{n} \sum_{j=1}^n X_j\right] = \frac{1}{n} \sum_{j=1}^n \mathbb{E}[X_j] = \lambda.$$

Ensuite, comme les  $X_j$  ne sont pas corrélées, les  $\frac{1}{n}X_j$  non plus, donc

$$\text{Var}\left(\frac{1}{n} \sum_{j=1}^n X_j\right) = \sum_{j=1}^n \text{Var}\left(\frac{1}{n}X_j\right) = \frac{1}{n^2} \sum_{j=1}^n \text{Var}(X_j) = \frac{\sigma^2}{n}.$$

En utilisant Bienaimé-Tchebichev, il vient alors

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n} \sum_{j=1}^n X_j - \lambda\right| \geq \varepsilon\right) &= \mathbb{P}\left(\left|\frac{1}{n} \sum_{j=1}^n X_j - \mathbb{E}\left[\frac{1}{n} \sum_{j=1}^n X_j\right]\right| \geq \varepsilon\right) \\ &\leq \frac{1}{\varepsilon^2} \text{Var}\left(\frac{1}{n} \sum_{j=1}^n X_j\right) = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \end{aligned}$$

lorsque  $n \rightarrow \infty$ . □

#### EXEMPLE 1.34. Grande déviation

Soient  $X_i$  des variables de Bernoulli  $p$  indépendantes *i.e.*  $\mathbb{P}(X_i = 0) = 1 - p$  et  $\mathbb{P}(X_i = 1) = p$ . Alors (voir Exercice 1.1)  $\mathbb{E}[X_i] = p$ ,  $\text{Var}(X_i) = p(1 - p)$  et  $S_n = \sum_{i=1}^n X_i$  est le nombre de 1 parmi  $X_1, \dots, X_n$ .

Notons que

$$\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| \geq \varepsilon \Leftrightarrow \left|\sum_{i=1}^n X_i - np\right| \geq n\varepsilon \Leftrightarrow |\{X_i = 1\}| \notin [np - n\varepsilon, np + n\varepsilon].$$

La loi des grands nombre nous dit que le nombre de 1 après  $n$  tirages est dans l'intervalle  $[np - n\varepsilon, np + n\varepsilon]$  avec grande probabilité (il faut toutefois que  $n\varepsilon \rightarrow 0$  pour que cette probabilité tende vers 0).

Par exemple, pour  $p = 1/2$ , on obtient

$$\mathbb{P}[|\{X_i = 1\}| \notin [np - n\varepsilon, np + n\varepsilon]] = \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}.$$

Cette estimation est pessimiste et on peut faire bien mieux: avec bienaymé-Tchebichev, on obtient pour  $\lambda > 0$

$$\begin{aligned}
\mathbb{P}\left(S_n - \frac{n}{2} \geq n\varepsilon\right) &= \mathbb{P}\left(e^{\lambda(S_n - \frac{n}{2})} \geq e^{\lambda n\varepsilon}\right) \leq \frac{1}{e^{\lambda n\varepsilon}} \mathbb{E}\left[e^{\lambda(S_n - \frac{n}{2})}\right] \\
&= \frac{1}{e^{\lambda n\varepsilon}} \mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^n (X_i - 1/2)\right)\right] = \frac{1}{e^{\lambda n\varepsilon}} \mathbb{E}\left[\prod_{i=1}^n \exp(\lambda(X_i - 1/2))\right] \\
&= \frac{1}{e^{\lambda n\varepsilon}} \prod_{i=1}^n \mathbb{E}[\exp(\lambda(X_i - 1/2))] \quad \text{par indépendance} \\
&= \frac{1}{e^{\lambda n\varepsilon}} \prod_{i=1}^n \left(e^{-\lambda/2} \mathbb{P}(X_i = 0) + e^{\lambda/2} \mathbb{P}(X_i = 1)\right) \\
&= \frac{(\cosh h)^n}{e^{\lambda n\varepsilon}} = e^{-n(\lambda\varepsilon - \ln \cosh \lambda)}.
\end{aligned}$$

Notons  $f(\lambda) = \lambda\varepsilon - \ln \cosh \lambda$ . On voit facilement que  $f$  est minimal lorsque  $\tanh \lambda/2 = \varepsilon$  et alors  $f(\lambda) = (\frac{1}{2} + \varepsilon) \ln(1 + 2\varepsilon) + (\frac{1}{2} - \varepsilon) \ln(1 - 2\varepsilon)$ . On en déduit que

$$\mathbb{P}\left(S_n - \frac{n}{2} \geq n\varepsilon\right) \leq (1 + 2\varepsilon)^{n(\frac{1+2\varepsilon}{2})} (1 - 2\varepsilon)^{n(\frac{1-2\varepsilon}{2})}.$$

Une estimation similaire est valable pour  $\mathbb{P}(S_n - \frac{n}{2} \leq -n\varepsilon)$ . Comme les évènements  $S_n - \frac{n}{2} \leq -n\varepsilon$  et  $S_n - \frac{n}{2} \geq n\varepsilon$  sont disjoints, on en déduit que

$$\mathbb{P}\left(\left|S_n - \frac{n}{2}\right| \geq n\varepsilon\right) \leq 2(1 + 2\varepsilon)^{n(\frac{1+2\varepsilon}{2})} (1 - 2\varepsilon)^{n(\frac{1-2\varepsilon}{2})}.$$

Cette estimation est bien meilleure que celle donnée par le théorème central limite. Par exemple, avec  $n = 1000$  et  $\varepsilon = 0.1$ , cette estimation donne  $\mathbb{P}(S_{1000} \notin [400, 600]) \leq 3.610^{-9}$  contre 0,025 avec la loi des grands nombres.

### 3.3. Théorème central limite.

**THÉORÈME 1.35.** *Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires définies sur un même espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$  et à valeurs dans  $\mathbb{R}^n$ . On les suppose de plus indépendantes, de même loi et admettant un moment d'ordre deux  $\mathbb{E}[\|X\|^2] < +\infty$ . Alors la suite*

$$Y_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \mathbb{E}[X_j])$$

converge en loi vers une v.a. de loi normale  $\mathcal{N}(0, \text{Cov}(X_1))$ .

Pour démontrer ce théorème, nous aurons besoin du lemme suivant:

**LEMME 1.36.** *Si la variable aléatoire  $X$  admet un moment d'ordre deux, sa fonction caractéristique admet un développement limité à l'ordre deux en 0 donné par*

eq: c1lemme

$$(3) \quad \varphi_X(t) = 1 + it\mathbb{E}[X] - \frac{t^2}{2}\mathbb{E}[X^2] + o(t^2).$$

**DÉMONSTRATION DU THÉORÈME.** Les variables aléatoires  $X_j$  étant indépendantes, la fonction caractéristique de  $Y_n$  est donnée par

$$\varphi_{Y_n}(u) = \mathbb{E}[e^{iuY_n}] = \mathbb{E}\left[\prod_{j=1}^n e^{i\langle u, X_j - \mathbb{E}[X_j] \rangle / \sqrt{n}}\right] = \prod_{j=1}^n \mathbb{E}\left[e^{i\langle u, X_j - \mathbb{E}[X_j] \rangle / \sqrt{n}}\right] = \prod_{j=1}^n \varphi_{\langle u, X_j - \mathbb{E}[X_j] \rangle}(1/\sqrt{n}).$$

Comme les  $X_j$  ont même loi, les  $\langle u, X_j - \mathbb{E}[X_j] \rangle$  aussi, donc

$$\varphi_{Y_n}(u) = \left(\varphi_{\langle u, X_1 - \mathbb{E}[X_1] \rangle}\left(\frac{1}{\sqrt{n}}\right)\right)^n.$$



Mais  $\mathbb{E}[\langle u, X_1 - \mathbb{E}[X_1] \rangle] = \langle u, \mathbb{E}[X_1 - \mathbb{E}[X_1]] \rangle = 0$  et  $\mathbb{E}[\langle u, X_1 - \mathbb{E}[X_1] \rangle^2] = \langle u, \text{Cov}(X_1)u \rangle$ . Il en résulte que

$$\varphi_{Y_n}(u) = \left(1 - \frac{1}{2n} \langle u, \text{Cov}(X_1)u \rangle o(n^{-1})\right)^n \rightarrow \exp\left(-\frac{1}{2} \langle u, \text{Cov}(X_1)u \rangle\right)$$

ce qui donne le résultat.  $\square$

**DÉMONSTRATION DU LEMME.** On utilise la formule de Taylor avec reste intégral qui n'est qu'une suite d'intégrations par parties du théorème fondamental de l'intégration:

$$\begin{aligned} e^{itx} &= 1 + \int_0^x ite^{its} ds = 1 + it[(s-x)e^{its}]_{s=0}^{s=x} - (it)^2 \int_0^x (s-x)e^{its} ds \\ &= 1 + itx + t^2 \int_0^x (s-x)e^{its} ds = 1 + itx + t^2 x^2 \int_0^1 (u-1)e^{itxu} du \end{aligned}$$

avec le changement de variable  $s = xu$ . On en déduit

$$e^{itx} = 1 + itx - \frac{1}{2}t^2x^2 + t^2x^2 \int_0^1 (u-1)(e^{itxu} - 1) du$$

donc

$$\boxed{\text{eq:c1fub}} \quad (4) \quad e^{itX(\omega)} = 1 + itX(\omega) - \frac{1}{2}t^2X(\omega)^2 + t^2 \int_0^1 (u-1)(e^{itX(\omega)u} - 1)X(\omega)^2 du.$$

Notons que

$$\boxed{\text{eq:c1leb}} \quad (5) \quad \left| (u-1)(e^{itX(\omega)u} - 1)X(\omega)^2 \right| \leq 2X(\omega)^2$$

on peut donc intégrer en  $\omega$  par rapport à la mesure  $\mathbb{P}$  pour obtenir, avec Fubini,

$$\boxed{\text{eq:c1}} \quad (6) \quad \varphi_X(t) = 1 + it\mathbb{E}[X] - \frac{1}{2}t^2\mathbb{E}[X^2] + t^2 \int_0^1 \int_{\Omega} (u-1)(e^{itX(\omega)u} - 1)X(\omega)^2 du d\mathbb{P}(\omega).$$

Mais, pour presque tout  $\omega$ ,  $|X(\omega)| < +\infty$  donc  $e^{itX(\omega)u} - 1 \rightarrow 0$  lorsque  $t \rightarrow 0$ . Avec l'aide de  $\boxed{\text{eq:c1leb}}$ , on applique le théorème de convergence dominée pour obtenir

$$\int_0^1 \int_{\Omega} (u-1)(e^{itX(\omega)u} - 1) du d\mathbb{P}(\omega) \rightarrow 0$$

lorsque  $t \rightarrow 0$ . On peut donc bien réécrire  $\boxed{\text{eq:c1}}$  sous la forme  $\boxed{\text{eq:c1lemme}}$ .  $\square$

#### 4. Espérance conditionnelle

Dans tout ce chapitre,  $(\Omega, \mathcal{A}, \mathbb{P})$  sera un espace de probabilité

##### 4.1. Loi conditionnelle et espérance conditionnelle: définitions.

4.1.1. *Cas discret.* Soit  $X : \Omega \rightarrow \mathcal{F}$  et  $Y : \Omega \rightarrow \mathcal{F}$  deux variables aléatoires *discrètes*. Notons  $p_X$  la loi de  $X$  et  $p_{X,Y}$  la loi du couple  $(X, Y)$ .

**DÉFINITION 1.37.** La *loi conditionnelle* de  $Y$  sachant que  $X = x$ , notée  $p_{Y|X}(\cdot|x)$  est définie par

$$\forall y \in \mathcal{F}, \quad p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)} = \frac{\mathbb{P}(\{X=x\} \cap \{Y=y\})}{\mathbb{P}(\{X=x\})}$$

pour chaque  $x \in \mathcal{F}$  tel que  $p_X(x) > 0$ .

Les réels  $\{p_{Y|X}(y|x)\}$  définissent une probabilité sur  $\mathcal{F}$ , représentant la distribution de  $Y$  sachant que  $X = x$ .

La loi conditionnelle n'est pas définie pour des valeurs de  $x$  telles que  $p_X(x) = \mathbb{P}(\{X=x\}) = 0$ , ce qui n'est pas gênant dans la mesure où l'évènement  $\{X=x\}$  se produit alors avec probabilité nulle.

DÉFINITION 1.38. Dans le cas où  $\mathcal{F} \subset \mathbb{R}$  et sous réserve de sommabilité, l'espérance de  $Y$  pour la loi conditionnelle  $\mathbb{P}(Y = \cdot | X = x)$  vaut

$$\mathbb{E}(Y|X = x) = \sum_{y \in \mathcal{F}} y p_{Y|X}(y|x)$$

et est appelée *espérance conditionnelle* de  $Y$  sachant que  $X = x$ . La variable aléatoire (fonction) *psi* définie par  $\psi(x) = \mathbb{E}(Y|X = x)$  est appelée *espérance conditionnelle de  $Y$  sachant  $X$* .

4.1.2. *Cas continu.* Soit  $(X, Y)$  un couple de variables aléatoires de densité  $p_{X,Y}$  sur  $\mathbb{R}^2$ . Soit  $I$  un intervalle de  $\mathbb{R}$ . Si  $p_X(x) > 0$ , on peut écrire pour  $\Delta > 0$  (et  $\Delta \rightarrow 0$ ):

$$\begin{aligned} \mathbb{P}(Y \in I | X \in [x, x + \Delta]) &= \frac{\mathbb{P}(\{Y \in I\} \cap \{X \in [x, x + \Delta]\})}{\mathbb{P}(\{X \in [x, x + \Delta]\})} \\ &\simeq \frac{\int_I p_{X,Y}(x, y) dy \Delta}{p_X(x) \Delta} \\ &= \int_I \frac{p_{X,Y}(x, y)}{p_X(x)} dy. \end{aligned}$$

Cela justifie la définition suivante:

DÉFINITION 1.39. Soit  $X$  une variable aléatoire sur  $\mathbb{R}^n$  et  $Y$  une variable aléatoire sur  $\mathbb{R}^d$ . On suppose que le couple  $(X, Y)$  a une loi de densité  $p_{X,Y}$ . Soit  $x$  tel que  $p_X(x) > 0$ . On appelle *densité conditionnelle de  $Y$  sachant que  $X = x$*  la fonction  $p_{Y|X}(\cdot|x)$  définie par

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

La *loi conditionnelle* de  $Y$  sachant  $X$  est la loi de probabilité sur  $\mathbb{R}^d$  ayant  $p_{Y|X}(\cdot|x)$  pour densité.

Enfin, l'*espérance conditionnelle* de  $Y$  sachant que  $X$  est la variable aléatoire  $\mathbb{E}(Y|X)$  définie sur  $\mathbb{R}^n$  par

$$x \rightarrow \mathbb{E}(Y|X = x) := \int_{\mathbb{R}^d} y p_{Y|X}(y|x) dy.$$

Si  $g$  est une fonction suffisamment régulière, alors

$$\mathbb{E}(g(Y)|X = x) = \int_{\mathbb{R}^d} g(y) p_{Y|X}(y|x) dy.$$

EXERCICE 1.13. Considérons le couple  $(X, Y)$  de loi jointe  $p_{X,Y}(x, y) = \frac{1}{x} \mathbf{1}_{0 < y < x < 1}$ .

Déterminer  $p_X$ ,  $p_{Y|X}$  et  $\mathbb{E}(Y|X)$ .

**4.2. Cas général.** Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité.

DÉFINITION 1.40. Soit  $X$  une variable aléatoire *intégrable i.e.*  $\mathbb{E}[|X|] < +\infty$ . Soit  $B \in \mathcal{A}$  avec  $\mathbb{P}(B) > 0$ . On appelle *espérance conditionnelle de  $X$  sachant  $B$*  la quantité

$$\mathbb{E}(X|B) = \int X d\mathbb{P}(\cdot|B) = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P}.$$

REMARQUE 1.41. Si  $X$  est intégrable, alors  $\frac{1}{\mathbb{P}(B)} \int_B |X| d\mathbb{P} < +\infty$ ,  $\mathbb{E}(X|B)$  est donc bien défini. Par ailleurs, si  $X = \mathbf{1}_A$  alors  $\mathbb{E}(\mathbf{1}_A|B) = \mathbb{P}(A|B)$ .

Supposons maintenant que  $B \in \mathcal{A}$  vérifie  $\mathbb{P}(B) > 0$  et  $\mathbb{P}(B^c) > 0$  (soit, de façon équivalente,  $0 < \mathbb{P}(B) < 1$ ). On définit  $\mathcal{B} = \sigma(B) = \{B, B^c, \emptyset, \Omega\}$ , la tribu engendrée par  $B$ . Une variable aléatoire  $Y$  est  $\mathcal{B}$ -mesurable. Alors, pour tout  $x \in \mathbb{R}$ ,  $Y^{-1}(\{x\}) \in \{B, B^c, \emptyset, \Omega\}$ . Il existe au moins un  $x$  tel que  $Y^{-1}(\{x\}) \neq \emptyset$ , on distingue alors 3 cas

–  $Y^{-1}(\{x\}) = \Omega$  et alors  $Y(\omega) = x$  pour tout  $\omega \in \Omega$ ;

–  $Y^{-1}(\{x\}) = B$  donc pour  $\omega \in B$ ,  $Y(\omega) = x$ . Soit alors  $\omega_0 \in B^c$ , alors  $y := Y(\omega_0) \neq x$  (sinon  $\omega_0 \in B$ ) donc  $Y^{-1}(\{y\}) \cap B = \emptyset$  et  $Y^{-1}(\{y\}) \neq \emptyset$  (car  $\omega_0 \in Y^{-1}(\{y\})$ ) donc  $Y^{-1}(\{y\}) = B^c$  et alors  $Y(\omega) = x\mathbf{1}_B + y\mathbf{1}_{B^c}$ ;

–  $Y^{-1}(\{x\}) = B$ , il existe alors  $y$  tel que  $Y(\omega) = y\mathbf{1}_B + x\mathbf{1}_{B^c}$ .

Soit  $X$  une variable aléatoire telle que  $\mathbb{E}[|X|^2] < +\infty$ .

Parmi tous les  $Y$   $\mathcal{B}$ -mesurable, cherchons maintenant celui qui est le plus proche de  $X$  au sens de la norme  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ . On veut donc minimiser sur les  $x, y \in \mathbb{R}$

$$\begin{aligned} \mathbb{E}[(X - x\mathbf{1}_B + y\mathbf{1}_{B^c})^2] &= \mathbb{E}[(X - x)\mathbf{1}_B + (X - y)\mathbf{1}_{B^c}]^2 \\ &= \mathbb{E}[(X - x)^2\mathbf{1}_B] + \mathbb{E}[(X - y)^2\mathbf{1}_{B^c}] \\ &= \mathbb{P}(B) \frac{1}{\mathbb{P}(B)} \int_B (X(\omega) - x)^2 d\mathbb{P}(\omega) + \mathbb{P}(B^c) \frac{1}{\mathbb{P}(B^c)} \int_{B^c} (X(\omega) - y)^2 d\mathbb{P}(\omega). \end{aligned}$$

où on a utilisé le fait que  $\mathbf{1}_E^2 = \mathbf{1}_E$  et que  $\mathbf{1}_B\mathbf{1}_{B^c} = 0$ . Il suffit donc de minimiser

$$\frac{1}{\mathbb{P}(B)} \int_B (X(\omega) - x)^2 d\mathbb{P}(\omega) \quad \text{et} \quad \frac{1}{\mathbb{P}(B^c)} \int_{B^c} (X(\omega) - y)^2 d\mathbb{P}(\omega).$$

Mais un calcul simple montre que ce minimum est atteint pour

$$x = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P} = \mathbb{E}[X|B] \quad \text{et} \quad y = \frac{1}{\mathbb{P}(B^c)} \int_{B^c} X d\mathbb{P} = \mathbb{E}[X|B^c].$$

On note donc

$$E[X, \mathcal{B}] = \mathbb{E}[X|B]\mathbf{1}_B + \mathbb{E}[X|B^c]\mathbf{1}_{B^c}$$

la projection orthogonale de  $X$  sur  $L^2(\Omega, \mathcal{B}, \mathbb{P})$  l'espace des fonctions  $\mathcal{B}$ -mesurables de carré intégrable.

De façon générale, si  $\mathcal{B}$  est une sous-tribu de  $\mathcal{A}$ , alors  $L^2(\Omega, \mathcal{B}, \mathbb{P})$  est un sous-espace vectoriel fermé de l'espace de Hilbert  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ . On peut donc définir la projection orthogonale sur ce sous-espace.

**DÉFINITION 1.42.** Soit  $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ . L'espérance condition de  $X$  par rapport à  $\mathcal{B}$  est la projection orthogonale de  $X$  sur  $L^2(\Omega, \mathcal{B}, \mathbb{P})$  et on la note  $\mathbb{E}[X|\mathcal{B}]$ .

Dans le cas où  $\mathcal{B}$  est la tribu engendrée par un vecteur aléatoire  $Y = (Y_1, \dots, Y_n)$ , c'est-à-dire quand  $\mathcal{B} = \{Y^{-1}(U) \mid U \text{ borélien de } \mathbb{R}^n\}$ , on la note  $\mathbb{E}[X|Y]$ .

L'espérance conditionnelle vérifie les propriétés suivantes:

- (a)  $X \rightarrow \mathbb{E}[X|\mathcal{B}]$  est linéaire.
- (b) Si  $X \geq 0$  alors  $\mathbb{E}[X|\mathcal{B}] \geq 0$  et plus généralement, si  $X \geq Y$  alors  $\mathbb{E}[X|\mathcal{B}] \geq \mathbb{E}[Y|\mathcal{B}]$ .
- (c)  $\mathbb{E}[X|\mathcal{B}]$  est l'unique variable aléatoire  $Y$   $\mathcal{B}$ -mesurable telle que

def:cond

$$(7) \quad \forall A \in \mathcal{B}, \quad \int_A X(\omega) d\mathbb{P}(\omega) = \int_A Y(\omega) d\mathbb{P}(\omega).$$

En particulier

- (a) Si  $X$  est  $\mathcal{B}$ -mesurable, alors  $\mathbb{E}[X|\mathcal{B}] = X$  p.s.
- (b) Si  $X$  et  $\mathcal{B}$  sont indépendantes *i.e.* si  $\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A] \mathbb{E}[X]$  pour tout  $A \in \mathcal{B}$ , alors  $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$ .
- (c) L'équation (7) sert de définition si  $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$  (au lieu de  $L^2$ ).
- (d) Si  $\mathcal{C}$  est une sous-tribu de  $\mathcal{B}$ , alors  $\mathbb{E}[X|\mathcal{C}] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}]|\mathcal{C}]$ .
- (e)  $\mathbb{E}[X|\{\emptyset, \Omega\}] = \mathbb{E}[X]$  et  $\mathbb{E}[\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[X]$ .
- (f) (Convergence conditionnelle) Si  $X_n \geq 0$  et  $X_n$  croît p.s. vers  $X \in L^1$ , alors  $\lim \mathbb{E}[X_n|\mathcal{B}] = \mathbb{E}[X|\mathcal{B}]$  p.s.
- (g) (convergence dominée conditionnelle) Si  $|X_n| \leq Z$  avec  $Z \in L^1(\Omega, \mathcal{A}, \mathbb{P})$  et si  $X_n \rightarrow X$  p.s. alors  $\mathbb{E}[X_n|\mathcal{B}] \rightarrow \mathbb{E}[X|\mathcal{B}]$  p.s.
- (h) (Lemme de Fatou conditionnel), si  $X_n \geq 0$ , alors  $\mathbb{E}[\liminf X_n|\mathcal{B}] \leq \liminf \mathbb{E}[X_n|\mathcal{B}]$ .
- (i) Soient  $X, Y$  deux variables aléatoires telles que  $X, Y$  et  $XY$  soient intégrables. Si  $X$  est  $\mathcal{B}$ -mesurable, alors  $\mathbb{E}[XY|\mathcal{B}] = X\mathbb{E}[Y|\mathcal{B}]$  p.s.
- (j) Si  $X$  est une variable aléatoire réelle, alors  $X$  est indépendante par rapport à  $\mathcal{B}$  si et seulement si, pour tout  $u \in \mathbb{R}$ ,  $\mathbb{E}[e^{iuX}|\mathcal{B}] = \mathbb{E}[e^{iuX}]$ .

exo:cond10

EXERCICE 1.14.

- (1) Soient  $X$  et  $Y$  deux v.a. Montrer qu'elles sont indépendantes si et seulement si, pour toute fonction  $g : \mathbb{R} \rightarrow \mathbb{R}$  borélienne et bornée, on a

$$\mathbb{E}[g(Y)|X] = \mathbb{E}[g(Y)] p.s.$$

- (2) Application: on considère un vecteur aléatoire  $(X, Y)$  dont la loi possède une densité par rapport à la mesure de Lebesgue

$$\rho(x, y) = e^{-y} \mathbf{1}_{0 < x < y}.$$

Calculer la loi conditionnelle de  $Y$  sachant que  $X = x$ . En déduire que  $X$  et  $Y - X$  sont indépendantes.

exo:cond3

EXERCICE 1.15. Rappelons que  $X$  suit une loi de Poisson de paramètre  $\lambda$  si  $\mathbb{P}(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}$ . Dans ce cas  $\mathbb{E}[X] = \lambda$ .

- (1) Le nombre  $N$  de voitures passant devant une station d'essence suit une loi de Poisson de paramètre  $\lambda > 0$ . Chaque voiture s'arrête à la station avec une probabilité  $p$ , indépendamment des autres. On note  $K$  le nombre de voitures s'arrêtant à la station. Trouver l'espérance de  $K$ .
- (2) Un auto-stoppeur attend à un péage d'autoroute à Bordeaux. Le nombre de véhicules passant par ce péage durant une heure est une variable aléatoire  $X$ . Pour chaque véhicule il y a une probabilité  $p \in ]0, 1[$  qu'il vienne de la direction "Paris" et  $1 - p$  qu'il vienne de la direction "Espagne". On note  $Y$  (resp.  $Z$ ) le nombre de véhicules venant de Paris (resp. d'Espagne) donc  $X = Y + Z$ . On suppose que  $X$  suit une loi de Poisson de paramètre  $\lambda$ . Déterminer les lois de  $Y$  et de  $Z$  et montrer que  $Y$  et  $Z$  sont indépendantes.

EXERCICE 1.16. On considère l'espace de probabilité  $([0, 1], \mathcal{A}, \lambda)$  avec  $\mathcal{A}$  la tribu borélienne et  $\lambda$  la mesure de Lebesgue.

- (1) Considérons la tribu  $\mathcal{B}$  sur  $[0, 1]$  engendrée par les intervalles  $]1/4, 2/3]$  et  $]2/3, 1]$ . On note  $X$  la variable aléatoire  $\omega \rightarrow \omega^2$ . Calculer  $\mathbb{E}[X|\mathcal{B}]$ .
- (2) Pour tout entier  $n$ , et tout entier  $0 \leq k \leq 2^n - 1$ , on pose  $I_{n,k} = [2^{-n}k, 2^{-n}(k+1)]$ . Considérons les tribus  $\mathcal{F}_n = \sigma(I_{n,k}, 0 \leq k \leq 2^n - 1)$  et  $\mathcal{F}_\infty = \bigcup_n \mathcal{F}_n$ .

**4.3. Approximation linéaire.** Revenons provisoirement sur le cas où la tribu  $\mathcal{B}$  est engendré par un vecteur aléatoire  $X = (X_1, \dots, X_n)$ . Dans ce cas, les fonctions  $\mathcal{B}$ -mesurables sont les fonctions de la forme

$$Z = f(X_1, \dots, X_n) \quad f \text{ borélienne.}$$

Ainsi, l'espérance conditionnelle vérifie:

$$\|Y - \mathbb{E}[Y|X]\|_{L^2(\Omega, \mathcal{A}, \mathbb{P})} = \inf\{\|Y - Z\|_{L^2(\Omega, \mathcal{A}, \mathbb{P})} : Z \in L^2(\Omega, \mathcal{A}, \mathbb{P}) - \exists f \text{ borélienne } Z = f(X_1, \dots, X_n)\}.$$

Comme l'espérance conditionnelle n'est pas, en général, une fonction *linéaire* des v. a. r. connues  $X_1, \dots, X_n$ , elle est difficile à calculer. On peut donc être amené à chercher la v. a. r. la plus proche de  $Y$  dans le sous-espace engendré par  $X_1, \dots, X_n$ .

On pose encore  $H = \text{Vect}(X_1, \dots, X_n)$  et on appelle *meilleure approximation linéaire* de  $Y$  par rapport à  $X_1, \dots, X_n$ , la v. a. r.  $Z = \text{Proj}_H(Y)$ . On a

$$\|Y - Z\| = \inf\{\|Y - U\| : U \in H\}.$$

Comme une fonction linéaire des  $X_1, \dots, X_n$  est mesurable pour la tribu engendrée par les  $X_1, \dots, X_n$ , on a que

$$\|Y - \mathbb{E}[Y|X]\| \leq \|Y - Z\|$$

donc l'approximation linéaire est moins bonne que l'espérance conditionnelle, mais c'est une fonction plus simple, parce que linéaire, des variables  $X_1, \dots, X_n$ .

Les espaces gaussiens sont ici particulièrement intéressants puisque les deux notions y coïncident:

THÉORÈME 1.43. Soit  $X = (X_1, \dots, X_n)$  un vecteur gaussien de  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ . Si  $Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ , alors l'approximation linéaire de  $Y$  par rapport à  $X$  et l'espérance conditionnelle de  $Y$  par rapport à  $X$  coïncident. En d'autres termes

$$\mathbb{E}[Y|X] = a_1 X_1 + \dots + a_n X_n.$$

### 5. Solution de certains exercices

**5.1. Exercice 1.1.** <sup>exo:esppardisc</sup> Dans les cas discrets  $\mathbb{P}_X = \sum_{k \in \mathbb{Z}} \mathbb{P}(X = k) \delta_k$ .

Pour vérifier que  $\mathbb{P}_X$  est une probabilité, il suffit de voir que  $\mathbb{P}(X = k) \geq 0$  (ce qui est trivial pour les 4 exemples) et que  $\sum_{k \in \mathbb{Z}} \mathbb{P}(X = k) = 1$ . L'espérance et la variance s'obtiennent alors avec la formule

$$\mathbb{E}[X] = \sum_{k \in \mathbb{Z}} k \mathbb{P}(X = k) \quad \text{Var}(X) = \sum_{k \in \mathbb{Z}} (k - \mathbb{E}[X])^2 \mathbb{P}(X = k) = \sum_{k \in \mathbb{Z}} k^2 \mathbb{P}(X = k) - \mathbb{E}[X]^2.$$

i) *Bernouilli*( $p$ ): c'est bien une probabilité puisque  $\mathbb{P}(X = 0) + \mathbb{P}(X = 1) = 1 - p + p = 1$ ,  $\mathbb{E}[X] = 0(1 - p) + 1p = p$  et  $\text{Var}(X) = (0 - p)^2(1 - p) + (1 - p)^2 * p = p(1 - p)(p + 1 - p) = p(1 - p)$ .

ii) *Binomiale*( $n, p$ ): Nous allons utiliser le fait que

$$(x + a)^n = \sum_{k=0}^n \binom{n}{k} x^k a^{n-k}$$

donc en dérivant

$$\sum_{k=0}^n \binom{n}{k} k x^k a^{n-k} = x \sum_{k=0}^n \binom{n}{k} k x^{k-1} a^{n-k} = x \sum_{k=0}^n \binom{n}{k} \partial_x (x^k)' a^{n-k} = x \partial_x (x + a)^n = nx(x + a)^{n-1}$$

et

$$\begin{aligned} \sum_{k=0}^n \binom{n}{k} k^2 x^k a^{n-k} &= x \sum_{k=0}^n \binom{n}{k} k \partial_x (x^k) a^{n-k} = x \partial_x (nx(x + a)^{n-1}) \\ &= nx(x + a)^{n-1} + n(n-1)x^2(x + a)^{n-2} = nx(x + a)^{n-2}(nx + a). \end{aligned}$$

Ainsi, on a bien une mesure de probabilité

$$\sum_{k=0}^n \mathbb{P}(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1.$$

L'espérance est donnée par

$$\mathbb{E}[X] = \sum_{k=0}^n \binom{n}{k} k p^k (1-p)^{n-k} = np(p + 1 - p)^{n-1} = np$$

alors que la variance est donnée par

$$\text{Var}(X) = \sum_{k=0}^n \binom{n}{k} (k^2 - 2nkp + n^2 p^2) p^k (1-p)^{n-k} = np(p + 1 - p)^{n-2} (np + 1 - p) - 2n^2 p^2 + n^2 p^2 = np(1-p).$$

iii) *Géométrique*( $p$ ): On utilise la somme de la série géométrique et ses dérivées: si  $|r| < 1$

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}, \quad \sum_{k=0}^{\infty} k r^{k-1} = \frac{\partial}{\partial r} \frac{1}{1-r} = \frac{1}{(1-r)^2}$$

donc  $\sum_{k=0}^{\infty} k r^k = \frac{r}{(1-r)^2}$  et, en dérivant,

$$\sum_{k=0}^{\infty} k^2 r^{k-1} = \frac{\partial}{\partial r} \frac{r}{(1-r)^2} = \frac{1}{(1-r)^2} + \frac{2r}{(1-r)^3} = \frac{1+r}{(1-r)^3}.$$

Ainsi  $\mathbb{P}_X$  est bien une probabilité puisque

$$\sum_{k=1}^{\infty} \mathbb{P}(X = k) = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p \sum_{j=0}^{+\infty} (1-p)^j = \frac{p}{1-(1-p)} = 1.$$

L'espérance de  $X$  est donnée par

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \frac{1}{(1-(1-p))^2} = \frac{1}{p}$$

et

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} - \frac{1}{p^2} = p \frac{1+1-p}{(1-(1-p))^3} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

iv) *Poisson*( $\lambda$ ) :  $\mathbb{P}_X$  est bien une probabilité puisque

$$\sum_{k=0}^{\infty} \mathbb{P}(X = k) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

L'espérance est donnée par

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} e^{-\lambda} k \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!} = \lambda \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} = \lambda$$

alors que pour la variance

$$\begin{aligned} \text{Var}(X) &= \sum_{k=0}^{\infty} e^{-\lambda} k^2 \frac{\lambda^k}{k!} - \lambda^2 = \sum_{k=0}^{\infty} e^{-\lambda} (k-1)k \frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} e^{-\lambda} k \frac{\lambda^k}{k!} - \lambda^2 \\ &= \sum_{k=2}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-2)!} \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!} - \lambda^2 \\ &= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda - \lambda^2 = \lambda. \end{aligned}$$

## 5.2. Exercice <sup>ex:loicont</sup> 1.2. Pour la loi uniforme

$$\int_{\mathbb{R}} \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b dx = 1,$$

puis

$$\mathbb{E}[X] = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

et enfin

$$\begin{aligned} \text{Var}(X) &= \frac{1}{b-a} \int_a^b x^2 dx - \frac{(a+b)^2}{4} = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} \\ &= \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{a^2 + b^2 - 2ab}{12} = \frac{(a-b)^2}{12}. \end{aligned}$$

La loi normale a été traitée en détail dans la section <sup>sec:normal</sup> 2.

Pour la loi exponentielle

$$\int_{\mathbb{R}} \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x) dx = \lambda \int_0^{+\infty} e^{-\lambda x} dx = 1.$$

Une intégration par parties donne

$$\mathbb{E}[X] = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = \left[ \frac{-1}{\lambda} e^{-\lambda x} \right]_0^{+\infty} = \frac{1}{\lambda}.$$

puis

$$\mathbb{E}[X^2] = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda} \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

donc  $\text{Var}(X) = \frac{1}{\lambda^2}$ .

Pour Cauchy, notons que  $\mathbb{P}_X(x) = \frac{1}{a\pi(1+(x/a)^2)} dx$  d'où

$$\int_{\mathbb{R}} \frac{1}{a\pi(1+(x/a)^2)} dx = \int_{\mathbb{R}} \frac{1}{\pi(1+t^2)} dt = \frac{1}{\pi} [\arctan t]_{-\infty}^{+\infty} = 1.$$

Ensuite  $|x|/(a^2+x^2)$  n'est pas intégrable, donc une variable de Cauchy n'a pas d'espérance et a fortiori pas de variance.

**5.3. Exercice** <sup>exo:repart</sup> **1.3.** Notons que  $F_X(\lambda) = \mathbb{P}(X \leq \lambda) = \mathbb{P}_X((-\infty, \lambda])$ .  
 $F_X$  est croissante puisque, si  $x < y$ , alors  $F_X(y) = F_X(x) + \mathbb{P}_X(]x, y]) \geq F_X(x)$ .  
 Ensuite,

$$\lim_{n \rightarrow -\infty} (-\infty, n] := \bigcap_{n \in \mathbb{N}} (-\infty, -n] = \emptyset$$

comme limite décroissante d'ensembles et

$$\lim_{n \rightarrow +\infty} (-\infty, n] := \bigcup_{n \in \mathbb{N}} (-\infty, n] = \mathbb{R}$$

comme limite croissante d'ensembles. Donc, par continuité de l'application probabilité,

$$\lim_{n \rightarrow -\infty} F_X(n) = \lim_{n \rightarrow -\infty} \mathbb{P}(X \leq n) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} (-\infty, -n]\right) = P(\emptyset) = 0$$

et

$$\lim_{n \rightarrow +\infty} F_X(n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq n) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} (-\infty, n]\right) = P(\mathbb{R}) = 1.$$

Comme  $\lim_{n \rightarrow +\infty} (-\infty, x + 1/n] = \bigcap_{n \in \mathbb{N}} (-\infty, x + 1/n] = (-\infty, x]$  (limite décroissante),  $F_X(x + 1/n) = \mathbb{P}_X((-\infty, x + 1/n]) \rightarrow \mathbb{P}_X((-\infty, x]) = F_X(x)$  i.e.  $F_X$  est continue à droite. Par ailleurs,  $\lim_{n \rightarrow +\infty} (-\infty, x - 1/n] = \bigcup_{n \in \mathbb{N}} (-\infty, x - 1/n] = (-\infty, x[$  (limite croissante) donc  $F_X(x + 1/n) = \mathbb{P}_X((-\infty, x - 1/n]) \rightarrow \mathbb{P}_X((-\infty, x])$  donc  $F_X$  a une limite à gauche  $F_X(x^-)$ .

Le saut de  $F_X$  est  $F_X(x) - F_X(x^-) = \mathbb{P}_X((-\infty, x]) - \mathbb{P}_X((-\infty, x]) = \mathbb{P}_X(\{x\})$ . Comme  $X$  est discret, cela détermine bien la loi de  $X$ .

**5.4. Exercice** <sup>exo:gener</sup> **1.4.** Comme  $|\mathbb{P}(X = k)s^k| \leq \mathbb{P}(X = k)$  pour  $|s| \leq 1$  et que  $\sum_{k \in \mathbb{N}} \mathbb{P}(X = k)$  converge ( $= \mathbb{P}(\mathbb{N}) = 1$ ), la série  $\sum_{k \in \mathbb{N}} \mathbb{P}(X = k)s^k$  converge i.e.  $G_X(s)$  est bien définie sur  $[-1, 1]$ .

De plus,  $G_X$  est une fonction de classe  $\mathcal{C}^\infty$  sur  $] -1, 1[$  puisque développable en série entière et  $\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}$ . La loi de  $X$  est donc caractérisée par  $G_X$ .

Si  $X$  est intégrable,  $\sum_{k \in \mathbb{N}} k\mathbb{P}(X = k)$  converge (et vaut  $\mathbb{E}[X]$ ). Mais  $G'_X(s) = \sum_{k \in \mathbb{N}} k\mathbb{P}(X = k)s^{k-1}$ .

La convergence de cette série en 1 implique l'existence de  $G'_X(1)$  et  $G'_X(1) = \sum_{k \in \mathbb{N}} k\mathbb{P}(X = k) = \mathbb{E}[X]$ .

On montre de la même manière que, si  $\mathbb{E}[X^2]$  existe, alors  $\mathbb{E}[X(X-1)]$  aussi et donc  $G''_X(s) = \sum_{k \in \mathbb{N}} k(k-1)\mathbb{P}(X = k)s^{k-2} \rightarrow_{s \rightarrow 1} \sum_{k \in \mathbb{N}} k(k-1)\mathbb{P}(X = k) = \mathbb{E}[X(X-1)]$ . Ainsi  $\mathbb{E}[X^2] = \mathbb{E}[X] + \mathbb{E}[X(X-1)] = G'_X(1) + G''_X(1)$  et enfin  $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = G''_X(1) - G'_X(1)^2 + G'_X(1)$ .

Enfin, pour la loi Bernouilli( $p$ )  $G_X = 1 - p + sp$ , pour la loi binomiale( $n, p$ )

$$G_X(s) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} s^k = (1-p+sp)^n.$$

Pour Poissons( $\lambda$ ),  $G_X(s) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} s^n = e^{\lambda s} e^{-\lambda} = e^{\lambda(s-1)}$ . Enfin pour la loi géométrique( $p$ ),

$$G_X(s) = \sum_{k=1}^{\infty} s^k (1-p)^{k-1} p = \frac{sp}{1-s(1-p)}.$$

exo:espvardisc  
Le lecteur vérifiera qu'on retrouve bien les espérances et variances calculées ci-dessus à l'exercice 1.1.

ex:lois  
**5.5. Exercice 1.5.** On considère  $X_1, X_2, \dots$  indépendantes de loi  $\mathbb{P}(X_i = 1) = p$  (succès avec probabilité  $p$ ) et  $\mathbb{P}(X_i = 0) = 1 - p$  (échec avec probabilité  $1 - p$ ).

On note  $S_n$  le nombre de succès parmi les  $n$  premières expériences. Pour  $k \in \{1, \dots, n\}$ ,

$$\begin{aligned} \mathbb{P}(S_n = k) &= \sum_{S \subset \{1, \dots, n\}, |S|=k} \mathbb{P}(\forall i \in S, X_i = 1, \forall i \notin S, X_i = 0) \\ &= \sum_{S \subset \{1, \dots, n\}, |S|=k} \prod_{i \in S} \mathbb{P}(X_i = 1) \prod_{i \notin S} \mathbb{P}(X_i = 0) \\ &= \sum_{S \subset \{1, \dots, n\}, |S|=k} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

L'instant  $N$  de premier succès est tel que les  $N - 1$  expériences précédentes sont des échecs donc

$$\mathbb{P}(N = k) = \mathbb{P}(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) = p(1-p)^{k-1}.$$

Ainsi  $N$  suit une loi géométrique.

ex:bool  
**5.6. Exercice 1.6.** Soit  $p_n$  la probabilité qu'à la  $n$ -ième transmission l'information soit correcte. Il y a deux possibilités pour que cette information soit correcte:

- soit elle était correcte à l'étape précédente et il n'y a pas d'erreur de transmission
- soit elle était fautive mais la transmission est erronée.

Ainsi

$$p_0 = 1, \quad p_n = p.p_{n-1} + (1-p)(1-p_{n-1}) = 1-p + (2p-1)p_{n-1}.$$

Pour  $p = 1/2$ , on trouve  $p_n = 1/2$ . Pour  $p \neq 1/2$ , on se ramène à une suite géométrique en posant

$$q_n = p_n - \alpha = 1 - p + (2p-1)p_{n-1} - \alpha = 1 - p + (2p-2)\alpha + (2p-1)q_{n-1}$$

Ainsi, en prenant  $\alpha = 1/2$ ,  $q_0 = 1 - \alpha = 1/2$ ,  $q_n = (2p-1)q_{n-1}$  donc  $q_n = (2p-1)^n/2$  donc

$$p_n = \frac{1 + (2p-1)^n}{2}.$$

Notons que  $-1/2 \leq 2p-1 \leq 1/2$  donc  $(2p-1)^n \rightarrow 0$  et  $p_n \rightarrow \frac{1}{2}$ .

exo:cond10  
**5.7. Exercice 1.14. 1)** Supposons d'abord que  $X$  et  $Y$  sont indépendantes. Alors  $X$  et  $g(Y)$  le sont aussi - car  $\sigma(g(Y)) \subset \sigma(Y)$ . Donc  $\mathbb{E}[g(Y)|X] = \mathbb{E}[g(Y)]$  p.s.

Réciproquement, supposons que  $\mathbb{E}[g(Y)|X] = \mathbb{E}[g(Y)]$  p.s. pour toute  $g$  borélienne bornée. Soit alors  $h$  une fonction borélienne bornée,

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

d'où l'indépendance.

2) Pour trouver la loi conditionnelle de  $Y$  sachant que  $X = x$ , commençons par calculer la loi marginale de  $X$ :

$$\rho_X(x) = \int_x^{+\infty} e^{-y} dy \mathbf{1}_{x>0} = e^{-x} \mathbf{1}_{x>0}.$$

On a alors

$$\mathbb{P}(y|x) = \frac{\rho(x, y)}{\rho_X(x)} = e^{-y+x} \mathbf{1}_{x>0}$$



et

$$\mathbb{P}(Y \leq t | X = x) = \int_x^t \mathbb{P}(y|x) dy = (t-x)(e^{x-t} - 1).$$

Pour montrer que  $X$  et  $X - Y$  sont indépendantes, considérons une fonction borélienne bornée  $g$ . On a

$$\begin{aligned} \mathbb{E}[g(Y - X) | X = x] &= \int_{\mathbb{R}} g(y-x) \mathbb{P}(y|x) dy \\ &= \int_x^{+\infty} g(y-x) e^{-y+x} dy \\ &= \int_0^{+\infty} g(v) e^{-v} dv \end{aligned}$$

avec le changement de variable  $v = y - x$ . Cette dernière intégrale est une constante (indépendante de  $x$ ) ce qui démontre l'indépendance cherchée.

**5.8. Solution de l'exercice 1.15.** <sup>exo:cond3</sup> 1) Par hypothèse  $\mathbb{P}(N = n) = \frac{\lambda^n}{n!} e^{-\lambda}$  et  $\mathbb{P}(K = k | N = n) = C_n^k p^k (1-p)^{n-k}$   $0 \leq k \leq n$ : on suppose que  $n$  voitures  $v_1, \dots, v_n$  passent. La probabilité que chacune des voitures  $v_{j_1}, \dots, v_{j_k}$  arrêtent est de  $p$ , celle que chacune des autres ne s'arrête pas est de  $(1-p)$ . Comme ces événements sont indépendants, la probabilité que  $v_{j_1}, \dots, v_{j_k}$  arrêtent et que les autres passent est de  $p^k (1-p)^{n-k}$ . Il y a  $C_n^k$  telles configurations, d'où le résultat.

On en déduit que

$$\mathbb{E}[K | N = n] = \sum_{k=0}^n k \mathbb{P}(K = k | N = n) = \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} = np$$

ce qui donne  $\mathbb{E}[K | N] = pN$ .

Les propriétés de l'espérance conditionnelle montrent que  $\mathbb{E}[K] = \mathbb{E}[\mathbb{E}[K | N]] = p\mathbb{E}[N] = \lambda p$ .

2) Posons  $Y = \sum_1^X Y_i$  avec  $Y_i = 1$  si la  $i$ -ième voiture vient de Paris et  $Y_i = 0$  sinon. À nouveau

$$\mathbb{P}[Y = k | X = n] = C_n^k p^k (1-p)^{n-k}$$

pour  $k \leq n$  (et = 0 pour  $k > n$ ). Ainsi, comme dans la question précédente

$$\mathbb{P}(Y = k) = \sum_{n \geq k} \mathbb{P}[Y = k | X = n] \mathbb{P}(X = n) = \sum_{n \geq k} C_n^k p^k (1-p)^{n-k} \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda p} \frac{(\lambda p)^k}{k!}.$$



## Chaînes de Markov et processus gaussien

### 1. Introduction

Pour tenir compte des “bruits” provenant, par exemple, de la transmission des signaux, on va reprendre ce que l’on a déjà vu au premier semestre avec un terme supplémentaire : l’aléatoire !

Ainsi si  $f(t)$  est un signal analogique, on l’a échantillonné en une série  $\{f(nT)\}_{n \in \mathbb{Z}}$  numérique. Maintenant, du aux “bruits”, notre série sera une série  $\{f(nT) + X(n)\}_{n \in \mathbb{Z}}$ , avec  $\{X(n)\}_{n \in \mathbb{Z}}$  une suite de v. a. r..

**DÉFINITION 2.1.** Un *processus aléatoire discret* est une suite de v. a. r.  $\{X(n)\}_{n \in \mathbb{Z}}$  définies sur l’espace de probabilités  $(\Omega, \mathcal{A}, \mathbb{P})$ . La suite  $\{X(n)\}_{n \in \mathbb{Z}}$  est aussi appelée série chronologique, ou série temporelle.

Attention les v. a. r.  $X(n)$  sont à *valeurs réelles (ou complexes)*, ce qui est discret est le fait que l’on traite une *suite* de telles v. a. r..

En fait un processus est un vecteur aléatoire de dimension infinie. Sa loi est définie par la famille de toutes les lois de ses sous-vecteurs de dimension finie  $k$ , pour toutes les valeurs de l’entier  $k$ . Ainsi la loi de tous les vecteurs de la forme

$$Y := (X(n_1), \dots, X(n_k)),$$

doit être connue. En fait, souvent, la loi du processus est donnée de façon récurrente à partir d’un état initial.

Dans ce cours, on ne s’intéressera qu’aux processus discrets  $\{X(n)\}_{n \in \mathbb{N}}$  contenus dans  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ , qui est un espace de Hilbert. À un processus discret, on peut associer différents sous-espaces fermés. Notons

- le *passé linéaire* de  $X$  au temps  $n$  :  $H(X, n) = \overline{\text{Vect}\{X(n), X(n-1), \dots\}}$ ;
- l’*histoire linéaire* de  $X$  :  $H(X) = \overline{\text{Vect}\{X(n), n \in \mathbb{Z}\}}$ .

**DÉFINITION 2.2.**

Le processus discret  $X$  est *strictement stationnaire* si sa loi de probabilité est invariante par décalage temporel *i.e.* si pour tout  $k$  et tout  $(n_1, \dots, n_k) \in \mathbb{Z}^k$ , la loi du vecteur  $Y_\tau := (X(n_1 + \tau), \dots, X(n_k + \tau))$  est la loi de  $Y_0$ , quel que soit  $\tau \in \mathbb{Z}$ .

Comme l’espérance ne dépend que de la loi, cette propriété implique que l’espérance de  $X(n)$  est *indépendante* de  $n$ .

De même, la covariance de  $X(m)$  et de  $X(n)$  ne dépend que de la loi du vecteur  $(X(m), X(n))$  qui est aussi la loi de  $(X(0), X(n-m))$ . Ainsi

$$\text{Cov}(X(m), X(n)) = \text{Cov}(X(0), X(n-m)) := R(n-m).$$

En général, on se limitera à ces deux seuls paramètres, on ne demandera donc la stationnarité que pour eux:

**DÉFINITION 2.3.**

On dira qu’un processus  $X(n)$  est *stationnaire au sens large* (SSL) si on a

- (i) pour tout  $n$ ,  $\mathbb{E}(X(n)) = \lambda$
- (ii) pour tous  $m, n$ ,  $\text{Cov}(X(m), X(n)) = R(n-m)$  — donc  $\text{Var}(X(n)) = \text{Cov}(X(n), X(n)) = R(0) = \sigma^2$ .

Si  $X(n)$  est SSL et si on prend le vecteur  $Y = (X(n_1), \dots, X(n_k))$ , sa matrice de covariance a donc des coefficient constants sur les parallèles à la diagonale principale. On dit que c'est une matrice de Toeplitz et on peut lui associer un opérateur sur un espace de Hilbert de fonction holomorphes sur le disque unité de  $\mathbb{C}$ .

## 2. Chaîne de Markov homogène

Rappelons la définition de la probabilité conditionnelle :

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

DÉFINITION 2.4. Supposons que les v. a. r.  $\{X(n)\}_{n \in \mathbb{N}}$  prennent leurs valeurs dans un ensemble fini ou dénombrable, l'ensemble des états  $\mathcal{F} \subset \mathbb{N} \subset \mathbb{R}$  et que l'on a,  $\forall n \in \mathbb{N}, \forall j, k, k_0, \dots, k_{n-1} \in \mathcal{F}$

$$\begin{aligned} \mathbb{P}(\{X(n+1) = j\} \mid \{X(n) = k\} \cap \{X(n-1) = k_{n-1}\} \cap \dots \cap \{X(0) = k_0\}) \\ = \mathbb{P}(\{X(n+1) = j\} | \{X(n) = k\}) := p_{j,k}, \end{aligned}$$

alors on dit que le processus  $\{X(n)\}_{n \in \mathbb{N}}$  est une *chaîne de Markov de matrice de transition*  $\mathcal{P} = \{p_{j,k}\}_{j,k \in \mathcal{F}}$ .

Si en plus, la probabilité  $\mathbb{P}(\{X(n+1) = j\} | \{X(n) = k\})$  ne dépend pas de  $n$  i.e. si

$$\forall n, \quad \mathbb{P}(\{X(n+1) = j\} | \{X(n) = k\}) = \mathbb{P}(\{X(1) = j\} | \{X(0) = k\})$$

on dit que la chaîne de Markov est *homogène*.

REMARQUE 2.5. On suppose ici que  $\mathbb{P}(\{X(n+1) = j\} | \{X(n) = k\} \cap \{X(n-1) = k_{n-1}\} \cap \dots \cap \{X(0) = k_0\})$  a bien un sens, c'est-à-dire que  $\mathbb{P}(\{X(n) = k\} \cap \{X(n-1) = k_{n-1}\} \cap \dots \cap \{X(0) = k_0\}) > 0$ .

Ainsi, pour connaître la probabilité que  $\{X(n+1) = k\}$ , on n'a pas besoin de connaître tout le passé du processus jusqu'au temps  $n$ , il suffit de connaître l'étape précédente. Plus précisément, supposons qu'on connaisse la *loi initiale* du processus, c'est à dire la fonction  $\lambda : \mathcal{F} \rightarrow [0, 1]$  donnée par  $\lambda(j) = \mathbb{P}(\{X(0) = j\})$ . Alors, tout d'abord

$$\mathbb{P}(\{X(1) = k_1\} \cap \{X(0) = k_0\}) = \mathbb{P}(\{X(1) = k_1\} | \{X(0) = k_0\}) \mathbb{P}(\{X(0) = k_0\}) = p_{k_1, k_0} \lambda(k_0)$$

donc

$$\begin{aligned} \mathbb{P}(\{X(1) = k_1\}) &= \mathbb{P}\left(\{X(1) = k_1\} \cap \bigcup_{k_0 \in \mathcal{F}} \{X(0) = k_0\}\right) \\ &= \mathbb{P}\left(\{X(1) = k_1\} \cap \bigcup_{k_0 \in \mathcal{F}} (\{X(1) = k_1\} \cap \{X(0) = k_0\})\right) \\ &= \sum_{k_0 \in \mathcal{F}} \mathbb{P}(\{X(1) = k_1\} \cap \{X(0) = k_0\}) = \sum_{k_0 \in \mathcal{F}} p_{k_1, k_0} \lambda(k_0). \end{aligned}$$

Cela peut s'écrire matriciellement, en notant  $\pi_0 = [\lambda(j)]_{j \in \mathcal{F}}$  et  $\pi_k = [\mathbb{P}(\{X(k) = j\})]_{j \in \mathcal{F}}$  alors  $\mathbb{P}(\{X(1) = k_1\}) = (\mathcal{P}\pi_0)(k_1)$  i.e.  $\pi_1 = \mathcal{P}\pi_0$ .

De même,

$$\begin{aligned} \mathbb{P}(\{X(2) = k_2\} \cap \{X(1) = k_1\} \cap \{X(0) = k_0\}) \\ = \mathbb{P}(\{X(2) = k_2\} | \{X(1) = k_1\} \cap \{X(0) = k_0\}) \mathbb{P}(\{X(1) = k_1\} \cap \{X(0) = k_0\}) \\ = p_{k_2, k_1} p_{k_1, k_0} \lambda(k_0) \end{aligned}$$

d'où on déduit immédiatement (en sommant respectivement sur  $k_0$ , sur  $k_1$  et sur  $k_0, k_1$ )<sup>1</sup>

$$\begin{aligned}\mathbb{P}(\{X(2) = k_2\} \cap \{X(1) = k_1\}) &= p_{k_2, k_1} \sum_{k_0 \in \mathcal{F}} p_{k_1, k_0} \lambda(k_0) = p_{k_2, k_1} (\mathcal{P}\pi_0)(k_1) = p_{k_2, k_1} \mathbb{P}(\{X(1) = k_1\}) \\ \mathbb{P}(\{X(2) = k_2\} \cap \{X(0) = k_0\}) &= \left( \sum_{k_1 \in \mathcal{F}} p_{k_2, k_1} p_{k_1, k_0} \right) \lambda(k_0) = \mathcal{P}^2(k_2, k_0) \lambda(k_0) \\ \mathbb{P}(\{X(2) = k_2\}) &= (\mathcal{P}^2\pi_0)(k_2) = (\mathcal{P}\pi_1)(k_2).\end{aligned}$$

Une récurrence simple montre alors que

**eq:markov**

$$(8) \quad \mathbb{P}(\{X(n) = k_n\} \cap \dots \cap \{X(0) = k_0\}) = p_{k_n, k_{n-1}} \dots p_{k_1, k_0} \lambda(k_0),$$

donc en sommant sur  $k_0, \dots, k_{n-1}$

$$\begin{aligned}\mathbb{P}(\{X(n) = k_n\}) &= \sum_{k_{n-1} \in \mathcal{F}} p_{k_n, k_{n-1}} \dots \left( \sum_{k_1 \in \mathcal{F}} p_{k_2, k_1} \left( \sum_{k_0 \in \mathcal{F}} p_{k_1, k_0} \lambda(k_0) \right) \right) \\ &= \sum_{k_0 \in \mathcal{F}} \left( \sum_{k_1, \dots, k_{n-1} \in \mathcal{F}} p_{k_n, k_{n-1}} \dots p_{k_1, k_0} \right) \lambda(k_0) \\ &= \sum_{k_0 \in \mathcal{F}} \mathcal{P}^n(k_n, k_0) \lambda(k_0) = (\mathcal{P}^n\pi_0)(k_n).\end{aligned}$$

On en déduit que  $\pi_n = \mathcal{P}^n\pi_0$  d'où  $\pi_n = \mathcal{P}^{n-k}\pi_k$  pour  $0 \leq k \leq n$ .

Enfin, en sommant (8) sur tous les  $k_0, \dots, k_{\ell-1} \in \mathcal{F}$  on obtient

$$\mathbb{P}(\{X(n) = k_n\} \cap \dots \cap \{X(\ell) = k_\ell\}) = p_{k_n, k_{n-1}} \dots p_{k_{\ell+1}, k_\ell} \mathbb{P}(\{X(\ell) = k_\ell\})$$

puis en sommant sur tous les  $k_{\ell+1}, \dots, k_{n-1} \in \mathcal{F}$ , on trouve

**eq:markov2**

$$(9) \quad \mathbb{P}(\{X(n) = k_n\} \cap \{X(\ell) = k_\ell\}) = \mathcal{P}^{n-\ell}(k_n, k_\ell) \mathbb{P}(\{X(\ell) = k_\ell\}).$$

Il en résulte que

$$\mathbb{P}(\{X(n) = k_n\} | \{X(\ell) = k_\ell\}) = \mathcal{P}^{n-\ell}(k_n, k_\ell).$$

Ainsi, le processus obtenu à partir de  $\{X(n)\}$  en allant de  $k = n - \ell$  pas en  $k$  pas est encore un processus de Markov, mais de matrice de transition  $\mathcal{P}^k$ .

EXEMPLE 2.6. Vérifiez dans chacun des exemples suivants que  $X(n)$  est une chaîne de Markov homogène et précisez sa matrice de transition.

- (1) *Promenades aléatoires*: soit  $(Y_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes et de même loi à valeurs dans  $\mathbb{Z}$  (ou  $\mathbb{Z}^d$ ), soit  $X_0$  une variable aléatoire à valeurs dans  $\mathbb{Z}$  (ou  $\mathbb{Z}^d$ ), indépendante des  $(Y_n)$ . On pose alors  $X(n) = X_0 + \sum_{j=1}^n Y_j$ .
- (2) *File d'attente*: Soit  $(T_n)_{n \in \mathbb{N}}$  la suite des instants (aléatoires) d'arrivée des clients à un guichet. Un seul client est servi à la fois. On note  $X(n)$  le nombre de clients en attente ou en cours de service juste avant l'instant  $T_n$  et  $D_n$  le nombre de clients dont le service se termine dans l'intervalle  $[T_n, T_{n+1}[$ . On suppose les variables  $D_n$  indépendantes et de même loi donnée: pour tout  $k \in \mathbb{N}$ ,  $p_k = \mathbb{P}(\{D_0 = k\})$ . On note  $a_+ = \max(a, 0)$  la partie positive du réel  $a$ , et on note  $X(n+1) = (X(n) - 1 + D_n)$ .
- (3) *Gestion de stock*: on s'intéresse au nombre de pièces d'un même type en stock dans un entrepôt, à différents instants  $(t_n)_{n \in \mathbb{N}}$ , par exemple à chaque fin de journée ou de semaine. La demande pour ce type de pièces dans l'intervalle  $[t_n, t_{n+1}[$  est une variable aléatoire entière  $D_n$ . La suite des v.a.  $(D_n)_{n \in \mathbb{N}}$  est supposée indépendante et de même loi connue.

La politique de gestion est la suivante: lorsque le niveau du stock à un des instants  $t_n$  descend en dessous d'un seuil  $s$  fixé on se réapprovisionne de façon à ramener le stock à son niveau maximal  $S$  (déterminé par exemple par la taille de l'entrepôt ou les moyens financiers

<sup>1</sup>On notera  $\mathcal{P}^j(k, \ell)$  est la  $(k, \ell)$ -ième entrée de la matrice  $\mathcal{P}^j$ .

de l'entreprise). On admet que la livraison intervient sans délai, c'est-à-dire avant le début de la période suivante. La taille  $X(n)$  du stock à l'instant  $t_n$  vérifie

$$\forall n \in \mathbb{N}, X(n+1) = \begin{cases} (X(n) - D_n)_+ & \text{si } s \leq X(n) < S \\ (S - D_n)_+ & \text{si } s \leq X(n) < s \end{cases}.$$

REMARQUE 2.7. Si  $(Y(n))_{n \in \mathbb{N}}$  est une suite de variables aléatoires indépendantes et de même loi à valeurs dans  $\mathcal{F}$  et si  $f : \mathcal{F} \times \mathcal{F} \rightarrow \mathcal{F}$  est une fonction quelconque, alors la suite  $(X(n))_{n \in \mathbb{N}}$  définie par  $X(n+1) = f(X(n), Y(n))$  et  $X(0)$  donnée, indépendante des  $(Y(n))_{n \in \mathbb{N}}$  est une chaîne de Markov homogène. Ce résultat (à démontrer en exercice) fournit un moyen assez général pour établir qu'une suite de variables aléatoires est une chaîne de Markov homogène.

DÉFINITION 2.8.

Soit maintenant  $\{X(n)\}$  une chaîne de Markov de matrice de transition  $\mathcal{P}$  et de loi initiale  $\pi_0$ . On dira que  $\pi_0$  est *stationnaire* si  $\mathcal{P}\pi_0 = \pi_0$  i.e. si  $\pi_0$  est vecteur propre de  $\mathcal{P}$  pour la valeur propre 1.

Une telle probabilité n'existe pas forcément.

Notons que si  $\pi_0$  est stationnaire alors pour tout  $k \in \mathcal{F}$

$$\mathbb{P}(\{X(1) = j\}) = \pi_1(j) = \sum_{k \in \mathcal{F}} p_{j,k} \pi_0(k) = \pi_0(j) = \mathbb{P}(\{X(0) = j\})$$

et plus généralement,  $\mathbb{P}(\{X(k) = j\}) = \mathbb{P}(\{X(0) = j\}) = \pi_0(j)$ . Ainsi, en généralisant le calcul menant à (9), pour  $k, m \in \mathbb{N}$ ,  $n_1 > n_2 > \dots > n_k$  et  $j_1, \dots, j_k \in \mathcal{F}$ ,

$$\begin{aligned} \mathbb{P}(\{X(n_1 + m) = j_1\} \cap \dots \cap \{X(n_k + m) = j_k\}) \\ = \mathcal{P}^{(n_1 + m) - (n_2 + m)}(j_1, j_2) \dots \mathcal{P}^{(n_{k-1} + m) - (n_k + m)}(j_{k-1}, j_k) \mathbb{P}(\{X(n_k + m) = j_k\}) \\ = \mathcal{P}^{n_1 - n_2}(j_1, j_2) \dots \mathcal{P}^{n_{k-1} - n_k}(j_{k-1}, j_k) \mathbb{P}(\{X(0) = j_k\}) \end{aligned}$$

si  $\pi_0$  est stationnaire. Cette quantité ne dépend plus de  $m$ , le processus  $\{X(n)\}$  est donc stationnaire.

DÉFINITION 2.9. Soient  $i$  et  $j$  deux états de  $\mathcal{F}$ . On dit que l'état  $j$  est accessible à partir de  $i$  si et seulement si

$$\exists n \in \mathbb{N}, \mathcal{P}^n(j, i) = \mathbb{P}(\{X(n) = j\} | \{X(0) = i\}) > 0.$$

On dit que les états  $i$  et  $j$  communiquent si et seulement si  $j$  est accessible à partir de  $i$  et  $i$  est accessible à partir de  $j$ .

Le processus est dit *irréductible* si tous les états communiquent entre eux.

REMARQUE 2.10. L'accessibilité est une relation d'équivalence: si  $i$  et  $j$  sont accessibles et  $j$  et  $k$  sont accessibles, il en va de même pour  $i$  et  $k$ .

### 3. Processus Gaussien

DÉFINITION 2.11. La série chronologique  $\{X(n)\}_{n \in \mathbb{Z}}$  est dite *gaussienne* si elle est contenue dans un espace gaussien, i.e. si tous les vecteurs  $Y := (X(n_1), \dots, X(n_k))$  sont des vecteurs gaussiens.

Si la série  $\{X(n)\}_{n \in \mathbb{Z}}$  est gaussienne et SSL, l'espérance et la matrice de covariance du vecteur  $Y_\tau := (X(n_1 + \tau), \dots, X(n_k + \tau))$  sont les mêmes que celles de  $Y_0$ . Mais comme  $Y_\tau$  est un vecteur gaussien, sa loi est complètement déterminée par son espérance et sa covariance, donc la loi de  $Y_\tau$  est la même que celle de  $Y_0$ , et donc le processus est strictement stationnaire. Ainsi, on a:

PROPOSITION 2.12. *Le signal  $\{X(n)\}_{n \in \mathbb{Z}}$  est stationnaire si et seulement les deux conditions suivantes sont remplies*

- (i)  $\forall n \in \mathbb{Z}, \mathbb{E}(X(n)) = \mathbb{E}(X(0))$ ,
- (ii) *les matrice de covariance  $\Gamma(m, n) = \text{Cov}(X(m)X(n))$  ne dépendent que de  $m - n$ , i.e.  $\forall m, n \in \mathbb{Z}, \Gamma(m, n) = R(m - n)$ .*

#### 4. Opérateur de covariance

**4.1. Définition.** Soit  $\{X(n)\}_{n \in \mathbb{Z}}$  une série chronologique, considérons les deux suites  $\{a_n\}_{n \in \mathbb{Z}}$  et  $\{b_n\}_{n \in \mathbb{Z}}$  de  $\ell^2(\mathbb{Z})$ . Associons leurs les v. a. r. suivantes :

$$A = \sum_{j \in \mathbb{Z}} a_j X(j) \quad \text{et} \quad B = \sum_{j \in \mathbb{Z}} b_j X(j)$$

(en supposant que les séries convergent).

On a

$$\mathbb{E}(A) = \sum_{j \in \mathbb{Z}} a_j \mathbb{E}(X(j)) \quad \text{et} \quad \mathbb{E}(B) = \sum_{j \in \mathbb{Z}} b_j \mathbb{E}(X(j))$$

d'où

$$\begin{aligned} \text{Cov}(A, B) &= \mathbb{E} \left( \sum_{j \in \mathbb{Z}} a_j (X(j) - \mathbb{E}(X(j))) \sum_{k \in \mathbb{Z}} b_k (X(k) - \mathbb{E}(X(k))) \right) \\ &= \sum_{j \in \mathbb{Z}} a_j \sum_{k \in \mathbb{Z}} b_k \mathbb{E} \left[ (X(j) - \mathbb{E}(X(j))) (X(k) - \mathbb{E}(X(k))) \right] \\ &= \sum_{j \in \mathbb{Z}} a_j \sum_{k \in \mathbb{Z}} b_k \Gamma(j, k) \end{aligned} \tag{10}$$

eq:cov

avec  $\Gamma(j, k) = \text{Cov}(X(j), X(k))$ .

Définissons l'opérateur de covariance  $\Gamma : \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$ , par

$$(\Gamma b)_j := \sum_{k \in \mathbb{Z}} \Gamma(j, k) b(k).$$

On a alors

$$\text{Cov}(A, B) = \langle a, \Gamma b \rangle,$$

le produit scalaire étant celui de  $\ell^2(\mathbb{Z})$ .

Comme la covariance de deux variables aléatoires est symétrique et positive,  $\Gamma^* = \Gamma$  il en est de même de l'opérateur de covariance :

$$\text{Cov}(A, B) = \langle a, \Gamma b \rangle = \langle \Gamma a, b \rangle = \langle b, \Gamma a \rangle = \text{Cov}(B, A)$$

et

$$\mathbb{V}(A) := \text{Cov}(A, A) = \langle a, \Gamma a \rangle = \mathbb{E} \left[ \left( \sum_{j \in \mathbb{Z}} a_j (X(j) - \mathbb{E}(X(j))) \right)^2 \right] \geq 0.$$

Enfin, si les v. a. r.  $A$  et  $B$  sont telles que la suites  $\{a_j\}_{j \in \mathbb{Z}}$  et  $\{b_j\}_{j \in \mathbb{Z}}$  sont à support fini, *i.e.* si, par exemple,  $a_j = b_j = 0$  sauf pour  $j = 1, \dots, n$ , on retrouve dans l'opérateur de covariance, la matrice de covariance  $\tilde{\Gamma}$  du vecteur  $\tilde{X} := (X(1), \dots, X(n))$ .

**4.2. Puissance spectrale.** Comme l'opérateur de covariance  $\Gamma$  est un opérateur positif, il est caractérisé par ses vecteurs propres et les valeurs propres associées. Supposons maintenant que la série chronologique  $\{X(n)\}_{n \in \mathbb{Z}}$  soit stationnaire au sens large, alors

$$\forall m, n \in \mathbb{Z}, \quad \Gamma(m, n) = R(m - n),$$

donc, si  $\{a_n\}_{n \in \mathbb{Z}} \in \ell^2(\mathbb{Z})$ , on a

$$(\Gamma a)(n) = \sum_{j \in \mathbb{Z}} a_j \Gamma(n, j) = \sum_{j \in \mathbb{Z}} a_j R(n - j) = (a * R)(n).$$

Ainsi  $\Gamma$  est un opérateur de convolution discret. On sait (Cours d'analyse de base) que les vecteurs propres d'une convolution discrète sont les exponentielles discrètes  $e_\omega := \{e^{i n \omega}\}_{n \in \mathbb{Z}}$ . On en déduit que

$$\Gamma e_\omega = \hat{R}(e^{i \omega}) e_\omega,$$

avec la valeur propre  $\hat{R}(e^{i\omega})$ , qui est positive puisque  $\Gamma$  l'est, et qui est donnée par la série de Fourier

$$\hat{R}(e^{i\omega}) := \sum_{j \in \mathbb{Z}} R(j) e^{-ij\omega} \geq 0.$$

Quand cette série converge vers une fonction, elle est appelée *densité spectrale de puissance du processus*. En général ce n'est pas une fonction mais une mesure  $\lambda$  positive sur le cercle unité  $\mathbb{T}$  du plan complexe, que l'on appelle *puissance spectrale du processus*.

Ainsi on a une densité spectrale  $f = \hat{R}(e^{i\omega})$  si la mesure  $\lambda$ , puissance spectrale, est absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{T}$ ,  $d\lambda = f(\omega) d\omega$ .

On retrouve l'auto-covariance du processus par Fourier inverse de la puissance spectrale :

$$R(j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{R}(e^{i\omega}) e^{ij\omega} d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ij\omega} d\lambda(\omega).$$

En particulier on a la *variance* du processus :

$$\sigma^2 = R(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{R}(e^{i\omega}) d\omega.$$

### 4.3. Bruit blanc.

DÉFINITION 2.13. On appelle bruit blanc un processus SSL dont les valeurs à des instants différents sont décorréélées, *i.e.*  $R(j) = \sigma^2 \delta_0(j)$ .

La densité spectrale du bruit blanc est donc constante:

$$\hat{R}(e^{i\omega}) = \sum_{j \in \mathbb{Z}} R(j) e^{-ij\omega} = \sigma^2.$$

## 5. Filtrage des processus stationnaires

**5.1. Filtrage homogène.** On a déjà vu que le filtrage d'un signal discret consistait en une convolution discrète avec le signal. Comme cette convolution est homogène dans le temps, cela ne modifie pas la stationnarité du processus. On a le lien suivant entre processus et processus filtré.

THÉORÈME 2.14 (Filtrage des processus SSL). *Soient  $H$  et  $G$  deux filtres discrets de réponse impulsionnelle respective  $\{h(n)\}_{n \in \mathbb{Z}}$ ,  $\{g(n)\}_{n \in \mathbb{Z}}$ , supposées dans  $\ell^1(\mathbb{Z})$ . Soit  $\{X(n)\}_{n \in \mathbb{Z}}$  un processus SSL tel que  $\{R_X(n)\}_{n \in \mathbb{Z}} \in \ell^1(\mathbb{Z})$ . Les deux processus*

$$Y(n) = h * X(n) = \sum_{j \in \mathbb{Z}} h(n-j) X(j) \quad \text{et} \quad Z(n) = g * X(n) = \sum_{j \in \mathbb{Z}} g(n-j) X(j)$$

sont SSL et on a

sigAlea51

$$(11) \quad \text{Cov}(Y(m), Z(n)) = h * \tilde{g} * R_X(m-n),$$

avec  $\tilde{g}(n) := g(-n)$ . La puissance spectrale de  $Y$  est

sigAlea52

$$(12) \quad \hat{R}_Y(e^{i\omega}) = \left| \hat{h}(e^{i\omega}) \right|^2 \hat{R}_X(e^{i\omega}).$$

DÉMONSTRATION. Tout d'abord, si  $\{X(n)\}$  est SSL,  $\mathbb{E}[X(n)] = \lambda$  et  $\mathbb{E}[(X(n) - \mathbb{E}[X(n)])^2] = \sigma^2$ . Mais alors, en utilisant l'inégalité triangulaire (avec  $X(n) = X(n) - \lambda + \lambda$ ) puis Cauchy-Schwarz,

$$\mathbb{E}[|X(n)|] \leq \mathbb{E}[|X(n) - \lambda|] + \lambda \leq \mathbb{E}[|X(n) - \lambda|^2]^{1/2} + \lambda = \sigma + \lambda.$$

REMARQUE 2.15. Nous venons en fait de montrer que si  $\{X(n)\}$  est un processus SSL alors la suite  $\{X(n)\}$  est uniformément bornée dans  $L^1(\Omega, \mathbb{P})$  et dans  $L^2(\Omega, \mathbb{P})$ .

Mais alors, la série  $Y(n) = \sum_{j \in \mathbb{Z}} h(n-j) X(j)$  converge dans  $L^1(\Omega, \mathbb{P})$  puisqu'elle est normalement convergente  $\sum_j |h(n-j)| \mathbb{E}[|X(j)|] < +\infty$ . De plus, on peut appliquer Fubini.

$$\mathbb{E}[Y(n)] = \sum_{j \in \mathbb{Z}} h(n-j) \mathbb{E}[X(j)] = \lambda \sum_{j \in \mathbb{Z}} h(n-j) = \lambda \hat{h}(0).$$



On utilise ensuite la formule (II0) qui donne

$$\begin{aligned}
\text{Cov}(Y(m), Z(n)) &= \sum_{j \in \mathbb{Z}} h(m-j) \sum_{k \in \mathbb{Z}} g(n-k) \text{Cov}(X(j), X(k)) \\
&= \sum_{j \in \mathbb{Z}} h(m-j) \sum_{k \in \mathbb{Z}} g(n-k) R_X(j-k) \\
&= \sum_{j \in \mathbb{Z}} h(m-j) \sum_{\ell \in \mathbb{Z}} \check{g}(\ell) R_X(j-n-\ell) \quad \ell = k-n \\
&= \sum_{j \in \mathbb{Z}} h(m-j) (\check{g} * R_X)(j-n) \\
&= \sum_{j \in \mathbb{Z}} h(k) (\check{g} * R_X)(m-n-k) \quad k = m-j \\
&= h * \check{g} * R_X(m-n).
\end{aligned}$$

En particulier, si on prend  $g = h$ , on obtient

$$\text{Cov}(Y(m), Y(n)) = h * \check{h} * R_X(m-n)$$

qui ne dépend que de  $m-n$  donc  $Y$  est bien SSL.

Enfin, comme la transformée de Fourier de  $h * \check{h}$  est  $|\widehat{h}(e^{i\omega})|^2$ , on obtient bien  $\widehat{R}_Y(e^{i\omega}) = |\widehat{h}(e^{i\omega})|^2 \widehat{R}_X(e^{i\omega})$ .  $\square$

**5.2. Densité spectrale de puissance.** Si la puissance spectrale  $\widehat{R}_X(e^{i\omega})$  d'un processus est à densité  $f$  par rapport à la mesure de Lebesgue, on peut retrouver cette *densité spectrale de puissance* (d.s.p.) de la manière suivante.

On fixe  $\xi \in ]0, \pi[$  et  $\Delta > 0$  (assez petit pour que  $0 < \xi - \Delta/2 < \xi + \Delta/2 < \pi$ ). Soit alors  $\{h_\Delta^\xi(n)\}$  le filtre défini par

$$\widehat{h}_\Delta^\xi(e^{i\omega}) = \begin{cases} \frac{\sqrt{\pi}}{\sqrt{\Delta}} & \text{si } |\omega - |\xi|| < \Delta/2 \\ 0 & \text{sinon} \end{cases}.$$

Notez que, avec Plancherel

$$\|h_\Delta^\xi\|_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \widehat{h}_\Delta^\xi(e^{i\omega}) \right|^2 d\omega = \frac{1}{2\pi} \left( \int_{-\xi-\Delta/2}^{-\xi+\Delta/2} \frac{\pi}{\Delta} d\omega + \int_{\xi-\Delta/2}^{\xi+\Delta/2} \frac{\pi}{\Delta} d\omega \right) = 1.$$

Soit maintenant  $\{X(n)\}$  un processus SSL et considérons le processus filtré par  $h_\Delta^\xi$ :  $Y(n) := X_\Delta^\xi(n) = h_\Delta^\xi * X(n)$ . C'est le processus obtenu en ne gardant que les fréquences dans le voisinage de  $\pm\xi$ . La formule (II2) permet de calculer la variance:

$$\text{Var}(X_\Delta^\xi) = R_{X_\Delta^\xi}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \widehat{R}_{X_\Delta^\xi}(e^{i\omega}) d\omega = \frac{1}{2\pi} \int_{|\omega-|\xi|| < \Delta/2} \frac{2\pi}{\Delta} \widehat{R}_X(e^{i\omega}) d\omega.$$

Mais  $R_X$  est une fonction continue (transformée de Fourier d'une mesure de probabilité). Il suffit alors de faire appel au lemme cidessous pour voir que

$$\lim_{\Delta \rightarrow 0} \text{Var}(X_\Delta^\xi) = \widehat{R}_X(\xi) + \widehat{R}_X(-\xi) = 2\widehat{R}_X(\xi).$$

Ainsi on a un moyen "expérimental" de trouver la d.s.p. d'un processus grâce à ce filtrage passe-bandes.

LEMME 2.16. Si  $g$  est une fonction continue en  $x_0$  alors  $\frac{1}{2a} \int_{x_0-a}^{x_0+a} g(t) dt \rightarrow g(x_0)$  lorsque  $a \rightarrow 0$ .

DÉMONSTRATION. Soit  $\varepsilon > 0$ , comme  $g$  est continue, il existe  $\eta > 0$  tel que si  $|t - x_0| \leq \eta$  alors  $|g(t) - g(x_0)| \leq \varepsilon$ . Mais alors, si  $a \leq \eta$ ,

$$\begin{aligned} \left| \frac{1}{2a} \int_{x_0-a}^{x_0+a} g(t) dt - g(x_0) \right| &= \left| \frac{1}{2a} \int_{x_0-a}^{x_0+a} g(t) - g(x_0) dt \right| \\ &\leq \frac{1}{2a} \int_{x_0-a}^{x_0+a} [g(t) - g(x_0)] dt \leq \frac{1}{2a} \int_{x_0-a}^{x_0+a} \varepsilon dt = \varepsilon \end{aligned}$$

ce qui est bien le résultat cherché.  $\square$

EXERCICE 2.1. Soit  $\alpha = (\alpha_k)_{k \in \mathbb{Z}}$  une suite non constante. Soit  $\{Y(k)\}_{k \in \mathbb{Z}}$  un processus SSL centré.

- (1) Soit  $\{X(k)\}$  le processus défini par  $X(k) = Y(k) + \alpha_k$ . Calculer  $\mathbb{E}[X(k)]$  et  $\text{Cov}(X(m), X(n))$ .  $\{X(k)\}$  est-il stationnaire ?
- (2) Soit  $\{Z(k)\}$  le processus obtenu à partir de  $X$  par filtrage  $Z(k) = X(k) - X(k-d)$  ( $d$  entier fixé). Quelle est la fonction de transfert du filtre ? Calculer  $\mathbb{E}[Z(k)]$  et  $\text{Cov}(Z(m), Z(n))$ .  $\{Z(k)\}$  est-il stationnaire au sens large ?

**5.3. Filtrage adapté.** Le théorème du filtrage des processus SSL permet de calculer les filtres optimaux pour détecter un signal *connu* dégradé par un bruit stationnaire. Par exemple une cible repérée par une onde radar, donc de *fréquences connues*, produit une onde réfléchie, donc de fréquences *connues* a priori, mais qui peut être perturbée par le bruit du milieu de propagation. Ce bruit peut être approximé par un processus aléatoire stationnaire dont on connaît la puissance spectrale.

Pour détecter la présence et la position d'un signal particulier, on filtre le signal bruité de façon à amplifier la réponse du signal par rapport à celle du bruit.

On considère ici que le signal bruité est un processus  $X(n) := f(n) + W(n)$ , où  $W(n)$  est un bruit de puissance spectrale  $\hat{R}_W(e^{i\omega})$  connue et de moyenne nulle, alors que  $f(n)$  est le signal déterministe, que l'on suppose centré en  $p$ , i.e.  $f(n) = g(n-p)$  où  $g$  est centrée (concentrée) en 0, que l'on cherche à repérer. Soit  $h(n)$  la réponse impulsionnelle du filtre.

$$(X * h)(n) = (f * h)(n) + (W * h)(n) = (g * h)(n-p) + (W * h)(n).$$

Lorsque  $n = p$ , on veut maximiser le rapport entre la réponse du signal et l'énergie moyenne du bruit:

$$\frac{S}{B} = \frac{(f * h)(p)^2}{\mathbb{E}[(W * h)(p)^2]} = \frac{(g * h)(0)^2}{\mathbb{E}[(W * h)(p)^2]}.$$

Avant d'énoncer un résultat, étudions un peu les deux quantités qui interviennent.

Tout d'abord, le bruit étant stationnaire, la puissance spectrale de  $Y := W * h$  est, grâce au théorème de filtrage des processus SSL,

$$(13) \quad \hat{R}_Y(e^{i\omega}) = |\hat{h}(e^{i\omega})|^2 \hat{R}_W(e^{i\omega})$$

et sa variance est

$$\boxed{\text{eq:snr1}} \quad (14) \quad \mathbb{E}[(W * h)(p)^2] = \mathbb{E}[(W * h)(0)^2] = \frac{1}{2\pi} \int_0^{2\pi} |\hat{h}(e^{i\omega})|^2 \hat{R}_W(e^{i\omega}) d\omega.$$

D'autre part

$$\boxed{\text{eq:snr2}} \quad (15) \quad g * h(0) = \frac{1}{2\pi} \int_0^{2\pi} \hat{g}(e^{i\omega}) \hat{h}(e^{i\omega}) d\omega = \frac{1}{2\pi} \int_0^{2\pi} \frac{\hat{g}(e^{i\omega})}{\sqrt{\hat{R}_W(e^{i\omega})}} \hat{h}(e^{i\omega}) \sqrt{\hat{R}_W(e^{i\omega})} d\omega.$$

Deux cas se présentent alors.

**Premier cas.**  $\hat{R}_W$  s'annule sur un ensemble (de mesure positive) sur lequel  $\hat{g}$  ne s'annule pas.

En d'autres termes, l'ensemble  $D = \{\omega \in [0, 2\pi] : \hat{R}_W(e^{i\omega}) = 0 \text{ \& } \hat{g}(e^{i\omega}) \neq 0\}$  est de mesure positive. On prend alors  $\hat{h}(e^{i\omega}) = \overline{\hat{g}(e^{i\omega})} \mathbf{1}_D$  et (15) montre que  $g * h(0) = \frac{1}{2\pi} \int_D |\hat{g}(e^{i\omega})|^2 d\omega \neq 0$  alors que (14) donne  $\mathbb{E}[(W * h)(p)^2] = 0$ . Le rapport signal/bruit est donc  $S/B = +\infty$ .

**Deuxième cas.**  $\widehat{R}_W$  et  $\widehat{g}$  s'annulent en même temps.

Dans ce cas, on utilise l'inégalité de Cauchy-Schwarz dans (II5) qui donne

$$g * h(0)^2 \leq \frac{1}{(2\pi)^2} \int_0^{2\pi} \frac{|\widehat{g}(e^{i\omega})|^2}{\widehat{R}_W(e^{i\omega})} d\omega \int_0^{2\pi} |\widehat{h}(e^{i\omega})|^2 \widehat{R}_W(e^{i\omega}) d\omega$$

et donc

$$\frac{S}{B} \leq \frac{1}{2\pi} \int_0^{2\pi} \frac{|\widehat{g}(e^{i\omega})|^2}{\widehat{R}_W(e^{i\omega})} d\omega.$$

De plus, cette inégalité est une égalité si on a une égalité dans Cauchy-Schwarz, c'est-à-dire si

$$\widehat{h}(e^{i\omega}) \sqrt{\widehat{R}_W(e^{i\omega})} = \frac{\widehat{g}(e^{i\omega})}{\sqrt{\widehat{R}_W(e^{i\omega})}}.$$

On a ainsi démontré le théorème suivant:

**THÉORÈME 2.17.** Soit  $D = \{\omega \in [0, 2\pi] : \widehat{R}_W(e^{i\omega}) = 0 \text{ \& } \widehat{g}(e^{i\omega}) = 0\}$ .

– Si  $D$  est de mesure  $> 0$  alors le filtre  $h$  donné par  $\widehat{h}(e^{i\omega}) = \widehat{g}(e^{i\omega}) \mathbf{1}_D$  donne un rapport signal sur bruit infini.

– Si  $D$  est de mesure nulle, alors le rapport signal/bruit est maximisé par le filtre  $h$  donné par

$$\widehat{h}(e^{i\omega}) \sqrt{\widehat{R}_W(e^{i\omega})} = \frac{\widehat{g}(e^{i\omega})}{\widehat{R}_W(e^{i\omega})}.$$

Dans ce cas, le rapport signal/bruit est  $\frac{S}{B} = \frac{1}{2\pi} \int_0^{2\pi} \frac{|\widehat{g}(e^{i\omega})|^2}{\widehat{R}_W(e^{i\omega})} d\omega$ .

## 6. Processus ARMA

**6.1. Définitions.** Pour analyser les réalisations d'un processus il est souvent nécessaire d'approximer ce processus par un modèle faisant intervenir le moins de paramètres possible. Dans le cas des processus SSL on utilise des modèles obtenus par filtrage d'un bruit blanc *i.e.* un bruit blanc  $W = \{W(n)\}_{n \in \mathbb{Z}}$  de variance  $\text{Var}(W(n)) = \sigma^2$  filtré par une réponse impulsionnelle  $h = \{h(n)\}_{n \in \mathbb{Z}}$ ,  $\{X(n)\}_{n \in \mathbb{Z}}$ , où  $\forall n \in \mathbb{Z}$ ,  $X(n) = (W * h)(n)$ . C'est encore un processus SSL dont la densité spectrale de puissance est  $\widehat{R}_X(e^{i\omega}) = |\widehat{h}(e^{i\omega})|^2 \sigma^2$ .

Les principaux processus étudiés sont les processus ARMA dans lesquels le filtrage est effectué par une fraction rationnelle. Rappelons que la transformée en  $z$  d'une série  $\{X(n)\}_{n \in \mathbb{Z}}$  a été définie au premier semestre comme :

$$\mathcal{Z}[X](z) = \sum_{n \in \mathbb{Z}} X(n) z^n.$$

Rappelons que si  $T_k X(n) = X(n - k)$  alors  $\mathcal{Z}[T_k X](z) = z^k \mathcal{Z}[X](z)$ .

**DÉFINITION 2.18.** Soit  $p$  et  $q$  deux entiers. On dit que  $X = \{X(n)\}_{n \in \mathbb{Z}}$  est un processus  $ARMA(p, q)$  s'il existe deux polynômes  $P$  de degré  $p$  et  $Q$  de degré  $q$ ,

$$P(z) = 1 + a_1 z + \dots + a_p z^p \quad \text{et} \quad Q(z) = 1 + b_1 z + \dots + b_q z^q$$

et un bruit blanc  $W = \{W(n)\}_{n \in \mathbb{N}}$  tel que

$$\boxed{\text{eq: arma1}} \quad (16) \quad P(z) \mathcal{Z}[X](z) = Q(z) \mathcal{Z}[W](z).$$

Notons que (II6) est équivalente à

$$\boxed{\text{eq: arma2}} \quad (17) \quad X(n) + a_1 X(n-1) + \dots + a_p X(n-p) = W(n) + b_1 W(n-1) + \dots + b_q W(n-q).$$

Si on définit  $a_0 = 1$ ,  $a_k = 0$  si  $k < 0$  et si  $k > p$  et  $b_0 = 1$ ,  $b_k = 0$  pour  $k < 0$  et pour  $k > q$ , alors (II6)-(II7) est aussi équivalent à

$$\boxed{\text{eq: arma3}} \quad (18) \quad a * X = b * W.$$

Dans le cas particulier où  $q = 0$  et  $\sigma^2 > 0$ , le processus est dit *Auto-Regressif d'ordre  $p$  AR( $p$ )*. Dans le cas particulier où  $p = 0$  et  $\sigma^2 > 0$ , le processus est dit à *moyenne mobile* (Moving Average)  $MA(q)$ .

EXERCICE 2.2. Soit  $\{X(n)\}_{n \in \mathbb{Z}}$  un processus ARMA. Montrer que sa fonction d'autocorrélation décroît exponentiellement vite vers 0 en l'infini.

DÉFINITION 2.19. Soit maintenant  $\{X(n)\}$  un processus stochastique (pas forcément stationnaire).  
 – Le *passé (linéaire)* de  $X(n)$  est le sous-espace de  $L^2(\Omega, \mathbb{P})$  engendré par les événements antérieurs

$$\mathcal{H}(n) = \overline{\text{Vect}\{X(k) : k < n\}}.$$

– Le *processus d'innovation* de  $\{X(n)\}$  est le processus défini par  $U(n) = X(n) - \mathbf{P}_{\mathcal{H}(n)}X(n)$  où  $\mathbf{P}_{\mathcal{H}(n)}$  est la projection orthogonale sur  $\mathcal{H}(n)$ .

– Un processus est dit *déterministe* (prévisible) si, pour tout  $n$ ,  $U(n) = 0$ , c'est-à-dire si  $X(n) \in \overline{\text{Vect}\{X(k) : k < n\}}$ .

– Un processus est dit *purement non déterministe* ou *régulier* si  $\bigcap_{j \in \mathbb{Z}} \mathcal{H}(j) = \{0\}$ .

EXEMPLE 2.20. Soit  $\{X(n)\}_{n=0, \dots, p-1}$  un processus aléatoire fini et on définit  $X(n)$  pour  $n \notin \{0, \dots, p-1\}$  par  $p$ -périodicité :  $X(n+p) = X(n)$ . Alors ce processus est déterministe. En particulier, le processus constant  $X(n) = X(0)$  est déterministe.

Notons aussi qu'une combinaison linéaire de processus déterministes  $Y(n) = \lambda_1 X_1(n) + \dots + \lambda_k X_k(n)$ , avec  $\lambda_1, \dots, \lambda_k$  des constantes, est encore déterministe.

EXERCICE 2.3. Soient  $A$  et  $B$  deux variables aléatoires gaussiennes centrées. Soit  $\lambda = p/q$  un rationnel et soit  $X(n) = A \cos 2\pi \lambda n + B \sin 2\pi \lambda n$ .

- (1) Montrer que le processus  $\{X(n)\}$  est SSL.
- (2) Le processus  $\{X(n)\}$  est-il auto-régressif ?
- (3) Montrer que  $\{X(n)\}$  est un processus déterministe.

LEMME 2.21. Soit  $\{X(n)\}$  un processus SSL de moyenne nulle et soit  $\{U(n)\}$  son processus d'innovation. Alors pour tout  $n \neq m$ ,  $U(n)$  et  $U(m)$  sont de moyenne nulle et décorrés, c'est-à-dire  $\text{Cov}(U(m), U(n)) = 0$ .

En particulier, si on définit  $W(n) = (\text{Var } U(n))^{-1/2} U(n)$  alors  $W(n)$  est un bruit blanc de variance 1.

DÉMONSTRATION. Notons que si les  $X(n)$  sont de moyenne nulle, toutes leurs combinaisons linéaires aussi. En particulier, la base orthonormale de  $\mathcal{H}(n)$  obtenue par orthonormalisation de Gram-Schmidt de  $X(n), X(n-1), \dots$  est de moyenne nulle. Donc tous les éléments de  $\mathcal{H}(n)$  sont de moyenne nulle. En particulier les  $U(n)$  sont aussi de moyenne nulle.

Supposons  $n > m$ . On a  $U(n) = X(n) - \mathbf{P}_{\overline{\text{Vect}\{X(n-1)\}}}X(n) \in \mathcal{H}(n-1)^\perp$  alors que  $U(m) = X(m) - \mathbf{P}_{\overline{\text{Vect}\{X(m-1)\}}}X(m) \in \overline{\text{Vect}\{X(m)\}} = \mathcal{H}(m) \subset \mathcal{H}(n-1)$  donc

$$\text{Cov}(U(n), U(m)) = \mathbb{E}[U(n)U(m)] = \langle U(n), U(m) \rangle_{L^2(\Omega, \mathbb{P})} = 0.$$

La dernière partie est triviale. □

Notez que le même argument montre que  $U(n)$  est décorréolé avec tout élément de  $\mathcal{H}(n-1)$ .

Si on reprend la démonstration, on voit aussi que

$$\mathcal{V} \nabla U(n) = \|U(n)\|_{L^2(\Omega, \mathbb{P})}^2 = \left\| X(n) - \mathbf{P}_{\overline{\text{Vect}\{X(n-1)\}}}X(n) \right\|_{L^2(\Omega, \mathbb{P})}^2 = \|X(n)\|_{L^2(\Omega, \mathbb{P})}^2 + \left\| \mathbf{P}_{\overline{\text{Vect}\{X(n-1)\}}}X(n) \right\|_{L^2(\Omega, \mathbb{P})}^2$$

### 6.2. Décomposition de Wold.

THÉORÈME 2.22 (Décomposition de Wold). Soit  $\{X(n)\}$  un processus stationnaire au sens large. Alors, il existe  $(b_n)_{n \geq 0}$  tel que  $b_0 = 1$ ,  $b_n = 0$  si  $n < 0$  et  $\sum_{n \geq 0} |b_n| < +\infty$ , un bruit blanc  $W(n)$  et un processus déterministe  $S(n)$  tels que

$$X(n) = W(n) + \sum_{k=1}^{+\infty} b_k W(n-k) + S(n).$$

De plus, en notant  $\mathcal{H}(n) = \overline{\text{Vect} \{X(k) : k < n\}}$  le passé de  $X(n)$ ,

- (i) pour tout  $n$ ,  $W(n) \in \mathcal{H}(n)$ ;
- (ii)  $S(n) \in \bigcap_{k \in \mathbb{Z}} \mathcal{H}(k)$ ;
- (iii)  $\{W(n)\}$  et  $\{S(n)\}$  sont décorrélés, i.e. pour tous  $m, n \in \mathbb{Z}$ ,  $\text{Cov}(W(n), S(m)) = 0$ .

Ainsi ce théorème nous dit que dans tout processus stationnaire, l'aléatoire ne peut provenir que d'un bruit blanc filtré.

**6.3. Filtres AR.** Nous allons ici nous concentrer sur les processus  $\{X(n)\}_{n \in \mathbb{Z}}$  auto-régressifs,  $AR(p)$ , c'est-à-dire qui satisfont une équation de récurrence

sigAllea60

$$(19) \quad X(n) + \sum_{k=1}^N a_k X(n-k) = W(n),$$

où  $W(n)$  est un bruit blanc de variance  $\sigma^2$ .

On peut voir cette équation comme une régression linéaire de  $X(n)$  sur les  $p$  valeurs "passées"  $X(n-1), \dots, X(n-p)$  auxquelles est ajoutée  $W(n)$  appelée *innovation* au temps  $n$ , non prévue par la régression :

$$(20) \quad X(n) = - \sum_{k=1}^p a_k X(n-k) + W(n).$$

Les constantes de régression  $\{a_k\}_{k=1, \dots, p}$  ainsi que la variance  $\sigma^2$  du bruit sont les paramètres du processus.

On peut aussi voir l'équation (19) comme une équation de convolution  $X * a(n) = W(n)$ . La fonction de transfert du filtre  $X \rightarrow W$ , i.e.  $X$  en entrée et  $W$  en sortie, est alors

$$\mathcal{Z}a(z) = 1 + \sum_{k=1}^p a_k z^k.$$

On dit que le filtre récursif  $a$  "blanchit" le processus  $X$  : son action sur  $X$ , i.e. sa convolution avec  $X$ , donne un bruit blanc.

Nous nous intéressons maintenant au filtre inverse  $W \rightarrow X$  dont la fonction de transfert est donnée par

$$\mathcal{Z}b(z) = \frac{1}{\mathcal{Z}a(z)} = \frac{1}{1 + \sum_{k=1}^p a_k z^k}.$$

En écrivant  $u = \sum_{k=1}^p a_k z^k$ , en remarquant que  $u \rightarrow 0$  quand  $z \rightarrow 0$  et en utilisant  $(1+u)^{-1} = \sum_{k \geq 0} (-1)^k u^k$ , on obtient

$$\mathcal{Z}b(z) = \sum_{k=0}^{\infty} (-1)^k \left( \sum_{k=1}^p a_k z^k \right)^k = \sum_{k \in \mathbb{Z}} b^k z^k$$

avec  $b_k = 0$  si  $k < 0$ ,  $b_0 = 1$ ,  $b_1 = -a_1$ ,  $b_2 = a_2 - a_1^2, \dots$ . En particulier, le filtre  $b$  est *causal*.

Posons maintenant  $Y(n) = X(n) - b * W(n)$  c'est-à-dire  $X(n) = b * W(n) + Y(n)$  et  $S(n) = b * W(n)$ . La proposition suivante est un corollaire direct du théorème de filtrage des processus.

LEMME 2.23. *Le processus  $S(n)$  est stationnaire et est appelé composante stationnaire de  $\{X(n)\}_{n \in \mathbb{Z}}$ . Sa puissance spectrale est donnée par*

$$\widehat{R}_S(\omega) = \frac{\sigma^2}{\left|1 + \sum_{j=1}^p a_j e^{ij\omega}\right|^2}.$$

Notons que  $\mathcal{Z}[a * b] = \mathcal{Z}[a]\mathcal{Z}[b] = 1$  donc  $a * b = \delta_0$ . Ainsi

$$a * Y(n) = a * X(n) - a * b * W(n) = a * X(n) - \delta_0 * W(n) = a * X(n) - W(n) = 0.$$

Cette équation peut se ré-écrire

**eq:recar**

$$(21) \quad Y(n) + a_1 Y(n-1) + \dots + a_p Y(n-p) = 0.$$

Soient maintenant  $r_1, r_2, \dots, r_p$  les racines de  $P(z) = 1 + a_1 z + \dots + a_p z^p$  donc  $P(z) = a_p \prod_{k=1}^p (z - r_k)$

et remarquons que

$$z^p + a_1 z^{p-1} + \dots + a_p = z^p P(1/z) = a_p z^p \prod_{k=1}^p (z^{-1} - r_k) = a_p \prod_{k=1}^p (1 - z r_k).$$

Les racines du polynôme caractéristique de [\(21\)](#) sont donc  $1/r_1, \dots, 1/r_p$ . Supposons pour simplifier qu'elles soient toutes distinctes, alors il existe  $\alpha_1, \dots, \alpha_p$  tels que

$$Y(n) = \alpha_1 \left(\frac{1}{r_1}\right)^n + \dots + \alpha_p \left(\frac{1}{r_p}\right)^n.$$

Mais alors, si  $|r_j| > 1$  pour  $j = 1, \dots, p$ ,  $Y(n) \rightarrow 0$ . Notons que si  $r_j$  est une racine d'ordre  $k$ , alors le facteur  $\alpha_j (1/r_j)^n$  doit être remplacé par  $(\alpha_{j,0} + \alpha_{j,1}n + \dots + \alpha_{j,k}n^k)(1/r_j)^n$  et le résultat est inchangé. Nous avons donc démontré la partie directe du théorème suivant, la réciproque est admise.

THÉORÈME 2.24. *Soit  $\{X(n)\}_{n \in \mathbb{N}}$  un processus  $AR(p)$  associé au polynôme  $P(z) = 1 + a_1 z + \dots + a_p z^p$ . Soient  $r_1, \dots, r_p$  les racines de ce polynôme,  $b * W(n)$  la partie stationnaire du processus et  $Y(n) = X(n) - b * W(n)$ . Alors  $Y(n) \rightarrow 0$  presque partout si et seulement si  $|r_1|, \dots, |r_p| > 1$ .*

## Matrices stochastiques

### 1. Rappels d'algèbre linéaire

Rappelons qu'une matrice complexe est trigonalisable et, plus précisément, possède une forme de Jordan:

– on considère d'abord le block de Jordan de taille  $m$

$$\mathcal{J}_{\lambda,m} = \begin{pmatrix} \lambda & 1 & & (0) \\ & \lambda & 1 & \\ & & \ddots & \\ (0) & & & \lambda \end{pmatrix} \in \mathcal{M}_m(\mathbb{C}).$$

Il faut remarquer que  $\mathcal{J}_{\lambda,m}$  a une seule valeur propre  $\lambda$  et un seul vecteur propre  $(1, 0, \dots, 0)$ . De plus,  $\mathcal{J}_{\lambda,m} = \lambda I + J$  et que  $J^m = 0$  donc pour  $k \geq m$

$$\mathcal{J}_{\lambda,m}^k = (\lambda I + J)^k = \sum_{j=0}^{m-1} \binom{k}{j} N^j \lambda^{k-j}.$$

En particulier, si  $|\lambda| < 1$ , alors  $\mathcal{J}_{\lambda,m}^k \rightarrow 0$  quand  $k \rightarrow \infty$ .

On considère ensuite la matrice définie par blocs

$$\mathcal{J}_{\lambda_1, m_1, \dots, \lambda_s, m_s} = \begin{pmatrix} \mathcal{J}_{\lambda_1, m_1} & & & (0) \\ & \mathcal{J}_{\lambda_2, m_2} & & \\ & & \ddots & \\ (0) & & & \mathcal{J}_{\lambda_s, m_s} \end{pmatrix}.$$

On s'autorise à répéter les  $\lambda_i$ .

– Soit ensuite  $M \in \mathcal{M}_n(\mathbb{C})$ ,  $\lambda_1, \dots, \lambda_s$  ses valeurs propres (pas forcément distinctes). Alors il existe  $m_1, \dots, m_s$  et une matrice inversible  $P$  telle que

$$M = P \mathcal{J}_{\lambda_1, m_1, \dots, \lambda_s, m_s} P^{-1}.$$

Supposons que  $\lambda$  apparaît dans la suite  $\lambda_1, \dots, \lambda_s$   $k$  fois, pour simplifier, disons  $\lambda_1 = \dots = \lambda_k = \lambda$  et  $\lambda_j \neq \lambda$  si  $j > k$ . Alors l'espace propre associé à  $\lambda$  est de dimension  $k$  et  $\lambda$  est racine de multiplicité  $m_1 + \dots + m_k$  du polynôme caractéristique de  $M$ . En particulier, si  $\lambda$  est une valeur propre simple, alors  $k = 1$  et  $m_1 = 1$ .

On peut aussi considérer la forme de Jordan triangulaire inférieur.

**DÉFINITION A.1.** Le rayon spectral d'une matrice  $A \in \mathcal{M}_n(\mathbb{C})$  est le module de la plus grande valeur propre (complexe) de  $A$ . On le note  $\rho(A)$ .

Notez que si  $A$  est une matrice réelle, on la considère comme une matrice complexe.

**LEMME A.2.** Soit  $A \in \mathcal{M}_n(\mathbb{C})$ , alors  $\rho(A) = \rho(A^t)$ . De plus, si  $\|\cdot\|$  et  $\|\|\cdot\|\|$  sur  $\mathbb{R}^n$  sont deux normes sur  $\mathbb{R}^n$  et  $\|A\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|\|Ax\|\|}{\|x\|}$ , alors  $\rho(A) \leq \|A\|$  i.e. si  $\|\|Ax\|\| \leq C\|x\|$  alors  $\rho(A) \leq C$ .

DÉMONSTRATION. Si la forme de Jordan de  $A$  est  $P\mathcal{J}_{\lambda_1, m_1, \dots, \lambda_s, m_s}P^{-1}$ , la forme de Jordan triangulaire inférieure de  $A^t$  est  $(P^t)^{-1}\mathcal{J}_{\lambda_1, m_1, \dots, \lambda_s, m_s}^tP^t$  donc  $A$  et  $A^t$  ont mêmes valeurs propres, a fortiori, même rayon spectral.

La deuxième partie est évidente, soit  $x$  un vecteur propre pour la valeur propre  $\lambda$ , alors  $x \neq 0$  donc  $\frac{x}{\|x\|}$  est encore vecteur propre pour  $\lambda$  et de norme 1. Donc  $|\lambda| = \|\lambda x\| = \|Ax\| \leq \|A\|\|x\| = \|A\|$  et, en prenant le max sur les valeurs propres, on trouve bien  $\rho(A) \leq \|A\|$ .  $\square$

## 2. Matrices stochastiques

Dans cet appendice, nous identifierons un vecteur  $v = (v_1, \dots, v_n) \in \mathbb{R}^n$  et la matrice colonne  $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathcal{M}_{n,1}$ .

DÉFINITION A.3. Une matrice  $A = [a_{i,j}]_{1 \leq i, j \leq n}$  est dite stochastique (suivant les colonnes) si

(i) pour tous  $i, j \in \{1, \dots, n\}$ ,  $0 \leq a_{i,j} \leq 1$

(ii) pour tout  $j \in \{1, \dots, n\}$ ,  $\sum_{i=1}^n a_{i,j} = 1$ .

On dira que  $A$  est dite stochastique suivant les lignes si  $A^t$  est stochastique suivant colonnes.

Ainsi, une matrice stochastique est une matrice dont les colonnes sont des vecteurs de probabilité, i.e.  $\sum_{i=1}^n a_{i,j}\delta_{x_i}$  est une mesure de probabilité. Notons que  $a_{i,j} \geq 0$  et  $\sum_i a_{i,j} = 1$  implique  $a_{i,j} \leq 1$ .

LEMME A.4. Si  $A$  et  $B$  sont des matrices stochastiques suivant les colonnes (resp. les lignes), alors  $AB$  est une matrice stochastique suivant les colonnes (resp. les lignes).

Si  $v$  est un vecteur de probabilité (écrit comme un matrice colonne), alors  $Av$  aussi.

DÉMONSTRATION. Montrons d'abord la deuxième partie:  $(Av)_i = \sum_{j=1}^n a_{i,j}v_j \geq 0$  puisque  $a_{i,j} \geq 0$  et  $v_j \geq 0$ . Mais alors

$$\sum_{i=1}^n (Av)_i = \sum_{i=1}^n \sum_{j=1}^n a_{i,j}v_j = \sum_{j=1}^n v_j \sum_{i=1}^n a_{i,j} = \sum_{j=1}^n v_j = 1$$

puisque  $\sum_{j=1}^n a_{i,j} = 1$  et que  $v$  est un vecteur de probabilité.

Enfin, si  $v_1, \dots, v_n$  sont les colonnes de  $B$ , alors les colonnes de  $AB$  sont  $Av_1, \dots, Av_n$  qui sont des vecteurs de probabilité, donc  $AB$  est bien stochastique.

Pour les matrices stochastiques par ligne, on utilise  $(AB)^t = B^tA^t$  et tous ces matrices sont stochastiques par colonne.  $\square$

LEMME A.5. Soit  $A$  une matrice stochastique suivant les colonnes, alors  $(1, \dots, 1)$  est un vecteur propre de  $A^t$  pour la valeur propre 1 et  $\rho(A) = 1$ .

DÉMONSTRATION. Soit  $A$  une matrice stochastique par les colonnes.

Tout d'abord  $(1, \dots, 1)$  est bien vecteur propre de  $A^t = [a_{j,i}]_{1 \leq i, j \leq n}$  puisque

$$\begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{n,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{n,2} \\ \vdots & \vdots & & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{n,n} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n a_{i,1} \\ \sum_{i=1}^n a_{i,2} \\ \vdots \\ \sum_{i=1}^n a_{i,n} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Ainsi  $\rho(A) = \rho(A^t) \geq 1$ .

Pour montrer que  $\rho(A^t) \leq 1$ , il suffit de montrer que  $\|A^t x\|_\infty \leq 1$  si  $\|x\|_\infty = 1$ , ce qui est évident:

$$\max_j |(A^t x)_j| = \max_j \left| \sum_{i=1}^n a_{i,j}x_i \right| \leq \max_j \sum_{i=1}^n |a_{i,j}x_i| \leq \max_j \sum_{i=1}^n |a_{i,j}| \max_i |x_i| = \|x\|_\infty.$$



□

THÉORÈME A.6. Soit  $A$  une matrice stochastique suivant les colonnes et supposons que 1 soit une valeur propre simple (de multiplicité 1) et strictement dominantes (i.e. les autres valeurs propres sont de modules  $< 1$ ). En particulier, l'espace propre correspondant à la valeur propre 1  $E_1$  est de dimension 1,  $E_1 = \text{Vect}e_1$ . Alors

- (i) il existe un vecteur de probabilité  $\pi_0$  tel que  $E_1 = \text{Vect}\pi_0$ ;
- (ii) si  $\pi$  un vecteur de probabilité, alors  $A^k\pi$  converge vers  $\pi_0$ .

DÉMONSTRATION. Soit  $\pi$  un vecteur de probabilité. Comme  $A^k$  est une matrice stochastique,  $\pi^{(k)} := A^k\pi$  est un vecteur de probabilité. Si  $\pi^{(n)}$  converge vers un vecteur  $v$  alors ce vecteur est un vecteur de probabilité puisqu'en passant à la limite dans  $\pi_j^{(k)} \geq 0$  et  $\sum_{j=1}^n \pi_j^{(k)} = 1$  on obtient  $v_j \geq 0$  et

$$\sum_{j=1}^n v_j = 1.$$

Par ailleurs,  $\pi^{(k+1)} = A\pi^{(k)}$  donc, en passant à la limite,  $v = Av$ . Ainsi  $v \in E_1$ . Donc  $v$  est le  $\pi_0$  cherché.

Il reste donc à démontrer que  $A^k\pi$  converge.

Si 1 est valeur propre simple de  $A$ , sa décomposition de Jordan est de la forme

$$A = P \begin{pmatrix} 1 & & 0 \\ & \mathcal{J}_{\lambda_2, m_2, \dots, \lambda_s, m_s} & \\ 0 & & \end{pmatrix} P^{-1}$$

avec  $|\lambda_j| < 1$ . Mais alors

$$A^k = P \begin{pmatrix} 1 & & 0 & & \\ & \mathcal{J}_{\lambda_2, m_2}^k & & & (0) \\ & & \mathcal{J}_{\lambda_3, m_3}^k & & \\ & & & \ddots & \\ (0) & & & & \mathcal{J}_{\lambda_s, m_s}^k \end{pmatrix} P^{-1} \rightarrow P \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} P^{-1}.$$

Donc  $A^k\pi$  converge. □



## Travaux pratiques 1: Simulation de variables aléatoires (densités, répartition, histogrammes et corrélation)

### 1. Introduction

L'objet de cet atelier est de se familiariser avec la notion de variables aléatoires et de s'initier à la programmation de celles-ci sous Matlab. Une variable aléatoire (v.a.) est caractérisée par sa densité de probabilité ou sa fonction de répartition. Nous allons voir comment générer des réalisations d'une variable aléatoire et comment l'histogramme des réalisations permet d'obtenir une approximation de la densité de probabilité ou de la fonction de répartition.<sup>1</sup> Enfin, la troisième partie abordera les notions de couple de v.a. et de corrélation. Deux densités de probabilité serviront d'illustration : la loi uniforme et la loi normale (ou gaussienne).

Dans un premier temps, on considère une variable aléatoire réelle (v.a.r.)  $X$  uniformément distribuée sur l'intervalle  $[a, b]$ . Sa densité de probabilité s'écrit :

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}.$$

- (1) Représentez graphiquement  $f_X(x)$ . Vérifiez que  $f_X(x)$  définit bien une densité de probabilité. Calculez la moyenne et la variance de cette variable. Comparez le résultat obtenu avec matlab et le résultat théorique (en prenant  $a = 0$ ,  $b = 1$ ).
- (2) Calculez sa fonction de répartition  $F_X(x)$  et représentez-la graphiquement.

Dans un second temps, la variable  $X$  est distribuée selon une loi normale de moyenne  $m$  et de variance  $\sigma$ . Sa densité de probabilité s'écrit :

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

- (3) Représentez graphiquement  $f_X(x)$ . Vérifiez que  $f_X(x)$  définit bien une densité de probabilité. Calculez la moyenne et la variance de cette variable. Comparez le résultat obtenu avec matlab et le résultat théorique.
- (4) Déterminez sa fonction de répartition  $F_X(x)$  et représentez-la graphiquement.

### 2. Génération de réalisations d'une variable aléatoire

**2.1. Loi uniforme.** Le logiciel Matlab dispose d'un générateur de nombres pseudo-aléatoires `rand` qui simule des v.a. indépendantes de loi uniforme sur  $[0, 1]$  (voir l'aide de la fonction pour plus d'informations sur son utilisation).

- (1) Tapez la commande `rand` dans la fenêtre de commande Matlab. Recommencez plusieurs fois de suite. A quoi correspondent les valeurs renvoyées ?

Nous allons utiliser cette fonction pour générer une suite de réalisations  $x_1, x_2, \dots, x_N$  de la variable aléatoire  $X$ . Pour cela, créez un fichier de commandes et saisissez les lignes suivantes :

<sup>1</sup>Rappelons que la fonction de répartition  $F_X(x_0)$  de la v.a.  $X$  est définie par :

$$F_X(x_0) = \mathbb{P}(X < x_0) = \int_{-\infty}^{x_0} f_X(t) dt.$$

```

% nettoyage
close all, clear all,
% parametres de simulation
N=100;
n=1:N;
% realisations
x=rand(1,N);
% observations
figure(1)
plot(n,x,'*')
title('serie d'evenements')
xlabel('tirages')
ylabel('evenements')

```

- (2) Interprétez et commentez la figure obtenue.
- (3) Calculez la moyenne et la variance empiriques des  $N$  réalisations.<sup>2</sup>  
Comparez avec la moyenne et la variance théoriques. Conclure.
- (4) La fonction `rand` ne permet que de générer des réalisations d'une v.a. uniformément distribuée sur  $[0, 1]$ . Proposez une solution permettant de générer des réalisations d'une v.a. uniformément distribuée sur :
  - (a) l'intervalle  $[1, 2]$  ;
  - (b) l'intervalle  $[0, 3]$  ;
  - (c) l'intervalle  $[-1, 1]$ .
Vérifiez que vos propositions fournissent les résultats attendus.

**2.2. Loi normale.** Le logiciel dispose également d'un générateur de nombres pseudo-aléatoires `randn` qui simule des v.a. indépendantes distribuées selon une loi normale centrée et réduite (de moyenne  $m = 0$  et de variance  $\sigma = 1$ .)

- (5) Modifiez votre programme afin de générer une suite de réalisations d'une variable distribuée selon la loi normale centrée et réduite.
- (6) Interprétez et commentez la figure obtenue.
- (7) Comparez la moyenne et la variance empiriques des  $N$  réalisations avec celles théoriques. Conclure.
- (8) La fonction `randn` permet de générer uniquement des réalisations de v.a. de loi normale centrée et réduite. Proposez une solution permettant de générer des réalisations d'une v.a. distribuées selon une loi normale avec :
  - (a)  $m = 2$  et  $\sigma = 1$  ;
  - (b)  $m = 0$  et  $\sigma = 3$  ;
  - (c)  $m = 2$  et  $\sigma = 3$ .
Vérifiez que vos propositions fournissent les résultats attendus.

### 3. Densité de probabilité et histogramme

La probabilité pour que la v.a.  $X$  soit comprise entre  $x$  et  $x + \Delta x$  (avec  $\Delta x$  petit) peut être approchée par  $f_X(x)\Delta x$  :

$$\int_x^{x+\Delta x} f_X(t) dt \simeq f_X(x)\Delta x.$$

---

<sup>2</sup>La moyenne empirique est définie par  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$  et la variance empirique par  $\sigma = \frac{1}{N} \sum_{i=1}^N (x_i - m)^2$  où les  $x_i$  sont les tirages de  $X$  réalisés par matlab.

Ainsi, à partir des  $N$  réalisations de la v.a., on peut approcher  $f_X(x)\Delta x$  par le rapport entre le nombre de points  $x_k$  situés dans l'intervalle  $[x; x + \Delta x]$ , noté  $N_x$ , et le nombre de points total  $N$  :

$$f_X(x)\Delta x \simeq \frac{N_x}{N}.$$

Par conséquent, une approximation de la densité de probabilité peut être obtenue par un histogramme normalisé de la suite de réalisations de la v.a. :

$$f_X(x) \simeq \frac{N_x}{N\Delta x}.$$

- (1) Calculez l'histogramme normalisé de la suite de réalisations de  $X$  à l'aide de la commande `hist(x,M)` où  $M$  représente le nombre d'intervalles (ou de classes) (voir l'aide de la fonction `hist`). Faites-le avec une loi uniforme, puis avec une loi normale. Obtenez-vous une bonne estimation de la densité de probabilité de  $X$  ? Étudiez l'influence des paramètres  $N$  et  $M$ .

Rappelons que la fonction de répartition  $F_X(x_0)$  de la v.a.  $X$  est définie par :

$$F_X(x_0) = \mathbb{P}(X < x_0) = \int_{-\infty}^{x_0} f_X(t) dt.$$

Pour estimer  $F_X(x_0)$ , il suffit de calculer la somme cumulée de l'histogramme (normalisé), *i.e.* compter le nombre de réalisations  $x_n$  inférieures à  $x_0$  et de diviser par le nombre total de points  $N$  :

$$F_X(x_0) \simeq \sum_{x \leq x_0} \frac{N_x}{N}$$

les sommations étant effectuées sur les classes d'amplitude inférieures à  $x_0$ .

- (3) Approchez la fonction de répartition de  $X$  à partir de l'histogramme normalisé (en utilisant la fonction `cumsum` pour réaliser les sommations).

#### 4. Couple, transformation linéaire et corrélation

Dans cette partie, on s'intéresse à la construction d'un couple de v.a. corrélées à partir d'un couple de variables décorréelées (et même indépendantes). On considère un couple de v.a.  $(X_1, X_2)$  distribuées selon une loi normale centrée et réduite.

- (1) Quel est le coefficient de corrélation du couple  $(X_1, X_2)$ .
- (2) Générez 2000 réalisations de chacune des variables.
- (3) Tracez dans le plan  $(X_1, X_2)$  les réalisations d'une v.a. en fonction de l'autre (sans relier les points entre eux). Commentez le nuage de points obtenus.

On considère maintenant deux v.a. obtenues à partir de  $(X_1, X_2)$  par la transformation linéaire suivante :

$$\begin{aligned} Y_1 &= \alpha_1 X_1 + \varepsilon_1 X_2 \\ Y_2 &= \varepsilon_2 X_1 + \alpha_2 X_2 \end{aligned}$$

- (4) Calculez la covariance et le coefficient de corrélation du couple  $(Y_1, Y_2)$ .
- (5) Intéressez-vous au cas  $\alpha_1 = \alpha_2 = 1$  et  $\varepsilon_1 = \varepsilon_2 = \varepsilon$ . Étudiez le coefficient de corrélation  $\rho(\varepsilon)$ . Déterminez la ou les valeurs qui maximisent ou minimisent ce coefficient de corrélation. Déterminez les limites lorsque  $\varepsilon$  tend vers  $\pm\infty$ .
- (6) On fixe  $\varepsilon = 0, 1$ . Générez 2000 réalisations de chacune des variables. Tracez-les dans le plan  $(Y_1, Y_2)$  et comparez le nuage de points obtenu avec celui obtenu avec les v.a. indépendantes (donc décorréelées).
- (7) Recommencer avec les valeurs de  $\varepsilon$  déterminées à la question précédente. Conclure.

### 5. Algorithme de Box-Muller

Soient  $U$  et  $V$  deux variables aléatoires indépendantes, de loi uniforme sur  $[0, 1]$ . Soient  $X$  et  $Y$  les variables aléatoires

$$X = \sqrt{-2 \ln U} \cos 2\pi V \quad \text{et} \quad Y = \sqrt{-2 \ln U} \sin 2\pi V.$$

Montrer que  $X$  et  $Y$  sont deux variables aléatoires indépendantes de loi  $\mathcal{N}(0, 1)$ .

Utilisez le résultat précédent et la fonction `rand` pour simuler une v.a. de loi normale standard : effectuer  $N$  simulations, et comparer (sur une même figure) l'histogramme des valeurs obtenues, le résultat de la simulation à l'aide de la fonction `randn` et la fonction densité théorique. Que pensez-vous du résultat ?

### 6. Simulation d'un vecteur gaussien

On rappelle que si  $X = (X_1, \dots, X_n)$  est un vecteur gaussien de moyenne nulle et de covariance  $I$  (matrice identité), et si  $A \in \mathbb{M}_n(\mathbb{R})$  est telle que  $AA^t = C$ , alors  $Y = AX + \mu$  est un vecteur gaussien de moyenne  $\mu$  et de covariance  $C$ .

- (1) En déduire un algorithme de simulation de vecteurs gaussiens.
- (2) Application : faire  $N$  simulations d'un vecteur gaussien de covariance  $C$  et de moyenne  $\mu$  avec  $C = \begin{pmatrix} 5 & 3 & 2 \\ 3 & 4 & 1 \\ 2 & 1 & 2 \end{pmatrix}$  et  $\mu = (-2, 1, 4)$  (utiliser la fonction `chol`). Vérifier les résultats à l'aide des fonctions `mean` et `cov`.

### 7. Méthode d'inversion

Rappelons le lemme suivant fort utile pour simuler des variables aléatoires :

LEMME B.1. *Soit  $F$  une fonction  $\mathbb{R} \rightarrow [0, 1]$  croissante, ayant des limites à droite et à gauche en tout point. Soit  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ . Alors  $X = F^{-1}(U)$  est une variable aléatoire dont la fonction de répartition est  $F$  i.e.  $\mathbb{P}(X \leq \lambda) = F(\lambda)$ .*

- (1) Utiliser cette méthode pour simuler une loi exponentielle de paramètre  $\lambda$ . Faire  $N$  simulations et comparer sur une même figure l'histogramme des valeurs obtenues avec la densité théorique.
- (2) Faire de même pour une loi de Cauchy.
- (3) Faire de même pour une loi de Weibull de paramètres  $a$  et  $b$  strictement positifs, dont la densité est donnée par

$$f(x) = abx^{b-1}e^{-ax^b}.$$

- (4) On considère une variable aléatoire discrète  $X$  dont la loi est donnée par  $\mathbb{P}(\{X = 0\}) = 0.6$ ,  $\mathbb{P}(\{X = 1\}) = 0.2$  et  $\mathbb{P}(\{X = 2\}) = 0.2$ . Effectuez un tirage de  $X$ .
- (5) **Simulation de v.a. discrets.**
  - (a) *Représentation d'une loi de probabilité.* On se donne une loi de probabilité portée par les entiers de  $\{0, \dots, k\}$ . Pour représenter cette loi, il suffit de fournir les valeurs  $\mathbb{P}(\{X = i\}) = p_i$ ,  $i = 0, \dots, k$ . On les représente sous la forme d'un vecteur de longueur  $k + 1$  contenant  $(p_0, p_1, \dots, p_k)$ .
  - (b) *Fonction de répartition.* On représente la fonction de répartition d'une loi discrète en donnant le vecteur  $(a_0, a_1, \dots, a_k)$  tel que  $a_i = \mathbb{P}(\{X \leq i\})$ . Comment calculer avec `matlab` la fonction de répartition à partir de la loi ? (une seule ligne de code).
  - (c) *Tirage d'une loi discrète.* Il s'agit de calculer  $F^{-1}(U)$  où  $U$  est une variable aléatoire de loi uniforme. Pour cela, on cherche le plus grand indice  $i$  tel que  $F(i) < U$ . Implémenter une fonction `proba(p)` qui réalise le tirage. Vérifier en traçant un histogramme.

### 8. Méthode du rejet

La méthode du rejet est une méthode pour simuler des variables aléatoires uniformes sur un domaine  $\mathcal{D}$  de  $\mathbb{R}^n$ . Elle est basée sur le lemme suivant :

LEMME B.2. Soient  $\mathcal{D}$  et  $\mathcal{D}'$  deux domaines de  $\mathbb{R}^k$  mesurables, de mesure finie et non nulle, et tels que  $\mathcal{D} \subset \mathcal{D}'$ . Soit  $X$  un point aléatoire uniforme sur  $\mathcal{D}'$ , alors la loi conditionnelle de  $X$  sachant que  $X \in \mathcal{D}$  est la loi uniforme sur  $\mathcal{D}$ .

Pour utiliser ce lemme, on prend  $\mathcal{D}' = [-R/2, R/2]^k$  un cube qui contient  $\mathcal{D}$ . Il suffit alors de simuler  $k$  variables aléatoires indépendantes  $X_1, \dots, X_k$  de loi uniforme sur  $[0, 1]$  et alors  $X = (R(X_1 - 1/2), R(X_2 - 1/2), \dots, R(X_k - 1/2))$ .

EXERCICE B.1. Pour tirer une loi uniforme sur l'intérieur d'une ellipse (*i.e.* un domaine de la forme  $(x/a)^2 + (y/b)^2 \leq 1$ ), il suffit de simuler une loi uniforme sur un domaine rectangulaire contenant l'ellipse, et de rejeter les points hors de l'ellipse. Implémenter un tel tirage dans une fonction `rellipse`.

EXERCICE B.2. Toujours à propos de la simulation de la loi uniforme sur le disque, que pensez-vous de la méthode suivante: on commence par choisir  $x$  uniforme entre  $-1$  et  $1$  puis on prend  $y$  uniforme entre  $-\sqrt{1-x^2}$  et  $\sqrt{1-x^2}$ . Visualiser  $N$  simulations obtenues par cette méthode. La loi est-elle uniforme?

### 9. Solution: simulation de v.a. discrète

```

fonction [y] = rproba(p)
% Simule une variable alatoire de loi p
cdf = cumsum(p);
% cr la fonction de rpartition
u = rand();
% loi uniforme
y = sum(cdf < u) + 1;
% inverse la fonction de rpartition.

```

### 10. Théorème centrale limite

- (1) Faire  $N$  tirages indépendants de  $n$  variables aléatoires indépendantes uniformes sur  $[-1/2; 1/2]$ . En utilisant la fonction `sum`, obtenir  $N$  tirages indépendants de la somme de  $n$  variables aléatoires indépendantes uniformes sur  $[-1/2, 1/2]$ , normalisée par le facteur  $n^{-1/2}$  et par leur écart-type.
- (2) Faire l'histogramme des  $N$  valeurs obtenues, et le comparer sur un même graphique à la fonction densité de la loi normale standard  $\mathcal{N}(0, 1)$ .
- (3) Tracer sur un même graphique les fonctions de répartition empiriques (`cumsum`) obtenues pour des valeurs de  $n$  allant de 1 à 6, et comparer à la fonction de répartition de la loi normale standard (`erf`).





## Travaux pratiques 2: chaînes de Markov

### 1. Chaîne de Markov

Une matrice de transition  $\mathcal{P} = [p_{j,k}]_{1 \leq i, j \leq n}$  d'une chaîne de Markov (aussi appelé matrice de Markov) a les propriétés suivantes:

- (i) Comme  $p_{j,k} = \mathbb{P}(\{X = j\}|\{X = k\})$ , toutes ses entrées sont dans  $[0, 1]$ ,  $0 \leq p_{j,k} \leq 1$ .
- (ii) On a

$$\sum_k p_{j,k} = \mathbb{P}\left(\{X = j\}|\bigcup_k \{X = k\}\right) = \frac{\mathbb{P}(\{X = j\})\mathbb{P}(\bigcup_k \{X = k\})}{\mathbb{P}(\{X = j\})} = 1.$$

La somme des éléments d'une ligne vaut 1.

On peut donc simuler une matrice de Markov aléatoire comme suit: on prend une matrice  $(n, n)$  aléatoire  $R = \text{rand}(n, n)$ . On détermine la somme de ses ligne `ligne=sum(R,1)` et on divise chaque ligne par la somme des éléments de cette colonne `M=R./(ones(n,1)*ligne)`.

- (1) Implémentez cette procédure et vérifiez qu'elle fonctionne.
- (2) Écrire une procédure qui fournit une loi initiale au hasard (*indication*: utiliser `rand` et une normalisation).
- (3) Écrire une procédure qui, à partir d'une matrice de Markov et d'une loi initiale effectue les opérations suivantes:
  - (a) Tire une position initiale  $X(0)$  au hasard, suivant la loi initiale;
  - (b) Pendant  $n$  étape (paramètre de la procédure), tire la nouvelle position  $X(n)$ , connaissant la position  $X(n-1)$  suivant la loi  $p_{X(n),\cdot}$  donnée par la matrice de transition.
  - (c) Trace la trajectoire de  $X$  *i.e.* la courbe  $(n, X(n))$ .

Fixer une matrice de markov et une probabilité initiale et utiliser cette procédure pour tracer quelques trajectoires de longueur 500.

- (4) Déterminer l'évolution  $\pi_k = \mathcal{P}^k \pi_0$  de la loi de  $X$ , tracer l'histogramme.
- (5) À l'aide de la commande `eig` de `matlab`, déterminer la loi invariante  $\pi_i$  de  $\mathcal{P}$ .
- (6) Tracer l'évolution de  $\|\pi_k - \pi_i\|$ .
- (7) Même question dans le cas (plus déterministe) suivant: Les consommateurs de 3 produits sont répartis en 50% pour  $P_1$ , 30% pour  $P_2$  et 20% pour  $P_3$ . Après chaque mois, 60% restent fidèles à  $P_1$  contre 70% pour  $P_2$  et 90% pour  $P_3$ . Les autres se réorientent entre les deux produits de manière équiprobable.

### 2. Solution

```
function [m] = rmarkovmat(n)
% Genere une matrice de Markov de taille n
m = rand(n,n);
col = sum(m,2);
m = m ./ (col * ones(1,n));
end
```



## Travaux pratique: débruitage

Ces travaux pratiques sont basés sur l'utilisation de deux "toolbox" matlab disponibles ici  
[http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/toolbox\\_signal.zip](http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/toolbox_signal.zip)  
[http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/toolbox\\_general.zip](http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/toolbox_general.zip)

La première chose à faire est de les installer sur votre compte. Pour cela, commencer par les 'dé-zipper' (commande unix `unzip`) puis ajoutez ces commandes à votre 'path' matlab:

```
getd = @(p)path(p,path)
getd('toolbox_signal/');
getd('toolbox_general/');
```

### 1. Du bruit dans des signaux et images

Dans ce TP, nous allons simuler l'acquisition d'un signal en lui ajoutant du bruit (suivant un modèle que nous choisirons). Cela est utile car nous connaissons le signal ou l'image d'origine, qui sera propre, et que nous pourrions comparer l'efficacité des algorithmes de reconstruction.

Nous chargeons donc un signal:

```
name = 'piece-regular';
n = 1024;
x0 = load_signal(name,n);
x0 = rescale(x0);
```

puis nous lui ajoutons un bruit blanc gaussien et traçons le signal d'origine et sa version bruitée:

```
sigma = .04; % niveau de bruit
x = x0 + sigma*randn(size(x0));
clf;
subplot(3,1,1);
plot(f0); axis([1 n 0 1]);
title('Signal propre');
subplot(3,1,2);
plot(f-f0); axis([1 n -3*sigma 3*sigma]);
title('Bruit');
subplot(3,1,3);
plot(f); axis([1 n 0 1]);
title('Signal bruité');
```

Nous faisons de même pour une image

```
n = 256;
name = 'hibiscus';
M0 = load_image(name,n);
M0 = rescale( sum(M0,3) );
```

```
sigma = .08; % niveau de bruit
M = M0 + sigma*randn(size(M0));
clf
imageplot(M0, 'Image propre', 1,3,1);
```

```
imageplot(M-M0, 'Bruit', 1,3,2);
imageplot(clamp(M), 'Image bruitée', 1,3,3);
```

Regardons les statistiques du bruit

```
nbins = 51;
[h,t] = hist( M(:)-M0(:), nbins ); h = h/sum(h);
subplot(3,1,2);
bar(t,h);
axis([-sigma*5 sigma*5 0 max(h)*1.01]);
```

D'autres modèles de bruit peuvent être raisonnables. Par exemple, le bruit peut être uniformément distribué dans  $[-a, a]$  où  $a$  est choisi pour que la variance soit  $\sigma$ .

Recommencez les opérations ci-dessus en remplaçant les définitions de  $f$  et  $M$  par

```
a = sqrt(3)*sigma;
M = M0 + 2*(rand(N,N)-.5)*a;
f = f0 + 2*(rand(n,1)-.5)*a;
```

Enfin, pour un bruit suivant une distribution exponentielle de variance  $\sigma$

```
W = log(rand(n,1)).*sign(randn(n,1));
W = W/std(W(:))*sigma;
f = f0 + W;
W = log(rand(N,N)).*sign(randn(N,N));
W = W/std(W(:))*sigma;
M = M0 + W;
```

## 2. Débruitage par filtrage linéaire

Nous allons maintenant utiliser un filtrage pour débruiter le signal et l'image. La méthode utilisée ici est rudimentaire, mais est à la base de méthodes plus performantes.

Dans un premier temps, nous allons simplement faire une convolution du signal/de l'image par un filtre.

Dans un premier temps, nous considérons une gaussienne et nous utilisons la variance du filtre comme paramètre mesurant la largeur du filtre:

```
% largeur du filtre
mu = 4;
% calcul d'une gaussienne de variance \mu et de m\^eme longueur que le signal mesure x
t = (-length(x)/2:length(x)/2-1)';
h = exp( -(t.^2)/(2*mu^2) );
h = h/sum(h(:));
```

On calcule sa transformée de Fourier à l'aide de la FFT et on trace les deux

```
% Fourier transform of the (centered) filter
hf = real( fft(fftshift(h)) );
hf = ifftshift(hf);
% display
clf;
subplot(2,1,1);
plot( t,h ); axis('tight');
title('Filtre h');
subplot(2,1,2);
plot( t,hf ); axis('tight');
title('Transformée de Fourier');
```

On calcule la convolution du signal avec le filtre

```
% Transformée de Fourier du signal bruité
xf = fft(x);
```

```

% Transformee de Fourier du signal filtre
xhf = xf .* hf;
% Signal debruite
xh = real( ifft(xhf) );
Puis on trace le signal debruite.
clf;
subplot(2,1,1);
plot( t,x ); axis('tight');
title('bruite');
subplot(2,1,2);
plot( t,xh ); axis('tight');
title('Debruite');

```

Que constatez vous?

Regardons ce qui s'est passé du côté Fourier:

```

% log de la transformee de Fourier
epsilon = 1e-10;
L = log10(epsilon+abs(fftshift(xf)));
Lh = log10(epsilon+abs(fftshift(xhf)));
% trace de la transformee de Fourier
clf;
subplot(2,1,1);
plot( t, L, '-');
axis([-length(x)/2 length(x)/2 -4 max(L)]);
title('log de Fourier bruité. ');
subplot(2,1,2);
plot( t, Lh, '-');
axis([-length(x)/2 length(x)/2 -4 max(L)]);
title('log de Fourier debruite. ');

```

Le choix d'un bon paramètre  $\mu$  est difficile puisqu'il doit prendre en compte à la fois la variance du bruit et le signal d'origine. En faisant varier  $\mu$ , essayer de déterminer un paramètre optimal.

Nous allons faire la même chose pour les images (en utilisant les fonctionnalités de la "toolbox"). Le débruitage s'effectue ici par floutage gaussien (en considérant l'image comme périodique, ce qui est plus rapide)

```

options.bounds = 'per';
% nombre de pixels du filtre (largeur=variance de la gaussienne)
mu = 10;
Mh = perform_blurring(M,mu,options);
clf;
imageplot(clamp(M), 'Bruité', 1,2,1);
imageplot(clamp(Mh), 'Flouté', 1,2,2);

```

En faisant varier  $\mu$  et en calculant le rapport signal/bruit (fonction `snr`):

```

mulist = linspace(1,20,6);
clf;
for i=1:length(mulist)
    mu = multist(i);
    Mh = perform_blurring(M,mu,options);
    sb[i]=snr(M0,Mh);
    imageplot(clamp(Mh), strcat(['\mu=' num2str(mu) 'SNR=' num2str(sb[i]) 'dB']), 2,3,i);
end

```

Tracer la courbe du rapport signal sur bruit en fonction du  $\mu$

```

plot({mulist,sb})

```

Puis déterminer  $\mu$  pour lequel le débruitage est optimal et recalculer l'image correspondante

```
Mgauss = perform_blurring(M,mu,options);
%en remplaçant mu par la valeur optimale.
```

Le filtre optimal n'est pas une gaussienne, mais le filtre de wiener, qui est calculable à l'aide de la toolbox:

```
[Mwien,Hwien] = peform_wiener_filtering(M0,M,sigma);
et qu'on trace
k = 5;
clf;
imageplot(Hwien(n/2-k+2:n/2+k,n/2-k+2:n/2+k), 'filtre de Wiener');
```

Enfin, on affiche le résultat

```
clf;
imageplot( clamp(M0), strcat(['Image bruitée']), 1,3,1);
imageplot( clamp(Mgauss), strcat(['Debruitage Gaussien,
SNR=' num2str(snr(M0,Mgauss)) 'dB']), 1,3,2);
imageplot( clamp(Mwien), strcat(['Debruitage Wiener,
SNR=' num2str(snr(M0,Mwien)) 'dB']), 1,3,3);
```