Université de Montpellier II
Préparation Agrégation de Mathématiques.
Compléments pour l'épreuve de modélisation en probabilités statistiques.

Pascal Azerad ©

14 octobre 2009

Table des matières

1 Simulation de Variable Aléatoires **7**
1.1 Nombres pseudo-aléatoires 7
1.1.1 Un problème de définition 7
1.1.2 Un autre point de vue issu de la théorie de la complexité. 8
1.1.3 Le point de vue statistique. 8
1.1.4 Générateurs de nombres pseudo-aléatoires. 8
1.2 Lois non uniformes 10
1.2.1 Inversion de la fonction de répartition 10
1.2.2 Algorithme de rejet 12
1.3 Lois normales 13
1.3.1 Box-Muller 13
1.3.2 Vecteurs gaussiens 14
1.4 Méthode de Monte Carlo 15
1.5 Programmes scilab. 17

2 Chaînes de Markov. **25**
2.1 Matrice de transition 25
2.2 Propriété de Markov fort 26
2.3 Classification des états 27
2.4 Irréductibilité. 27
2.5 Mesure invariante 28
2.6 Convergence vers la loi stationnaire et propriété ergodique. . . . 28
2.7 Simulation de CDM et illustration en `Scilab`. 29

3 Espérance conditionnelle **31**
3.1 L'espérance mathématique : une moyenne probabiliste 31
3.2 Probabilité conditionnelle 32
3.3 Espérance conditionnelle : définitions et propriétés fondamentales 32
3.3.1 une moyenne relative 32
3.3.2 une projection orthogonale ou une prédiction optimale . 36
3.4 Loi conditionnelle. Brefs rappels. 36
3.5 Exercices 37

4 Processus stochastiques et temps d'arrêts **39**
4.1 Filtration : l'histoire d'un processus 39
4.2 Loi d'un processus 39
4.3 Trajectoires d'un processus 40
4.4 Temps d'arrêt 40

4.5 Exercices 42

5 Martingales **43**
5.1 Définition et interprétations 43
5.2 Martingales et temps d'arrêts 44
5.3 Convergence presque sûre 46
5.3.1 Le cas des surmartingales positives 46
5.3.2 Le cas des sous-martingales bornées dans L^1. 49
5.4 Martingales uniformément intégrables 51
5.4.1 Uniforme intégrabilité et convergence dans L^1 51
5.4.2 Martingales régulières 52
5.5 Exercices et problèmes. 54

Préambule

Ce fascicule volontairement succinct poursuit les buts suivants :
- être assimilable par un étudiant de master.
- servir de tremplin pour les cours de troisième cycle et de vade-mecum pour les agrégatifs.

Ce document est une première version, l'auteur est très reconnaissant pour toute erreur, coquille ou commentaire qu'on voudra bien lui adresser à

`azerad@math.univ-montp2.fr`

Chapitre 1

Simulation de Variable Aléatoires

Référence du chapitre : livre de B. Ycart [17]

1.1 Nombres pseudo-aléatoires

1.1.1 Un problème de définition

Supposons qu'on veuille simuler un jeu de pile ou face. Voici trois exemples de suites de longueur n

$$000000000\ldots$$

$$010101010\ldots$$

$$011011100\ldots$$

Ces trois suites ont la même probabilité d'occurrence $1/2^n$. Pourtant, au moins pour les deux premières, elles ne présentent aucun caractère aléatoire. La première suite ne respecte pas la loi des grands nombres : si $(X_i)_{i\geq 1}$ est une suite iid de variables de Bernoulli de paramètre $1/2$, presque toute réalisation de cette suite doit vérifier

$$\frac{\sum_1^n X_i(\omega)}{n} \to 1/2$$

La seconde suite vérifie cette loi mais pas la loi des grands nombres appliquée à la suite $Y_i = 1_{(X_i=1,X_{i+1}=1)}$ qui est aussi une suite iid de Bernoulli, de paramètre $1/4$. Cela motive un premier critère dit de k-uniformité.

Définition 1 *Une suite $x = (x_n) \in \{0,1\}^{\mathbb{N}}$ est dite k-uniforme si pour tout $(\epsilon_1, \epsilon_2, \ldots \epsilon_k) \in \{0,1\}^k$ on a*

$$\frac{\sum_1^n 1_{(X_{1+(i-1)k}=\epsilon_1,\ldots,X_{ik}=\epsilon_k)}}{n} \to 1/2^k$$

Selon ce critère, on peut prouver que la troisième suite (dite de Champernowne) est ∞-uniforme, pourtant elle n'a rien d'aléatoire, c'est la simple énumération des entiers écrits en base 2.

1.1.2 Un autre point de vue issu de la théorie de la complexité.

l'information nécessaire à générer n éléments de la suite : Pour la suite 1 il suffit de 0 bit Pour la suite 2 il suffit de 1 bit (parité de n) Pour la suite 3 il suffit de $\log_2 n$ bits (on énumère les chiffres en base 2). Cela motive *une* définition de suite de nombres aléatoires : Pour générer n termes d'une suite aléatoire, il faut de l'ordre de n bits. Les suites précédentes ne sont donc pas assez complexes pour être aléatoires. Mais cette *définition* n'est pas très pratique.

1.1.3 Le point de vue statistique.

Définition 2 *On appelle générateur uniforme pseudo-aléatoire une algorithme produisant une suite $u_1, u_2, \ldots u_n$ à valeurs dans $[0,1]$ fournissant les mêmes réponses qu'un échantillon uniforme (suite iid de loi $\mathcal{U}(0,1)$ à une batterie de tests statistiques usuels.*

Cette définition de nature statistique laisse vraiment à désirer du point de vue théorique : une suite pseudo-aléatoire l'est donc jusqu'à preuve du contraire... Parmi les tests de conformité usuels citons les Test du χ^2, de Kolmogorov-Smirnov, tests empiriques de Knuth et enfin batterie de tests Die Hard (sic !) de G. Marsaglia `http ://www.stat.fsu.edu/pub/diehard/`.

1.1.4 Générateurs de nombres pseudo-aléatoires.

Les instructions permettant de générer des suites pseudo-aléatoire sont `rand` et `grand` en `scilab`.[1]. Dans ce qui suit nous donnons les principes de fonctionnement de ces générateurs.
Pour des problèmes liés à la précision finie des ordinateurs, on préfère générer des suites d'entiers aléatoires. Les générateurs congruentiels sont basés sur l'algorithme de Lehmer :

$$x_{n+1} \equiv ax_n + b \mod M$$

Le module M est pris très grand. Par exemple, pour le générateur de base `rand`, dû à Knuth $M = 2^{31}$. La valeur initiale x_0 définit complètement la suite. On l'appelle la racine, ou la graine `'seed'`. Pensez en particulier à changer la racine, si vous voulez obtenir une suite aléatoire différente à chaque utilisation.

Remarque. Dans le cas particulier où $b = 0$, il faut choisir a de sorte que la période de la suite géométrique $a^n \cdot x_0$ soit la plus grande possible dans $\mathbb{Z}/M\mathbb{Z}$. D'après votre cours d'arithmétique, les inversibles de $\mathbb{Z}/M\mathbb{Z}$ forment un groupe multiplicatif d'ordre $\varphi(M)$, où φ est la fonction caractéristique d'Euler. Il suffit de prendre a un élément d'ordre maximum dans ce groupe (élément primitif) et x_0 impair.
Exercice. montrer que $a \equiv 3$ ou$5 \mod 8$ et que la période est alors 2^{m-2}. □

Remarque. La suite des points (x_n, x_{n+1}) est située sur le graphe de la droite $y = ax + b$ dans $\mathbb{Z}/M\mathbb{Z}$. Il faut choisir a et b de sorte que la droite remplissent le mieux possible le tore $[0, M] \times [0, M]$. L'idéal serait de prendre a irrationnel

[1] Voir également la démo `random`

mais c'est impossible sur une machine à précision finie. Il faut choisir a et b très judicieusement. □

Remarque. Tous les générateurs aléatoires basé sur un algorithme itératif $x_{n+1} = g(x_n)$ sur machine à précision finie sont **périodiques** ! □

On obtient ensuite un générateur de loi $\mathcal{U}(0,1)$ en divisant par M.

Remarque. Le générateur standard `rand` de Knuth ne passe pas tous les test de Die Hard, il ne faut pas l'utiliser pour des simulations sérieuses. □

De nombreux raffinements ont été introduits pour allonger la période des suites pseudo-aléatoires, citons en particulier le générateur KISS de George Marsaglia de période 2^{123} et dont l'état est caractérisé par 4 entiers, basé sur l'algorithme de Lehmer, avec $a = 69069, b = 23606797, M = 2^{32}$ et des déplacements de registres (décalages à droite ou à gauche de la suite) bien choisis. Nous donnons le code source en langage C, surtout pour mystifier le lecteur.
Exemple. Générateur KISS[2] de Marsaglia de période 2^{123}.

```
/*_________________Kiss, The generator___________________________*/

double kiss(i,j,k)
        unsigned long int *i,*j,*k;

/* Generator proposed by Marsaglia and Zaman, 1993. See
   Robert and Casella (1999, pages 41-43) for details.
   Watch out: the last line
        x = ((double) (*i+*j+*k)*exp(-32*log(2.0)));
   must be calibrated, depending on the precision
   of the computer. */
{
  double x;

  *j = *j ^ (*j<<17);
  *k = (*k ^ (*k<<18)) & 0x7FFFFFFF;
  *i = 69069*(*i)+23606797;
  *j ^= (*j>>15);
  *k ^= (*k>>13);
  x = ((double) (*i+*j+*k)*exp(-32*log(2.0)));
  return(x);
}
```

Le générateur `grand` utilise par défaut mt : le Mersenne-Twister de M. Matsumoto and T. Nishimura, periode environ 2^{19937} , état donné par un tableau de 624 entiers (plus un indice pointant sur ce tableau). On peut lui demander explicitement d'utiliser le générateur KISS par la commande `grand('setgen',kiss)`.

Terminons en soulignant qu'il n'est pas facile de construire un générateur aléatoire. Par exemple J. Von Neumann a essayé sans succès le système du carré

[2] Keep It Simple, Stupid.

médian. Les suites basées sur la théorie du Chaos du type `x<-a*x*(1-x)` pour a bien choisi, bien qu'uniformément répartie en théorie, finissent par être piégées dans des cycles courts à cause de la précision finie des ordinateurs.
Dans la suite, on suppose que l'on dispose d'un générateur de nombre aléatoire fiable.

1.2 Lois non uniformes

1.2.1 Inversion de la fonction de répartition

On rappelle une définition de la fonction de répartition d'une v.a. X.

$$x \in \mathbb{R} \mapsto F_X(x) = \mathbb{P}(X \leq x) \in [0,1]$$

C'est une fonction croissante, continue à droite et limitée à gauche, telle que $\lim_{x\to-\infty} F_X(x) = 0$ et $\lim_{x\to+\infty} F_X(x) = 1$.

Définition 3 *Sot F une fonction de répartition sur $\mathbb{R}$. L'inverse généralisée ou fonction quantile est*

$$F^-(u) = \inf\{x;\ F(x) \geq u\}$$

Remarque. Lorsque F est bijective, on a évidemment $F^- = F^{-1}$. □

Le théorème fondamental de la simulation est

Théorème 1 *Soit U de loi $\mathcal{U}(0,1)$, soit F une fonction de répartition alors $F^-(U)$ est une v.a. de fonction de répartition F.*

Preuve. En effet, posons $X = F^-(U)$

$$\mathbb{P}(X \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

■

Remarque. Cette méthode appelée parfois anamorphose est très pratique, mais elle nécessite de connaître l'inverse de la fonction de répartition de la loi à simuler. Si on connaît la densité, mais pas son inverse on peut résoudre l'équation $F(u) = t$ de maniére approchée par une méthode numérique appropriée (Newton, sécante, dichotomie) mais cela ralentit considérablement la simulation. On peut aussi utiliser des tables statistiques et effectuer une interpolation liénaire, mais le stockage des tables occupe de la place mémoire. □

Exemple. génération d'une v.a. exponentielle de paramètre λ de fonction de répartition $F(t) = (1-e^{-\lambda t})1_{(t>0)}$, donc $X = -\ln(1-U)/\lambda$, mais comme $1-U$ a même loi que U, on obtient que $Y = -\ln(U)/\lambda$ suit une loi exponentielle de paramètre λ. Cette loi modélise les temps d'attente, les durées de vie (sans phénomène d'usure ou de mémoire).
Les calculs sont analogues pour la loi de Weibull, très utilisée en théorie de la fiabilité car prenant en compte le phénomène d'usure (fonction de survie $G(t) = P(X > t) = e^{-t^\alpha}1_{(t>0)}$. On inverse aisément $F = 1 - G$ et cela donne $X = (-\ln(U))^{1/\alpha}$ qui suit une loi de Weibull.
Exemple. Loi géométrique de paramètre p, qui modélise le nombre de lancers nécessaires pour amener face au jeu de pile ou face. On vérifie que $N = [X]$ où

X est une loi exponentielle de paramètre $-\ln(1-p)$.
Exemple. Loi de Pareto la densité est $f(x) = \frac{ab^a}{x^{a+1}} 1_{(x \geq b)}$. Cette loi est utile en economie, elle modélise la richesse des individus. On obtient $X = \frac{b}{U^{1/a}}$
Exemple. génération d'une v.a. de Cauchy La densité est $f(x) = \frac{\sigma}{\pi(x^2+\sigma^2)}$, $\sigma \tan(\pi(U-1/2))$ suit une loi de Cauchy de paramètre σ.
Exemple. génération *approximative* d'une v.a. normale.
On peut essayer la loi normale en utilisant l'approximation à 10^-8 près de

$$F^{-1}(u) = t - \frac{a_0 + a_1 t}{1 + b_1 t + b_2 t^2}$$

où $t = \ln(u^-2)$ et $a_0 = 2.30753,\ a_1 = 0.27061,\ b_1 = 0.99229,\ b_2 = 0.04481.$

La méthode d'inversion de la fonction de répartition s'applique aussi aux v.a. discrètes. Supposons qu'on veuille simuler une v.a. X à valeurs x_k, $k \in \mathbb{N}$ de loi donnée par les $p_k = P(X = x_k)$. La fonction de répartition est une fonction en escalier. $F(x) = p_1 + p_2 + ... + p_i$ si $x_i \leq x < x_{i+1}$. En notant les probabilités cumulées $F_i = p_1 + p_2 + ... + p_i$, l'algorithme suivant simule $X = F^-(U)$.

$X \leftarrow$ `rand`
Chercher i tel que $F_{i-1} < X \leq F_i$
$X \leftarrow x_i$

Remarque. Si le nombre d'éventualités est petit, on peut choisir de stocker les valeurs F_i pour ne pas les recalculer à chaque fois. On a aussi intérêt à minimiser le nombre de tests pour trouver i tel que $F_{i-1} < X \leq F_i$. □

Exercice. Ecrire une fonction `scilab` réalisant l'algorithme précédent. Si F est telle qu'il existe un k tel que $F_k \approx 1$, on peut choisir de stocker les k premières valeurs de F_i et de ne calculer les autres que si besoin est, ce qui arrivera si `rand` $> F_k$ i.e. avec une probabilité très faible.
Exercice. génération d'une v.a. de Poisson

Remarque. Pour la loi de Poisson, on peut aussi utiliser le fait que cette loi est liée à la loi exponentielle par le processus de Poisson. Soit $(N_t)_t$ un processus de Poisson de paramètre λ, on a

$$\mathbb{P}(N_t = k) = \mathbb{P}(X_1 + X_2 + \ldots X_k \leq t < X_1 + X_2 + \ldots X_{k+1})$$

N_t modélise le nombre de clients servis à l'instant t dans une file d'attente, et X_j représente le temps d'attente du j^e client, suivant la loi exponentielle de paramètre λ. On sait que N_t suit une loi de Poisson de paramètre λt. Il suffit donc de simuler N_1. L'algorithme suivant réalise cela en `scilab`.

```
function = [Y] = simpois1(alpha)
//
//simule une v.a. de loi de Poisson de paramètre alpha
//
rand('uniform');
a = exp(-alpha);
k=0; // initialisation
M = rand(1,1);
while (M>=a) do
      M = M*rand(1,1); k = k+1;
```

```
end
Y = k;
endfunction
```

□

1.2.2 Algorithme de rejet

C'est une méthode extrêmement puissante et générale, qui ne nécessite pas l'inversion de la fonction de répartition, ni même la connaissance de celle-ci. Il faut seulement connaître la densité de la loi à simuler. Soit μ une mesure positive sur $\mathbb{R}$. Soient g une densité de probabilité par rapport à μ. Supposons qu'on sache générer une loi de densité g, et qu'on désire simuler une loi de densité f, telle que

$$f(x) \leq Mg(x)$$

. L'algorithme d'acceptation rejet est le suivant :

```
1. Générer X de loi g, U de loi uniforme sur (0,1)
2. Accepter Y = X si M*g(X)*U  est inférieur ou égal à  f(X)
   Retourner en 1 sinon.
```

Propriétés 1 *La procédure se termine p.s. et fournit une variable Y de densité f.*

Preuve. Prouvons d'abord que le nombre de passage N dans la boucle est fini p.s. La v.a. N suit une loi géométrique (loi du premier succès dans une suite de Bernoulli) :

$$P(N = k) = (1 - p)^{k-1}p$$

où le paramètre de succès vaut $p = 1/M$. En effet

$$p = \mathbb{P}(U \leq f(X)/Mg(X)) = \int_{-\infty}^{\infty} \int_0^{f(x)/Mg(x)} g(x)\, dudx = \int_{-\infty}^{\infty} f(x)/M\, dx = 1/M$$

Donc $\mathbb{E}N = 1/p = M < \infty$ et N est fini p.s. Prouvons maintenant que Y a la densité f. Il suffit de remarquer que par construction

$$\mathcal{L}_Y = \mathcal{L}_{X|\, U \leq f(X)/Mg(X)}$$

Calculons la fonction de répartition de Y.

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X \leq y | U \leq \frac{f(X)}{Mg(X)}) = \frac{\mathbb{P}(X \leq y, U \leq \frac{f(X)}{Mg(X)})}{\mathbb{P}(U \leq \frac{f(X)}{Mg(X)})}$$

Calculons les intégrales

$$\mathbb{P}(Y \leq y) = \frac{\int_{-\infty}^{y} \int_0^{f(x)/Mg(x)} g(x)\, dudx}{\int_{-\infty}^{\infty} \int_0^{f(x)/Mg(x)} g(x)\, dudx} = \frac{\int_{-\infty}^{y} f(x)/M\, dx}{\int_{-\infty}^{\infty} f(x)/M\, dx}$$

$$\mathbb{P}(Y \leq y) = \int_{-\infty}^{y} f(x)\, dx.$$

■

Remarque. On voit dans ce qui précède que le nombre moyen de passage dans la boucle vaut M et sera d'autant plus petit que la majoration $f(x) \leq Mg(x)$ sera précise. □

Donnons maintenant quelques exemples.
Exemple. Génération d'une loi normale à partir d'une loi de Cauchy. Dans ce cas $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ et $g(x) = \frac{1}{\pi(1+x^2)}$. Le rapport $\frac{f(x)}{g(x)} = \sqrt{\frac{\pi}{2}}(1+x^2)e^{-x^2/2}$ est majoré par $M = \sqrt{2\pi/e}$. Le paramètre de succès i.e. la probabilité d'acceptation est $1/M \approx 0.66$, ce qui implique un rejet pour trois simulations environ.
Exemple. Dans le cas pariculier où $\alpha \in \mathbb{N}$, la loi Gamma $\mathcal{G}(\alpha, \beta)$ peut se répresenter comme la loi de la somme de α v.a. iid de loi exponentielle de paramètre β. On peut alors chercher à simuler la loi $\mathcal{G}(\alpha, \beta)$ avec $\alpha \notin \mathbb{N}$ à partir de la loi $\mathcal{G}(a, b)$, où $a = [\alpha]$ désigne la partie entière de α. Sans perte de généralité, supposons $\beta = 1$, le rapport $\frac{f}{g_b} = b^{-a} x^{\alpha - a} e^{-(1-b)x}$ est borné pour $b < 1$ par

$$M = b^{-a} \left(\frac{\alpha - a}{(1-b)e} \right)^{\alpha - a}$$

On en déduit que le choix optimal est $b = a/\alpha$ qui donne d'ailleurs la même espérance pour $\mathcal{G}(\alpha, \beta)$ et $\mathcal{G}(a, b)$.
Exercice. (Variante). On peut aussi simuler une v.a. de loi gamma à partir d'une loi de Weibull facile à simuler par inversion.

$$f(t) = \frac{1}{\Gamma(\alpha)} t^{\alpha-1} e^{-t} 1_{(t>0)}$$

$$g(t) = \alpha t^{\alpha-1} e^{-t^\alpha} 1_{(t>0)}$$

Remarque. La méthode de rejet est souvent utilisée pour simuler des lois à support compact. Considérons $[0, 1]$ par exemple. On prendra alors $\mu(dx) = 1_{(0,1)}(x)dx$, $g = 1$ et $M = \sup_{x \in [0,1]} f(x)$. Dans ce cadre, la méthode de rejet consiste à choisir un point (x, y) de façon uniforme sur $[0, 1] \times [0, M]$ et à regarder s'il tombe dans l'aire sous la courbe de f. Si oui, l'algorithme sort la valeur de l'abscisse correpondante. Si non, le point est rejeté et on refait l'expérience.

En particulier si μ est la mesure de comptage sur $\mathbb{N}$, l'algorithme permet de simuler des variables discrètes. Par rapport à la méthode d'inversion, cela ne requiert pas le calcul ou le stockage de la fonction de répartition F. On l'utilisera pour des lois discrètes ayant un grand nombre d'issues et où les issues ont des probabilités comparables, ce qui exclut la possibilité de trouver un k tel que $F_k \approx 1$.

□

1.3 Lois normales

1.3.1 Box-Muller

Proposition 1 *Soit (U, V) iid de loi $\mathcal{U}(0, 1)$. Alors*

$$(X, Y) = (\sqrt{-2 \ln U} \cos 2\pi V, \sqrt{-2 \ln U} \sin 2\pi V)$$

sont iid de loi $\mathcal{N}(0,1)$

Preuve. On peut donner une preuve bourrine en calculant. Voici une preuve plus élégante et peut-être moins rigoureuse.
Faisons la preuve à l'envers. Soit (X,Y) iid de loi $\mathcal{N}(0,1)$. En effet soit (R,Θ) les coordonnées polaires de (X,Y). Posons $U = \exp -R^2/2$ et $V = \frac{\Theta}{2\pi}$
$R^2 = X^2 + Y^2$ suit une loi du $\chi^2(2)$ qui est aussi une loi exponentielle de paramètre $1/2$ donc est la transformée par $-2\ln$ d'une loi uniforme, autrement dit $R^2 = -2\ln U$. De plus, comme la loi de (X,Y) est invariante par rotation, Θ suit une loi uniforme sur $[0,2\pi]$ donc $\Theta = 2\pi V$ avec V uniforme (et independante de U). ■

Remarque. On simule évidemment $\mathcal{N}(\mu,\sigma^2)$ par $\mu + \sigma X$ où X est $\mathcal{N}(0,1)$. □

1.3.2 Vecteurs gaussiens

Pour générer un vecteur gaussien $X = (X_1,\ldots,X_n)$ de matrice de covariance Q et d'espérance μ, on utilise le résultat d'algèbre linéaire suivant. Q est symétrique définie positive donc on peut diagonaliser Q dans une base orthonormée $Q = ASA^T$ où A est une matrice orthogonale et S une matrice diagonale à terme diagonaux $\sigma_j^2 > 0$. On simule une suite de v.a. Y_j indépendante de loi $\mathcal{N}(0,\sigma_j^2)$
Alors $Z = \mu + AY$ a même loi que X.
Vérifions cela. Par linéarité, Y étant centré, $\mathbb{E}Z = \mu$. Pour la covariance

$$\mathbb{E}(AY)_i(AY)_j = \mathbb{E}(\sum_k a_{ik}X_k)(\sum_l a_{jl}X_l)$$

$$= \mathbb{E}\sum_{k,l} a_{ik}a_{jl}X_kX_l = \sum_{k,l} a_{ik}a_{jl}\mathbb{E}X_kX_l$$

$$= \sum_k a_{ik}a_{jk}\mathbb{E}X_k^2 = \sum_k a_{ik}a_{jk}\sigma_k^2$$

$$\mathbb{E}(AY)_i(AY)_j = (ASA^T)_{i,j}$$

donc la matrice de covariance de Z est bien ASA^T.
Pour réaliser la factorisation de Q, `scilab` fournit la fonction `svd` (singular value décomposition). L'aide de `scilab` précise

```
[U,S,V] = svd(X)  produces a diagonal matrix  S,
 of the same dimension as  X  and with nonnegative diagonal elements
in decreasing order,
and unitary matrices  U  and  V  so that  X = U*S*V' .
```

Lorsque Q est symétrique, `[A,S,B] = svd(Q)` fournit $A = B$ orthogonales telles que

$$Q = ASA^T$$

1.4 Méthode de Monte Carlo

Le terme générique "méthode de Montecarlo" recouvre les algorithmes de calcul de quantité déterministes (i.e. non aléatoire) par simulation d'échantillon pseudo-aléatoire. Le principe repose sur la loi des grands nombres.
Donnons l'exemple de calcul d'intégrales par cette méthode.

$$\int_a^b f(x)g(x)\,dx$$

Lorsque $g(x)$ est la densité d'une loi connue, l'intégrale n'est autre que

$$\mathbb{E}f(X)$$

où X est une v.a. de loi de densité g. Par la loi forte des grands nombres, si $(X_1, X_2, \dots X_n)$ désigne un échantillon de loi de densité g,

$$\frac{f(X_1) + f(X_2) + \dots f(X_n)}{n} \to \mu = \mathbb{E}f(X)$$

On approche donc

$$\int_a^b f(x)g(x)\,dx \approx \frac{f(X_1) + f(X_2) + \dots f(X_n)}{n}.$$

On simule donc un échantillon de taille *très grande* et on calcule une moyenne arithmétique. Le programme `scilab` suivant réalise ceci pour le calcul (!) de $\int_0^1 x\,dx = 1/2$.

```
function  [y] = Montecarlo(n)
//simulation d'un n-echantillon et
//calcul de la moyenne arithmétique
y=zeros(1,n);z=zeros(1,n); //initialisation
x = grand(1,n,'unf',0,1);//tirage aléatoire uniforme
for i=1:n
y(i) = mean(x(1:i));//calcul de la moyenne
end
clf();
plot(y);
xtitle("Montecarlo simulation");
endfunction
```

(Voir figure 1.1.)
Pour $n = 10000$, on trouve 0.5030378, soit deux décimales.

Remarque. La convergence de ces méthode est lente, la vitesse de convergence de la loi des grands nombres est en $O(1/\sqrt{n}$, à cause du théorème central limite.En effet

$$\frac{f(X_1) + f(X_2) + \dots f(X_n)}{n} = \mu + \frac{\sigma}{\sqrt{n}} Z_n$$

où Z_n converge en loi vers la loi $\mathcal{N}(0,1)$ (et là la convergence est rapide, en pratique $n \geq 30$). On a donc l'intervalle de confiance au niveau α

$$\mu \in [\frac{\sum_k f(X_k)}{n} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \frac{\sum_k f(X_k)}{n} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}]$$

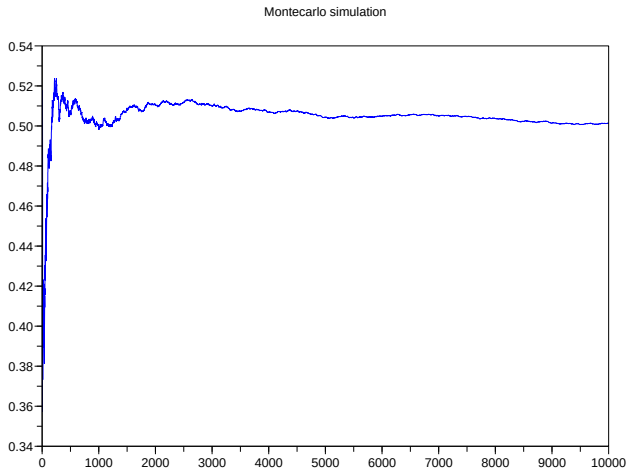

FIG. 1.1 – La convergence de la loi des grands nombres

où $z_{1-\alpha/2}$ est le quantile normal bilatéral correspondant. Cependant cette vitesse est la même *quelle que soit la dimension d'espace*. Cela rend la méthode de MonteCarlo compétitive pour le calcul d'intégrales *multiples en grande dimension* $d \geq 4$. En effet en analyse numérique on montre que la méthode de quadrature de Simpson appliquée à un intervalle de longueur h donne une erreur en $O(h^5)$. Pour un domaine $[a,b]^d$, si on subdivise $[a,b]$ en prenant un pas $h := (b-a)/n$, on découpera le pavé $[a,b]^d$ en $N := n^d$ petits pavés et l'erreur totale sera de l'ordre de $n^d \times \frac{1}{n^5}$ donc en $O(h^{5-d}) = O(\frac{1}{n^{5-d}}) = 0(\frac{1}{N^{5/d-1}})$. Pour $d = 3$, cela donne $0(\frac{1}{N^{2/3}})$, pour $d = 4$, cela fait $0(\frac{1}{N^{1/4}})$, la méthode de Monte Carlo est meilleure. A partir de $d = 5$ la méthode de MonteCarlo est la seule méthode utilisable. □

D'autre part, les formules d'erreur dans les quadratures numériques exigent une régularité élevée de la fonction (f de classe $\mathcal{C}^4$ (resp. $\mathcal{C}^2$) pour Simpson (resp trapèzes) alors que la méthode de MonteCarlo ne demande *aucune régularité de l'intégrand* (mesurable borné suffit).

Le programme `scilab` suivant trace l'évolution de l'erreur en coordonnées log log. Dans cette echelle, les erreurs pour des echantillons de taille croissante sont approximativement situés sur une droite de pense $-1/2$. Cela confirme l'évolution en $O(1/\sqrt{n})$.

```
function  [y] = Montecarlolog(n)
//simulation d'un n-echantillon et calcul de la moyenne arithmétique
//comparaison avec l'esperance
//illustration graphique de la convergence en 1/sqrt(n)
y=zeros(1,n);z=zeros(1,n); //initialisation
x = grand(1,n,'unf',0,1);//tirage aléatoire
for i=1:n
y(i) = mean(x(1:i));//calcul de la moyenne
z(i) = abs(y(i) - 0.5);//calcul de l'erreur absolue
```

```
end
clf();
t=1:n;
plot2d(t,z,logflag="ll");//trace log log
xtitle("Montecarlo simulation log log");
endfunction
```

(Voir figure 1.2.)

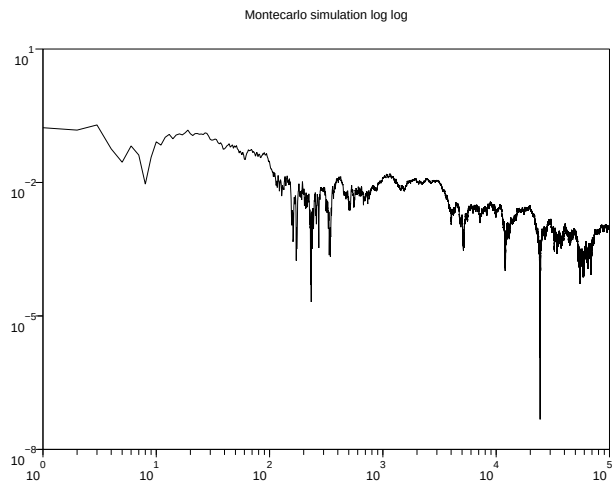

FIG. 1.2 – La convergence de la loi des grands nombres en coordonnées log log

1.5 Programmes scilab.

Les programmes qui suivent se trouvent dans la distribution `scilab` dans le répertoire `SCILAB/demos/random/`, où `SCILAB` désigne le répertoire où se trouve le logiciel. Voici le fichier `random.sci` qui permet de simuler certaines lois classiques.

```
function [y]=RndInt(m,n,imin,imax)
// discrete uniform random number
//-------------------------------
// Copyright ENPC
  y=rand(m,n,'uniform')
  y=int(floor(y*(imax+1-imin)+ imin ));
endfunction


function [y]=RndDisc(m,n,x,p)
// discrete law random number
// sum p_i delta_{x_i}
//-------------------------------
```

```
  p1=[0,p];p1=cumsum(p1);
  y=rand(m,n,'uniform')
  N=prod(size(x));
  res=0*ones(m*n);
  for i=1:N,z=0*ones(m*n,1),id=find( p1(i) <= y & y < p1(i+1) ),
    z(id)=x(i)*ones(prod(size(id))),res=res+z;
  end
  y=matrix(res,m,n);
endfunction

function [y]=Binomial(m,n,pb,nb)
// Binomial law (p,N)
// P{X=n} = C_N^n p^n (1-p)^(N-n)
//-----------------------------------
  res=[];
  // we use blocks of size 100 to avoid overflows
  ntir=100;
  ntirc=ntir;
  y=rand(ntir,nb,'uniform');
  indy= find( y < pb);
  y=0*ones(y);
  y(indy)=1;
  y=sum(y,'c')
  res=[res;y];
  while ( ntirc < m*n )
    y=rand(ntir,nb,'uniform');
    indy= find(y< pb);
    y=0*ones(y);
    y(indy)=1;
    y=sum(y,'c')
    res=[res;y];
    ntirc=ntirc+ntir;
  end
  y=matrix(res(1:m*n),m,n);
endfunction

function [y]=Geom(m,n,p)
// P(0)= 0 P(i) = p*(1-p)^{n-1} P(inf)=0
// E = 1/p ; sig2= (1-p)/p^2
//---------------------------------------
  if p >= 1 then write(%io(2),'p must be < 1');end
  y=0*ones(m,n)
  for i=1:m*n,
    samples=1
    z=rand(1,1,'uniform');
    while( z < 1-p) ,z=rand(1,1,'uniform'); samples=samples+1;end
    y(i)= samples;
  end
  y=matrix(y,m,n)
endfunction
```

```
function [y]=Poisson(m,n,pmean)
// P{n} = exp(-lambda)lambda^n/n!
// pmean =lambda
//----------------------------
  y=0*ones(m,n)
  bound= exp(-pmean);
  for i=1:m*n,
    count=0
    lprod=1
    while( lprod >= bound), lprod=lprod*rand(1,1,'uniform');
      count=count+1;end
      y(i)=count-1;
  end
  y=matrix(y,m,n)
endfunction

function [y]=Exp(m,n,lambda)
// lambda exp(-lambda x) x>=0
// ---------------------------
  y=(-1/lambda)* log(rand(m,n,'uniform'));
endfunction

function [y]=Weibull(m,n,Alpha,Beta)
//-------------------------------
  y=rand(m,n,'uniform')
  y= (Beta*( - log(1-y)))^(1/Alpha)
endfunction

function [y]=HyperGeom(m,n,Mean,var)
//-------------------------------
  z = var / (Mean * Mean);
  pP = 0.5 * (1.0 - sqrt((z - 1.0) / ( z + 1.0 )));
  y=rand(m,n,'uniform')
  zz=find( y > pP) ;
  y=pP*ones(y);
  y(zz) = (1-pP)*ones(zz);
  y1=rand(m,n,'uniform')
  y=-Mean * log(y1) ./ (2.0 * y) ;
endfunction

function [y]=Erlang(m,n,pMean,pVariance)
//-------------------------------
  k = int( (pMean * pMean ) / pVariance + 0.5 );
  if (k <= 0) then k=1;end
  a = k / pMean;
  // we use blocks of size 100 to avoid overflows
  res=[];
  ntir=100;
```

```
  ntirc=ntir;
  y=rand(ntir,k,'uniform');
  y= -log(prod(y,'r'))/a;
  res=[res;y];
  while ( ntirc < m*n )
    y=rand(ntir,k,'uniform');
    y= -log(prod(y,'r'))/a;
    res=[res;y];
    ntirc=ntirc+ntir;
  end
  y=matrix(res(1:m*n),m,n);
endfunction

function [y]=RndPerm(n)
//-------------------------------
// a uniform random permutation of (1:n)
  y=rand(1,n,'uniform')
  [us,z]=sort(y);
endfunction
```

Les programmes précédents sont utilisés dans les programmes de tracé d'histogramme d'échantillon et des densités correspondantes qui suivent. Ils sont contenus dans le fichier `randomT.sci`. Par exemple, l'appel de la fonction `WeibullT(10000)`, après avoir au préalable chargé les fichiers `SCILAB/demos/random/random.sci` et `SCILAB/demos/random/randomT.sci`, provoque l'affichage de la figure suivante :

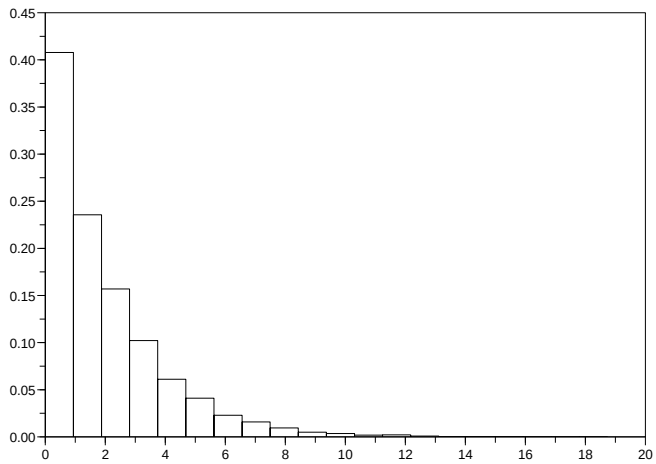

FIG. 1.3 – Histogramme d'une loi de Weibull simulée par anamorphose

```
function [z]=RndIntT(n)
//------------------------------- OK
  [lhs,rhs]=argn(0)
  if rhs <= 0 ; n=10000;end
  imin=-10;
  imax=10;
  y=RndInt(1,n,-10,10);
  i=imin-2:imax+2;
  z=[];
  for i1=i, z=[z,prod(size(find(y==i1)))],end
  plot2d3("onn",i',z'/n,[1,2],"151","Simulation ",[-12,0,12,0.1]);
  i1=(imin:imax)';
  plot2d1("onn",i1,ones(i1)/prod(size(imin:imax)),[-2,5],"100","Theory");
endfunction
function [z]=RndDiscT(n)
//------------------------------- OK
  [lhs,rhs]=argn(0)
  if rhs <= 0 ; n=10000;end
  x=[1,3,4,6,10,12];
  pr=[0.1,0.2,0.3,0.2,0.1,0.1];
  y=RndDisc(1,n,x,pr);
  i=0:13
  z=[];
  for i1=i, z=[z,prod(size(find(y==i1)))],end
  plot2d3("onn",i',z'/n,[1,3],"151","Simulation ",[0,0,14,0.5]);
  plot2d1("onn",x',pr',[-2,6],"100","Theory");
endfunction
function [zt]=BinomialT(n)
//------------------------------- OK
  [lhs,rhs]=argn(0)
  if rhs <= 0 ; n=10000;end
  prb=0.5;
  N=10;
  y=Binomial(1,n,prb,N);
  i=0:10;
  z=[];
  for i1=i, z=[z,prod(size(find(y==i1)))],end
  plot2d3("onn",i',z'/n,[1,3],"161","Simulation");
  deff('[y]=fact(n)','y=prod(1:n)');
  deff('[z]=C(N,n)','z= fact(N)/(fact(n)*fact(N-n))');
  i=0:N;
  zt=[];
  for j=i, zt=[zt, C(N,j)*prb^j*(1-prb)^(N-j)];end
  plot2d1("onn",i',zt',[-2,6],"100","Theory");
endfunction

function []=GeomT(n)
//------------------------------- OK
  [lhs,rhs]=argn(0)
  if rhs <= 0 ; n=10000;end
```

```
  pr=0.2
  y=Geom(1,n,pr)
  N=20
  i=0:N;
  z=[];
  for i1=i, z=[z,prod(size(find(y==i1)))],end
  plot2d3("onn",i',z'/n,[1,3],"161","Simulation");
  zt=[0];for i1=1:N; zt=[zt,pr*(1-pr)^(i1-1)];end
  plot2d1("onn",i',zt',[-2,6],"100","Theory");
endfunction
function [z]=PoissonT(n)
//------------------------------- OK
  [lhs,rhs]=argn(0)
  if rhs <= 0 ; n=1000;end
  pmean=3;
  y=Poisson(1,n,pmean);
  N=20;
  i=0:N;
  z=[];
  for i1=i, z=[z,prod(size(find(y==i1)))],end
  plot2d3("onn",i',z'/n,1,"161");
  deff('[y]=fact(n)','if n==0 then y=1;else y=n*fact(n-1);end');
  zt=[];for i1=0:N; zt=[zt, exp(-pmean) *pmean^i1/fact(i1)];end
  plot2d1("onn",i',zt',[-2,6],"100","Theory");
endfunction
function []=ExpT(n)
//------------------------------- OK
// lambda exp(-lambda x) x>=0
// ----------------------------
  [lhs,rhs]=argn(0)
  if rhs <= 0 ; n=1000;end
  lambda=3;
  y=Exp(1,n,lambda);
  histplot([0:0.1:10],y,[1,1],'051',' ',[0,0,5,3]);
  deff('[y]=f(x)','y=lambda*exp(-lambda*x);');
  x=[0:0.1:10]';plot2d(x,f(x),1,"000");
  titre= 'macro histplot : Histogram plot';
  xtitle(titre,'Classes','N(C)/Nmax');
endfunction
function []=WeibullT(n)
//-------------------------------
  [lhs,rhs]=argn(0)
  if rhs <= 0 ; n=10000;end
  y=Weibull(1,n,1,2)
  histplot(20,y,[1,1],'061');
endfunction

function []=HyperGeomT(n)
//-------------------------------
  [lhs,rhs]=argn(0)
```

```
  if rhs <= 0 ; n=10000;end
  y=HyperGeom(1,n,1,10)
  histplot([0:0.25:10],y,[1,1],'061',' ',[0,0,10,0.4]);
endfunction

function [y]=ErlangT(n)
//-------------------------------
  [lhs,rhs]=argn(0)
  if rhs <= 0 ; n=10000;end
  y=Erlang(1,n,10,1)
  histplot(20,y,[1,1],'061');
endfunction
```

Chapitre 2

Chaînes de Markov.

Référence du chapitre : livre de Norris [6]. Ce chapitre très bref présente un résumé des propriétés essentielles des CDM, illustrée par quelques exemples et simulations. On renvoie à la bibliographie [6],[7] pour les preuves et compléments. Des exercices corrigés sont présentés dans [16]. la rubrique Chaîne de Markov dans wikipedia est bien faite. Des exercices en lignes interactifs (avec WIMS) sont sur le site `http ://wims.unice.fr/wims/fr_U2~proba~oefmarkov.fr.html`

2.1 Matrice de transition

Soit $\Omega, \mathcal{F}, \mathbb{P}$ un espace probabilisé. Soit E un ensemble discret (fini ou dénombrable). On considère une suite de v.a. $(X_n)_n$ à valeurs dans E, autrement dit un processus aléatoire (à temps discret).

Définition 4 *le processus aléatoire $(X_n)_{n\in\mathbb{N}}$ est une CDM si*

$$\mathbb{P}(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = j | X_n = i_n)$$

pour tout $n \in \mathbb{N}$ et pour tous $i_0, i_1, \ldots i_n, j \in E$ tels que $\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) > 0$. Si de plus $\mathbb{P}(X_{n+1} = j | X_n = i)$ ne dépend pas de n, on dit que la CDM est homogène. On note alors

$$\mathbb{P}(X_{n+1} = j | X_n = i) = p_{i,j}$$

et $P = (p_{i,j})$ s'appelle la matrice de transition de la CDM.

Remarque. La matrice P est une matrice stochastique : pour chaque *ligne* i,

$$\sum_k p_{i,k} = 1.$$

□

Une CDM est un processus aléatoire $(X_n)_n$ sans mémoire.
Connaissant la loi de X_0 appelée loi *initiale* de la chaîne et les probabilités de transition on peut calculer la loi du processus :

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \mathbb{P}(X_0 = i_0) p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n}$$

Proposition 2 *Equations de Chapman-Kolmogorov :*

$$\mathbb{P}(X_{n+1} = j) = \sum_i \mathbb{P}(X_n = i) \cdot p_{i,j}$$

En notant $p_{i,j}^{(n)} = \mathbb{P}(X_n = j | X_0 = i)$,

$$p_{i,j}^{(n+k)} = \sum_l p_{i,l}^{(n)} \cdot p_{l,j}^{(k)}.$$

dans le cas homogène la matrice de transition en n unités de temps d'obtient en élevant à la puissance n la matrice de transition

$$P^n = (p_{i,j}^{(n)}).$$

Dans la suite, on étudie seulement les CDM *homogènes.* On voit donc que les propriétés algébriques de la matrice de transition déterminent la comportement de la chaîne. Dans le cas des CDM a nombre fini d'états, le problème se ramène donc à de l'algèbre linéaire. En effet en calculant les valeurs propres, il est aisé de calculer les puissances de P.

Remarque. La matrice P étant stochastique, le vecteur colonne $(1, 1, \ldots, 1)^T$ est vecteur propre de la matrice pour la valeur propre 1. □

On peut en particulier étudier très facilement les CDM a petit nombre d'états par la réduction des matrices.
Exercice. Soit $0 < p, q < 1$ Etudier la CDM à deux états de matrice de transition

$$P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$$

Etudier le comportement asymptotique de P^n. Que remarquez vous ? réponse :

$$P^n \to \begin{pmatrix} p/(p+q) & q/(p+q) \\ p/(p+q) & q/(p+q) \end{pmatrix}$$

Cette convergence vers la loi stationnaire se nomme la propriété ergodique, voir section 2.6.
Les cas particuliers sont instructifs.
Si $p = q = 1$, $P = Id$ et la chaîne n'est pas *irréductible* (voir section 2.4) Il n'y a pas unicité de la loi stationnaire.
Si $p = q = 0$,

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

et la chaîne est*périodique* de période 2. Dans ce cas, il n'y a pas convergence $P^{2n} = I$ et $P^{2n+1} = P$. Voir section 2.6.

2.2 Propriété de Markov fort

Notons $\mathcal{A}_n = \sigma(X_0, X_1, \ldots X_n)$ La définition des CDM à l'aide de probabilités conditionnelles élémentaires permet d'obtenir les propriétés suivantes.

Propriétés 2 *Soit* $A \in \mathcal{A}_n$

$$\mathbb{P}(X_{n+1} = j_1, X_{n+2} = j_2, \ldots, X_{n+k} = j_k | A, X_n = i)$$

$$= \mathbb{P}(X_1 = j_1, X_2 = j_2, \ldots, X_k = j_k | X_0 = i) = p_{i,j_1} \cdot p_{j_1,j_2} \cdots p_{j_{k-1},j_k}$$

La propriété précédente se généralise aux temps d'arrêts aléatoire (voir chapitre suivant).

Propriétés 3 *Soit* T *un temps d'arrêt pour le processus* $(X_n)_n$ *et* $A \in \mathcal{A}_T$

$$\mathbb{P}(X_{T+1} = j_1, X_{T+2} = j_2, \ldots, X_{T+k} = j_k | T < \infty, A, X_T = i)$$

$$= \mathbb{P}(X_1 = j_1, X_2 = j_2, \ldots, X_k = j_k | X_0 = i) = p_{i,j_1} \cdot p_{j_1,j_2} \cdots p_{j_{k-1},j_k}$$

On peut formaliser davantage en introduisant l'opérateur de décalage θ_T défini par

$$\theta_T(X) = (X_T, X_{T+1}, \ldots)$$

qui représente la chaîne observée à partir de l'instant T.

Propriétés 4 *Soit* T *un temps d'arrêt pour le processus* $(X_n)_n$, $A \in \mathcal{A}_T$ *et* C *une partie mesurable de* $E^{\mathbb{N}}$.

$$\mathbb{P}(\theta_T(X) \in C | T < \infty, A, X_T = i) = \mathbb{P}(X \in C | X_0 = i)$$

Exercice. Etudier le problème de la *ruine du joueur* dans [6] pages 15 et 21.

2.3 Classification des états

Définition 5 *un état* i *est appelé* récurrent *si* $\mathbb{P}_i(T_i < \infty) = 1$ transitoire *sinon. Parmi les états récurrents, on distingue les états* récurrents positifs *(ou ergodiques) qui vérifient* $\mathbb{E}_i T_i < \infty$, *des états* récurrents nuls *pour lesquels* $\mathbb{E}_i T_i = \infty$

Propriétés 5 *un état* i *est* récurrent *(resp.* transitoire *) ssi* $\sum_n p_{i,i}^{(n)} = \infty$ *(resp.* $\sum_n p_{i,i}^{(n)} < \infty$*)*

Remarque. (Pour retenir ce résultat.) Soit $V_i = \sum_n 1_{(X_n = i)}$ le nombre de visites en i. Par Beppo-Levi, $\mathbb{E}_i V_i = \sum_n p_{i,i}^{(n)}$. Si un état est récurrent alors partant de i la chaîne y revient une infinité de fois, donc $\mathbb{E}_i V_i = \infty$. □

Exemple. Marche aléatoire symétrique dans $\mathbb{Z}^n$. Polya (1928) a montré les résultats suivants (à connaître). Pour $n = 1, 2$ la marche est récurrente. Pour $n \geq 3$ elle est transiente. Pour les démontrer, appliquer la propriété précédente, en estimant $p_{0,0}^{(2n)}$ par la formule de Stirling, [6] page 30.

2.4 Irréductibilité.

Définition 6 *un état* i *mène à un état* j *si* $\exists n \geq 0, p_{i,j}^{(n)} > 0$. *On note alors* $i \to j$. *Lorsque* $i \to j$ *et* $j \to i$, *on dit que* i *et* j *communiquent et on note* $i \sim j$. *La relation* $\sim$ *est une relation d'équivalence.*

Lorsque deux états communiquent, ils sont de même nature (récurrent positif, récurrent nul ou transient. Une chaîne est **irréductible** si tous ses états communiquent. Ainsi il n'y que a trois types de CDM *irréductibles*, transitoire,récurrente nulle, récurrente positive selon que $P(T_i < \infty) < 1$, $P(T_i < \infty) = 1$ et $E_iT_i = +\infty$, $P(T_i < \infty) = 1$ et $E_iT_i < \infty$.

Remarque. Dans le cas *fini*, seul le cas récurrent positif est possible. □

2.5 Mesure invariante

Définition 7 *On dit que μ est une mesure invariante (ou stationnaire) si*

$$\mu_j = \sum_i \mu_i p_{i,j}.$$

Cela se traduit aussi par

$$\mu = \mu \cdot P$$

où $\mu = (\mu_1, \ldots, \mu_N)$ est un vecteur ligne.

Le terme stationnaire s'explique par le fait que lorsque X_n suit la loi μ alors X_{n+1} aussi.

Remarque. une mesure invariante est définie à une constante multiplicative près. □

Dans le cas récurrent positif, on démontre qu'il existe une seule *loi* invariante-donnée par

$$\pi_i = \frac{1}{E_iT_i}$$

Exemple. L'exemple de la CDM de diffusion d'Ehrenfest. Un mur sépare deux pièces contenant chacune k (resp. $N-k$) molécules de gaz.A $t = 0$, on perce une ouverture dans le mur. A chaque unité de temps, une particule (et une seule) change de pièce. Donner la matrice de transition du processus. Etudier la chaîne (irréductibilité, loi invariante). Montrer que la loi stationnaire est donnée par

$$\pi_j = \binom{N}{j} \frac{1}{2^N}.$$

Interprétez. Feller [7] commente ce résultat, en particulier la propriété de récurrence positive, qui semble violer le principe statistique de convergence vers l'équiilibre thermodynamique (répartition moitié-moitié dans chaque pièce) : Les temps de retour E_iT_i d'une répartition $i, N-i$ sont d'autant plus longs que la répartition diffère de $N/2, N/2$ (utiliser la formule de Stirling, $n! \sim \sqrt{2\pi n}(n/e)^n$.

2.6 Convergence vers la loi stationnaire et propriété ergodique.

Au bout d'un temps long, une CDM irréductible apériodique converge en loi vers la loi invariante et ce quelle que soit la loi initiale.

Théorème 2 *Soit une CDM irréductible apériodique alors quel que soit l'état initial i, quand $n \to \infty$*

$$p_{i,j}^{(n)} \to \frac{1}{\mathbb{E}_j T_j}$$

Si de plus la CDM est récurrente positive (et c'est toujours le cas lorsque E est fini et la chaîne irréductible apériodique) on a

$$p_{i,j}^{(n)} \to \pi_j$$

Remarque. Dans le cas fini, on peut obtenir ces résultats grace au théorème de Perron Frobenius aisément. Ce théorème affirme que si P est une matrice stochastique irréductible apériodique (i.e. la matrice de transition d'une CDM irréductible apériodique), alors $\lambda_1 = 1$ est valeur propre simple, associée à un vecteur propre à coefficients strictement positifs et que toutes les autres valeurs propres sont de module *strictement inférieur* à 1. On voit alors aisément que

$$|p_{i,j}^{(n)} - \pi_j| \leq C^{te} \rho^n$$

où $\rho = |\lambda_2| < 1$ est la valeur propre différente de $\lambda_1 = 1$ qui est de module maximum. On observe donc une vitesse de convergence *géométrique.*

□

Remarque. Les hypothèses d'irréductibilité et d'apériodicité sont nécessaires. Pour des contre-exemples simples voir la chaîne à deux états, voir section 2.1.

□

Soit $V_i(n) = \sum_{0 \leq k \leq n-1} 1_{(X_k = i)}$ le nombre de visites en i avant l'instant n. La proportion du temps passé en i avant n est alors $\frac{V_i(n)}{n}$.

Théorème 3 *Soit une CDM irréductible. Quelle que soit la loi initiale,*

$$\frac{V_i(n)}{n} \to \frac{1}{E_i T_i} \quad p.s.$$

De plus dans le cas où la chaîne est récurrente positive, pour toute fonction bornée f

$$\frac{\sum_{0 \leq k \leq n-1} f(X_k)}{n} \to E_\pi f = \sum_i f(i) \pi_i$$

où π désigne l'unique loi stationnaire.

Preuve. La preuve s'appuie sur la loi forte des grands nombres appliquée à la suite des temps d'atteinte successifs de l'état i notés $T_i^{(k)}$. Voir Norris,[6] ■

2.7 Simulation de CDM et illustration en `Scilab`.

Bien que `Scilab` possède des primitives permettant de le faire, nous donnons l'algorithme pour simuler une CDM d'ensemble d'état fini E, de loi initiale $\lambda = (\lambda_1, \ldots, \lambda_n)$ et de matrice de transition P. Voir Norris [6] chap 5, pp 206. Comme $\sum_k \lambda_k = 1$, on peut découper $[0, 1]$ en sous-intervalles disjoints I_k de longueur λ_k :

$$I_k = (\lambda_1 + \ldots + \lambda_{k-1}, \lambda_1 + \ldots + \lambda_k).$$

De la même façon, pour chaque $i \in E$, on peut découper $[0,1]$ en sous-intervalles disjoints $A_{i,j}$ de longueur $p_{i,j}$. On définit alors les fonctions $G_0 : [0,1] \to E$ et $G : E \times [0,1] \to E$ par

$$G_0(u) = i \quad \text{si} \quad u \in I_i$$

$$G(i,u) = j \quad \text{si} \quad u \in A_{i,j}.$$

Supposons que $U_0, U_1, \ldots$ est un échantillon uniforme sur $(0,1)$ (simulé par un générateur aléatoire), posons $X_0 = G_0(U_0)$ $X_{n+1} = G(X_n, U_{n+1})$.
Le processus $(X_n)_n$ est une CDM de loi initiale $\lambda = (\lambda_1, \ldots, \lambda_n)$ et de matrice de transition P.
Preuve. Nous avons alors

$$\mathbb{P}(X_0 = i) = P(U_0 \in I_i) = \lambda_i,$$

$$\mathbb{P}(X_{n+1} = j | (X_n = i) = \mathbb{P}(U_{n+1} \in A_{i,j}) = p_{i,j}.$$

■

En `Scilab`, les fonctions `genmarkov`, `classmarkov` permettent de créer des matrices de transition ayant les propriétés souhaitées et de classifier les CDM.
Exemple. $P = genmarkov([3], 2)$; crée une matrice de transition avec 3 états récurrents et 2 états transitoires.
Enfin, on peut simuler des trajectoires avec la fonction `grand`.
Exemple. $Y = grand(10000,' markov', P, [1,2,3,4,5])$; simule 10 000 états successifs à partir de $X_0 = 1, X_0 = 2, \ldots, X_0 = 5$.
Pour voir la loi invariante, on peut soit calculer P^{100}, à cause de la convergence géométrique ça suffit largement, ou bien utiliser `eigenmarkov`.
Exemple. $[M, Q] = eigenmarkov(P)$; donne $M = (\pi_1, \ldots \pi_n)$ et $Q = (1, \ldots, 1)$.

Chapitre 3

Espérance conditionnelle

Dans ce chapitre $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace probabilisé. Ω est appelé l'ensemble fondamental, $\mathcal{A}$ est une tribu ou σ-algèbre de parties de Ω. Le plus souvent, la tribu sera complète, i.e. contenant toutes les parties négligeables.
Il est bon de s'imprégner du vocabulaire de la théorie, afin de soutenir et développer l'intuition. Ainsi un élément $\omega \in \Omega$ est appelé une issue, une épreuve, un tirage ... selon le contexte, un élément $A \in \mathcal{A}$ est appelé un évènement. Lorsque $\omega \in A$, on dit que l'évènement A est réalisé.
★ A méditer : L'enjeu de la théorie ne consiste pas à connaître l'issue $\omega \in \Omega$ [1] mais à calculer des probabilités d'évènements ou des lois de variables aléatoires.

3.1 L'espérance mathématique : une moyenne probabiliste

Soit X une variable aléatoire $\mathbb{P}$-intégrable, ou positive. On définit son espérance[2] mathématique par

$$\mathbb{E}X = \int_\Omega X(\omega)\, d\mathbb{P}(\omega)$$

Cette expression se calcule à l'aide de *la loi* $\mathbb{P}_X$ de la v.a. X, qui est simplement la mesure image de $\mathbb{P}$ par X :

$$\mathbb{E}X = \int_\mathbb{R} x\, d\mathbb{P}_X(x)$$

En fait cette formule générale recouvre deux cas particuliers de natures différentes, selon que la v.a. X est discrète prenant ses valeurs dans l'ensemble dénombrable $\{x_k\}$

$$\mathbb{E}X = \sum_k x_k \mathbb{P}(X = x_k)$$

ou absolument continue de densité f_X

$$\mathbb{E}X = \int_{-\infty}^{+\infty} x f_X(x)\, dx$$

[1] la lettre Ω représentant l'inconnaissable...
[2] de l'anglais *expectation* : prévision

★ A méditer : Du fait que $\sum_k \mathbb{P}(X = x_k) = 1$ ou $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$, l'espérance représente *la moyenne des valeurs prises par la v.a., pondérées par leurs probabilités.*

Remarque. Dans le cas absolument continu, il est bon de voir que

$$f_X(x)\,dx = \mathbb{P}(x \leq X < x + dx) = dF_X(x),$$

où $F_X(x) = \mathbb{P}(X < x)$ désigne la fonction de répartition de X. □

Il faut bien garder à l'esprit que l'espérance est un nombre associé à une v.a. La propriété principale est

Proposition 3 *L'espérance est une forme linéaire positive sur $L^1(\Omega, \mathcal{A}, \mathbb{P})$*

Exercice. Retrouver l'espérance d'une loi binomiale $\mathcal{B}(n, p)$, d'une loi de Poisson $\mathcal{P}(\lambda)$ et d'une loi exponentielle de paramètre λ.

3.2 Probabilité conditionnelle

Soit un évènement $B \in \mathcal{A}$, tel que $\mathbb{P}(B) > 0$. Pour tout $A \in \mathcal{A}$, la probabilité de A sachant B est définie par

$$\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$$

★ A méditer : Ainsi $\mathbb{P}_B(\cdot) = \mathbb{P}(\cdot|B)$ induit une *nouvelle* probabilité pour laquelle B joue le même rôle que l'ensemble fondamental Ω vis à vis de $\mathbb{P}$: $\mathbb{P}_B(B) = 1$, $\mathbb{P}_B(A) = 0$ ssi $A \cap B = \emptyset$

Lorsqu'il y a différentes éventualités qui s'excluent mutuellement, on utilise :

Théorème 4 *la formule des probabilités totales. Soit $(B_j)j$ une partition de Ω,*

$$\mathbb{P}(A) = \sum_j \mathbb{P}(A|B_j)\mathbb{P}(B_j)$$

Lorsqu'il y a succession ou enchaînement d'évènements, on utilise

Théorème 5 *la formule des probabilités composées.*

$$\mathbb{P}(A \cap B \cap C \cap \ldots) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) \cdot \mathbb{P}(C|A \cap B) \ldots$$

Exercice. Retrouver la démonstration (très simple) des deux théorèmes.

3.3 Espérance conditionnelle : définitions et propriétés fondamentales

3.3.1 une moyenne relative

Théorème et Définition 1 *Soit X une variable aléatoire intégrable (resp. positive). Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. L'espérance conditionnelle de X sachant $\mathcal{B}$, notée $\mathbb{E}^{\mathcal{B}}X$ est l'unique (à équivalence $\mathbb{P}$ p.s. près) variable aléatoire intégrable (resp. positive) vérifiant*

(i) $\mathbb{E}^{\mathcal{B}}X$ *est* $\mathcal{B}$*-mesurable*

(ii) $\int_B \mathbb{E}^{\mathcal{B}}X\, d\mathbb{P} = \int_B X\, d\mathbb{P}, \quad \forall B \in \mathcal{B}$

Remarque. L'identité (ii) peut être remplacée par (ii')

$$\int_\Omega Z{\cdot}\mathbb{E}^{\mathcal{B}}X\, d\mathbb{P} = \int_\Omega Z{\cdot}X\, d\mathbb{P} \quad \text{pour toute v.a.} Z\ \mathcal{B}\text{-mesurable bornée (resp. positive).}$$

En effet, il suffit de commencer par prendre $Z = 1_B$ puis par linéarité d'étendre aux fonctions étagées, et enfin par passage à la limite aux v.a. bornées (resp. positives). □

Preuve. L'unicité de l'espérance conditionnelle résulte de (ii). En effet, une v.a.$\mathcal{B}$-mesurable Y telle que $\forall B \in \mathcal{B},\ \int_B Y\, d\mathbb{P} = 0$ est nulle $\mathbb{P}\, p.s.$
L'existence de l'espérance conditionnelle est une conséquence du théorème de Radon-Nykodym. En effet, l'application ν

$$B \in \mathcal{B} \mapsto \int_B X\, d\mathbb{P} \in \mathbb{R}$$

est une mesure de Borel *absolument continue* par rapport à $\mathbb{P}$: si $\mathbb{P}(B) = 0$ alors $\nu(B) = 0$. Il existe donc une application $\mathcal{B}$-mesurable Y telle que $\nu(B) = \int_B X\, d\mathbb{P}$ soit égale à $\int_B Y\, d\mathbb{P}$. Cet Y est l'espérance conditionnelle de X sachant $\mathcal{B}$. ■

Exemple. Le cas particulier où $\mathcal{B}$ est une tribu atomique : Soit $(B_j)_{1\le j\le n}$ une partition de Ω en évènements et $\mathcal{B} = \sigma(B_j)_j$

$$\mathbb{E}^{\mathcal{B}}X = \sum_j \frac{\int_{B_j} X\, d\mathbb{P}}{\mathbb{P}(B_j)} \cdot 1_{B_j}$$

Ceci est une conséquence directe du fait que les v.a. $\mathcal{B}$-mesurables sont *constantes* sur les atomes B_j. Cet exemple est fondamental. Conditionner par $\mathcal{B}$ signifie qu'on est à même de déterminer si tel évènement B_j de $\mathcal{B}$ est réalisé. Si c'est le cas, l'espérance conditionnelle est alors la moyenne de X sur l'évènement B_j en question.

Remarque. Si on prend comme cas "encore plus particulier" $\mathcal{B} = \{\emptyset, B, B^c, \Omega\}$, on retrouve

$$\mathbb{E}^{\mathcal{B}}X = \frac{\int_B X\, d\mathbb{P}}{\mathbb{P}(B)} \cdot 1_B + \frac{\int_{B^c} X\, d\mathbb{P}}{\mathbb{P}(B^c)} \cdot 1_{B^c}$$

Du fait que

$$\frac{\int_B 1_A\, d\mathbb{P}}{\mathbb{P}(B)} = \mathbb{P}(A|B) = \int_\Omega 1_A\, d\mathbb{P}_B$$

par extension aux v.a.

$$\frac{\int_B X\, d\mathbb{P}}{\mathbb{P}(B)} = \int_\Omega X\, d\mathbb{P}_B$$

et on notera donc

$$\frac{\int_B X\, d\mathbb{P}}{\mathbb{P}(B)} = \mathbb{E}(X|B)$$

□

Exercice. Considérer T la v.a. température définie sur $\Omega = \{1, \ldots, 365\}$, évidemment muni de la tribu discrète. Soit $\mathcal{S}$ la tribu engendrée par les quatre saisons S_1, S_2, S_3, S_4. Que vaut $\mathbb{E}^{\mathcal{S}} T$?
★ A méditer : Alors que l'espérance est un *nombre*, l'espérance conditionnelle est une *variable aléatoire*. C'est en quelque sorte une prévision moyennée de la variable aléatoire de départ, suivant l'information contenue dans la tribu par laquelle on conditionne. Ceci se reflète dans les propriétés suivantes.

Proposition 4 *Soit X une v.a. intégrable (resp.positive)*

1. *La moyenne de la prévision moyenne de X sachant $\mathcal{B}$ est la moyenne de X :*
$$\mathbb{E}(\mathbb{E}^{\mathcal{B}} X) = \mathbb{E} X$$
2. *Si X est indépendante de la tribu $\mathcal{B}$, la prévision n'apporte rien de plus que l'espérance simple :*
$$\mathbb{E}^{\mathcal{B}} X = \mathbb{E} X$$
3. *A l'opposé, si X est $\mathcal{B}$-mesurable, la prévision devient exacte*
$$\mathbb{E}^{\mathcal{B}} X = X$$
4. *si Z est $\mathcal{B}$-mesurable bornée (resp. positive), elle se comporte* comme une constante *vis à vis de l'espérance conditionnelle :*
$$\mathbb{E}^{\mathcal{B}}(ZX) = Z \cdot \mathbb{E}^{\mathcal{B}} X$$
5. *En particulier, $\mathbb{E}^{\mathcal{B}} 1 = 1$*
6. *Linéarité : $\mathbb{E}^{\mathcal{B}}(aX + bY) = a\mathbb{E}^{\mathcal{B}} X + b\mathbb{E}^{\mathcal{B}} Y$*
7. *Positivité : si $X \geq 0$ p.s. alors $\mathbb{E}^{\mathcal{B}} X \geq 0$ p.s.*

Preuve. Nous démontrons seulement le point 4, les autres sont laissés en exercice. D'abord $Z \cdot \mathbb{E}^{\mathcal{B}} X$ est $\mathcal{B}$-mesurable. Ensuite $\forall B \in \mathcal{B}$,
$$\int_B Z \mathbb{E}^{\mathcal{B}} X \, d\mathbb{P} = \int_\Omega 1_B Z \, \mathbb{E}^{\mathcal{B}} X \, d\mathbb{P}$$
La variante (ii') de la définition donne enfin
$$\int_\Omega 1_B Z \, \mathbb{E}^{\mathcal{B}} X \, d\mathbb{P} = \int_\Omega 1_B Z \, X \, d\mathbb{P} = \int_B Z \, X \, d\mathbb{P}$$
■

Remarque. La propriété 4 est encore vraie si $X \in L^p$ et $Z \in L^q(\mathcal{B})$, pour tous $1 \leq p, q \leq \infty$, $\quad 1/p + 1/q = 1$. Elle est vraie dès que $X \in L^1$ et $ZX \in L^1$. Pour le démontrer, commencer par tronquer Z, i.e. poser
$$Z_n = n \cdot 1_{(Z>n)} - n \cdot 1_{(Z<-n)} + Z \cdot 1_{(-n \leq Z \leq n)}$$
puis passer à la limite $n \to \infty$ en utilisant le théorème de convergence dominée. □

Insistons sur une propriété fondamentale pour la théorie des martingales (voir chapitre ultérieur).

Théorème 6 *Soit $\mathcal{C}$ et $\mathcal{B}$ des sous tribus de $\mathcal{A}$, si $\mathcal{C} \subset \mathcal{B}$ alors $\mathbb{E}^{\mathcal{C}}(\mathbb{E}^{\mathcal{B}}X) = \mathbb{E}^{\mathcal{C}}X$*

Preuve. Tout d'abord $\mathbb{E}^{\mathcal{C}}X$ est $\mathcal{C}$-mesurable donc a fortiori $\mathcal{B}$-mesurable. Ensuite, soit $C \in \mathcal{C}$, d'une part

$$\int_C \mathbb{E}^{\mathcal{C}}X \, d\mathbb{P} = \int_C X \, d\mathbb{P}.$$

D'autre part, comme $\mathcal{C} \subset \mathcal{B}$, $C \in \mathcal{B}$ donc

$$\int_C X \, d\mathbb{P} = \int_C \mathbb{E}^{\mathcal{B}}X \, d\mathbb{P}$$

Donc $\mathbb{E}^{\mathcal{C}}X$ vérifie bien les deux conditions caractérisant $\mathbb{E}^{\mathcal{C}}(\mathbb{E}^{\mathcal{B}}X)$ ■

Le théorème suivant est très utile pour obtenir des inégalités.

Théorème 7 (inégalité de Jensen) *Soit ϕ une fonction convexe réelle, telle que $\phi(X) \in L^1$ (resp. $\phi(X) \geq 0$),*

$$\phi(\mathbb{E}^{\mathcal{B}}X) \leq \mathbb{E}^{\mathcal{B}}\phi(X) \; p.s.$$

En particulier, pour $p \geq 1$

$$|\mathbb{E}^{\mathcal{B}}X|^p \leq \mathbb{E}^{\mathcal{B}}|X|^p, \quad \|\mathbb{E}^{\mathcal{B}}X\|_{L^p} \leq \|X\|_{L^p}$$

Preuve. Elle est très simple si on utilise le fait que toute fonction convexe est l'enveloppe supérieure de ses minorantes affines[3] :

$$\phi(x) = \sup\left\{ax + b; \; (a,b) \in \mathbb{R}^2, \forall t \; at + b \leq \phi(t)\right\}$$

Soit une fonction affine $t \mapsto at + b$ minorant ϕ, par *linéarité* on a :

$$a(\mathbb{E}^{\mathcal{B}}X) + b = \mathbb{E}^{\mathcal{B}}(aX + b) \; p.s.$$

Par *positivité*,

$$a(\mathbb{E}^{\mathcal{B}}X) + b \leq \mathbb{E}^{\mathcal{B}}\phi(X) \; p.s.$$

Comme ceci a lieu pour toute minorante affine, en prenant la borne supérieure du premier membre on obtient le résultat annoncé ■

Signalons enfin que les théorèmes fondamentaux de l'intégration s'étendent à l'espérance conditionnelle.

Théorème 8 *soit $(f_n)_n$ une suite croissante de v.a. positives, alors $(\mathbb{E}^{\mathcal{B}}f_n)_n$ est une suite croissante de v.a. positives et*

$$\lim_n(\mathbb{E}^{\mathcal{B}}f_n) = \mathbb{E}^{\mathcal{B}}(\lim_n f_n)$$

soit $(f_n)_n$ une suite de v.a. dominées par une v.a. g intégrable,

$$\lim_n(\mathbb{E}^{\mathcal{B}}f_n) = \mathbb{E}^{\mathcal{B}}(\lim_n f_n)$$

[3] appelées aussi droites d'appui

Preuve. Démontrons seulement le premier résultat. L'espérance conditionnelle étant positive, $(\mathbb{E}^{\mathcal{B}} f_n)_n$ est une suite croissante de v.a. positives, de plus pour tout n et pour tout $B \in \mathcal{B}$ on a

$$\int_B f_n \, d\mathbb{P} = \int_B \mathbb{E}^{\mathcal{B}} f_n \, d\mathbb{P}$$

Faisons tendre $n \to \infty$ et appliquons le théorème de convergence monotone à chaque membre, on obtient

$$\int_B \lim_n f_n \, d\mathbb{P} = \int_B \lim_n \mathbb{E}^{\mathcal{B}} f_n \, d\mathbb{P}$$

Or $\lim_n \mathbb{E}^{\mathcal{B}} f_n$ est $\mathcal{B}$-mesurable comme limite de fonctions $\mathcal{B}$-mesurables, d'où le résultat. ■

3.3.2 une projection orthogonale ou une prédiction optimale

Lorsqu'on la restreint à $L^2(\Omega, \mathcal{A}, \mathbb{P})$, l'inégalité de Jensen montre que l'espérance conditionnelle est contractante. Elle possède en outre une interprétation géométrique très importante : c'est l'opérateur de projection orthogonale de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ sur $L^2(\Omega, \mathcal{B}, \mathbb{P})$. En effet
(i) $\mathbb{E}^{\mathcal{B}}$ X est $\mathcal{B}$-mesurable
(ii) $(\mathbb{E}^{\mathcal{B}} X - X) \perp Z$, pour toute v.a. $Z \in L^2$ $\mathcal{B}$-mesurable.
A ce titre, il vérifie la propriété de meilleure approximation suivante :

Théorème 9 *Parmi toutes les v.a. Z $\mathcal{B}$-mesurable, $\mathbb{E}^{\mathcal{B}} X$ est la plus proche en norme L^2 de X.*

$$\|X - \mathbb{E}^{\mathcal{B}} X\| = \inf\{\|X - Z\|, \quad Z \in L^2(\Omega, \mathcal{B}, \mathbb{P})\}$$

3.4 Loi conditionnelle. Brefs rappels.

Dans le cas particulier où la tribu $\mathcal{B}$ est engendrée par une v.a. Y, l'espérance conditionnelle est notée

$$\mathbb{E}^{\sigma(Y)} X = \mathbb{E}(X|Y)$$

Rappelons le lemme très utile.

Lemma 1 *La v.a. Z est $\sigma(Y)$-mesurable si et seulement si Z est de la forme $g(Y)$, avec g mesurable.*

D'après la définition, $\mathbb{E}(X|Y)$ est une variable aléatoire qui est $\sigma(Y)$-mesurable, donc il existe une application mesurable g telle que

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

Ainsi il est naturel de noter

$$g(y) = \mathbb{E}(X|Y = y).$$

Cette notation n'est pas fortuite. En effet, elle correspond exactement au conditionnement par l'évènement $(Y = y)$ lorsque celui-ci a une probabilité non nulle.
Exemple. On lance d'abord une pièce équilibrée n fois. On compte le nombre de "face" obtenu, désigné par Y. On relance la pièce Y fois. Soit X le nombre de "face" obtenu au second tour. Déterminons $\mathbb{E}(X|Y)$. Pour cela $\mathbb{E}(X|Y = k)$ représente le nombre moyen de "face" obtenu au second tour, sachant que $Y = k$. La loi conditionnelle $\mathcal{L}_X(\cdot|Y = k)$ est facile à identifier ici : c'est la loi binomiale $\mathcal{B}(k, 1/2)$. D'où $\mathbb{E}(X|Y = k) = k/2$ et $\mathbb{E}(X|Y) = Y/2$.

Plus généralement, lorsque pour toute fonction borélienne bornée (resp. positive) on peut exprimer

$$\mathbb{E}(h(X)|Y = y) = \int_{\mathbb{R}} h(x)\, d\mu_y(x)$$

on dit que μ_y est la *loi conditionnelle* de X sachant $(Y = y)$ que l'on note désormais$\mathcal{L}_X(\cdot|Y = y)$.

Dans le cas où X et Y sont absolument continues, la densité conditionnelle de X sachant Y vaut :

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)} \tag{3.1}$$

★ A méditer : Cette formule se retient aisément, il suffit d'oser faire appel à l'intuition et aux *probabilités infinitésimales.*

$$f_{X|Y=y}(x)\, dx = \mathbb{P}(x \leq X < x + dx|Y = y) =$$

$$\frac{\mathbb{P}(x \leq X < x + dx,\ y \leq Y < y + dy)}{\mathbb{P}(y \leq Y < y + dy)} = \frac{f_{(X,Y)}(x, y)\, dxdy}{f_Y(y)\, dy}$$

Ainsi, la loi conditionnelle se calcule de façon analogue à la probabilité conditionnelle. Une variante intéressante de (3.1) est la version continue de la formule des probabilités totales

$$f_X(x) = \int f_{X|Y=y}(x)\, f_Y(y)dy \tag{3.2}$$

Exercice. Retrouver *à partir de la formule précédente* que la densité d'une somme de v.a. indépendante s'exprime comme le produit de convolution de leurs densités respectives : $f_{X+Y}(x) = \int f_X(x - y)\, f_Y(y)dy$

3.5 Exercices

1. **Propriété d'absence de mémoire.** Soit T v.a.r. suivant une loi exponentielle de paramètre λ. Montrer que pour tous t, s réels positifs,

   $$P(X > t + s|X > t) = P(X > s).$$

   Interpréter. Montrer que cette propriété caractérise la loi exponentielle parmi les lois continues. Quelle est l'analogue parmi les lois discrètes ?

2. **Indépendance partielle.** Soit $\mathcal{B}$ une tribu et X une v.a. On se donne $\mathcal{C}$ une tribu indépendante de $\sigma(X) \cup \mathcal{B}$. Montrer que

   $$E^{\sigma(\mathcal{B}\cup\mathcal{C})}X = E^{\mathcal{B}}X$$

   indications : $\sigma(\mathcal{B} \cup \mathcal{C})$ est engendrée par les $B \cap C$

3. **Echantillon de v.a.** Soit $X_1, X_2, \ldots, X_n$ des v.a. indépendantes et de même loi intégrable. Soit $S_n = \sum_{k \leq n} X_k$. Déterminer $E(X_1|S_n)$ ainsi que $E(S_n|X_1)$
4. **Une propriété d'approximation.** Soit $\mathcal{F}_n$ la tribu sur $[0,1]$ engendrée par les intervalles de la forme $[\frac{k}{n}, \frac{k+1}{n}[$ où $k = 1, \ldots, n-1$. Soit $f : [0,1] \to \mathbb{R}$ une fonction continue, calculer $E^{\mathcal{F}_n} f$. Que se passe-t-il lorsque $n \to \infty$?
5. **Formule des espérances conditionnelles totales.** Soit X une v.a.r. définie sur $(\Omega, \mathcal{A}, P)$. Soit $(A_i)_{i=1}^n$ une partition de Ω en évènements. Montrer que
$$E(X) = \sum_i P(A_i) E(X|A_i).$$
Application : David (fortune initiale 1F) joue à pile ou face contre Goliath (fortune initiale 1MF). La mise est de 1F à chaque partie. Le jeu s'arrête lorsque l'un des deux n'a plus rien. Quelle est la durée moyenne du jeu ? *indications :* Soit X_n (resp. Y_n) la fortune de David (resp. Goliath) à la n^e partie et T la v.a. durée du jeu. Soit $g(x) = E(T|X_0 = x)$. Montrer que $g(x) = 1/2(g(x-1) + g(x+1)) + 1$. Chercher $g(x)$ sous la forme $ax^2 + bx + c$, en déduire que $g(x) = x\,(10^6 + 1 - x)$

Chapitre 4

Processus stochastiques et temps d'arrêts

4.1 Filtration : l'histoire d'un processus

Définition 8 *Un processus stochastique*[1] *est une famille de variables aléatoires* $(X_t)_{t\in\mathbb{T}}$*, qui peut être indexée par* $\mathbb{T} = \mathbb{R}+$ *ou* $\mathbb{T} = \mathbb{N}$*. Dans le cas où* $\mathbb{T} = \mathbb{R}+$*, resp.* $\mathbb{T} = \mathbb{N}$*, on dit que le processus est à temps continu, resp. discret. Dans ce dernier cas, on dit aussi série temporelle ou série chronologique.*

★ A méditer : En général, un processus aléatoire modélise l'évolution au cours du temps d'une quantité aléatoire, par exemple le cours d'un actif financier.

Définition 9 *On appelle filtration naturelle du processus la suite* croissante *de tribus complètes*

$$\mathcal{F}_t = \sigma(X_s,\ s \leq t)$$

Plus généralement, on appelle filtration toute suite croissante de sous-tribus de $\mathcal{A}$*. On dit que le processus est* adapté *si pour tout* t*,* X_t *est* $\mathcal{F}_t$*-mesurable.*

★ A méditer : Le fait que les tribus d'une filtration soient emboîtées $\mathcal{F}_t \subset \mathcal{F}_{t+\delta t}$ reflète le fait que l'information accumulée sur le processus croît avec le temps. Dans certains cas, on aura intérêt à ce que $\mathcal{F}_0$ soit la tribu grossière $\{\emptyset, \Omega\}$. Enfin on désigne par $\mathcal{F}_\infty$ la tribu $\sigma(\bigcup_{t\in\mathbb{T}} \mathcal{F}_t)$.

4.2 Loi d'un processus

La loi d'un processus est la donnée des *lois conjointes* de $(X_{t_1}, X_{t_2} \ldots, X_{t_k})$ pour toutes les suites finies d'instants $t_1, t_2, \ldots t_k$. Ainsi la loi du processus (X_t) permet de calculer les probabilités :

$$\mathbb{P}(X_{t_1} \in B_1, X_{t_2} \in B_2, \ldots X_{t_k} \in B_k)$$

Exercice. Le processus (X_t) est *à accroissement indépendants* si pour tous $t_1 < t_2 < t_3$, $X_{t_3} - X_{t_2}$ est indépendant de $X_{t_2} - X_{t_1}$. Il est dit *stationnaire* si

[1] mot apparu vers 1945, du grec *stokhastês* : devin

la loi de $X_{t_2} - X_{t_1}$ ne dépend que de $(t_2 - t_1)$. On suppose de plus que $X_0 \equiv 0$. On note f_t la densité de X_t. Montrer que

$$\mathbb{P}(X_{t_1} \in B_1, X_{t_2} \in B_2, \dots X_{t_k} \in B_k) =$$

$$\int_{B_1 \times B_2 \times \dots B_k} dx_1 dx_2 \dots dx_k f_{t_1}(x_1) f_{t_2 - t_1}(x_2 - x_1) \dots f_{t_k - t_{k-1}}(x_k - x_{k-1})$$

Indication : penser à la formule des probabilités composées.

4.3 Trajectoires d'un processus

On peut voir aussi un processus stochastique comme une *fonction aléatoire* qui à chaque tirage $\omega \in \Omega$ associe une fonction

$$t \mapsto X_t(\omega)$$

On appelle cette dernière une trajectoire.
★ A méditer : Il y a donc deux points de vue sur un processus stochastique. A chaque instant t, on observe une v.a. X_t. Ou bien à chaque tirage ω, on tire une trajectoire $\{X_t(\omega),\, t \in \mathbb{T}\}$. C'est exactement la même chose en mécanique des fluides avec les points de vue eulerien et lagrangien.

4.4 Temps d'arrêt

Définition 10 *Une variable aléatoire $T : \Omega \to \mathbb{T} \cup \{+\infty\}$ est appelé temps d'arrêt si*

$$\forall t < \infty, \quad (T \leq t) \in \mathcal{F}_t \tag{4.1}$$

Dans le cas particulier où le processus est à temps discret, (4.1) équivaut à

$$\forall n \in \mathbb{N}, \quad (T = n) \in \mathcal{F}_n. \tag{4.2}$$

Preuve. D'une part on a

$$(T = n) = (T \leq n) \cap (T \leq n-1)^c.$$

Comme $(T \leq n) \in \mathcal{F}_n$ et $(T \leq n-1) \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$, (4.1) $\Rightarrow$ (4.2).
Réciproquement,

$$(T \leq n) = \bigcup_{k \leq n} (T = k)$$

Comme $\forall k \leq n,\ (T = k) \in \mathcal{F}_k \subset \mathcal{F}_n$, (4.2) $\Rightarrow$ (4.1). ■

★ A méditer : $(T \leq t) \in \mathcal{F}_t$ signifie qu'à un instant donné t, on sait si on s'est arrêté ou non. Mais on ne sait pas encore si on va s'arrêter avant une date ultérieure $t + s$: $(T \leq t + s) \notin \mathcal{F}_t$

Définition 11 *Soit T un temps d'arrêt. On dit qu'un évènement $A \in \mathcal{F}_\infty$ est* antérieur *à T si*

$$\forall t < \infty, \quad A \cap (T \leq t) \in \mathcal{F}_t. \tag{4.3}$$

L'ensemble de tels évènements constitue une sous-tribu de $\mathcal{F}_\infty$ appelée tribu antérieure *à T notée $\mathcal{F}_T$.*
De même, si le processus est à temps discret, on peut remplacer (4.3) par

$$\forall n \in \mathbb{N}, \quad A \cap (T = n) \in \mathcal{F}_n. \tag{4.4}$$

Preuve. Il s'agit de voir que
– $\mathcal{F}_T$ contient Ω
– $\mathcal{F}_T$ est stable par complémentation
– $\mathcal{F}_T$ est stable par réunion dénombrable.
Montrons la stabilité par complémentation. Les autres points sont triviaux. Soit $A \in \mathcal{F}_T$, et $t < \infty$, l'idée clef est que

$$A^c \cap (T \leq t) = (A \cap (T \leq t))^c \cap (T \leq t) \tag{4.5}$$

En effet $(A \cap (T \leq t))^c = A^c \cup (T > t)$ donc

$$(A \cap (T \leq t))^c \cap (T \leq t) = (A^c \cap (T \leq t)) \cup ((T > t) \cap (T \leq t))$$

or $(T > t) \cap (T \leq t) = \emptyset$ d'où (4.5). Comme d'une part $(A \cap (T \leq t) \in \mathcal{F}_t$ donc aussi $(A \cap (T \leq t))^c$ et d'autre part $(T \leq t) \in \mathcal{F}_t$, le résultat en découle. ■

Exemple. Temps d'atteinte d'une partie. Soit D une partie borélienne de $\mathbb{R}$,

$$\tau_D = \inf\{n \in \mathbb{N}\ X_n \in D\}$$

avec la convention

$$\inf \emptyset = +\infty$$

est un temps d'arrêt.
En effet, pour tout $n \in \mathbb{N}$

$$(\tau_D = n) = (X_n \in D) \cap \bigcap_{k \leq n} (X_k \notin D)$$

Or $(X_n \in D) \in \mathcal{F}_n$ et $(X_k \notin D) = (X_k \in D)^c \in \mathcal{F}_k \subset \mathcal{F}_n$ d'où le résultat.
Illustrons par cet exemple la pertinence de la notion de temps d'arrêt. Soit un joueur dans un casino (ou un actif financier sur un marché...) qui mise initialement X_0. Soit X_n le gain algébrique (positif ou négatif) à la n^e partie (ou le n^e jour). Soit $S_n = \sum_{0 \leq k \leq n} X_k$ la fortune totale au temps n. Une stratégie peut être de s'arrêter (resp. vendre, acheter) dès que S_n dépasse un certain seuil b ou descend en dessous d'un seuil a. On est amené naturellement à introduire

$$T = \inf\{n \geq 1;\ S_n \in [a, b]^c\}$$

Exercice. Soit X une v.a. et T un temps d'arrêt. Montrer que, sur $(T = n)$, $\mathbb{E}^{\mathcal{B}_T} X = \mathbb{E}^{\mathcal{B}_n} X$

Proposition 5 *Soit S et T deux temps d'arrêts, $S \wedge T$ est aussi un temps d'arrêt.*

Preuve. Il suffit de voir que

$$((S \wedge T) \leq t) = (S \leq t) \cup (T \leq t).$$

■

Remarque. Attention, selon le tirage ω, $(S \wedge T)(\omega)$ peut valoir $S(\omega)$ ou $T(\omega)$. Le plus petit de S et T varie selon l'épreuve. □

Le terme antérieur qui qualifie la tribu $\mathcal{F}_T$ n'est pas fortuit.

Proposition 6 *Soit S et T deux temps d'arrêts, si $S \leq T$ p.s., alors $\mathcal{F}_S \subset \mathcal{F}_T$.*

Preuve. Soit $A \in \mathcal{F}_S$, cela signifie que $\forall t < \infty, \quad A \cap (S \leq t) \in \mathcal{F}_t$ Comme $S \leq T$ p.s.,

$$(T \leq t) \subset (S \leq t),$$

et $A \cap (T \leq t) = A \cap (S \leq t) \cap (T \leq t)$, à une partie négligeable près. Or $A \cap (S \leq t) \in \mathcal{F}_t$ et $(T \leq t) \in \mathcal{F}_t$ car T est un temps d'arrêt, de plus $\mathcal{F}_t$ est supposée complète, donc $A \cap (T \leq t) \in \mathcal{F}_t$ ■

A l'aide d'un temps d'arrêt T, d'une variable aléatoire $\mathcal{F}_\infty$-mesurable X_∞ et d'un processus adapté (X_n), on peut construire une nouvelle variable X_T par

$$X_T(\omega) = X_{T(\omega)}(\omega)$$

ou de façon équivalente

$$X_T(\omega) = \begin{cases} X_n(\omega) & si \quad T(\omega) = n \\ X_\infty(\omega) & si \quad T(\omega) = \infty \end{cases}$$

On vérifie aisément que

Proposition 7 *X_T est une variable aléatoire $\mathcal{F}_T$-mesurable*

Preuve. Il suffit d'écrire

$$X_T = \sum_n 1_{(T=n)} X_n + 1_{(T=\infty)} X_\infty$$

et d'utiliser le lemme

Lemma 2 *Soit Z une v.a. $\mathcal{F}_\infty$-mesurable , Z est $\mathcal{F}_T$-mesurable ssi $\forall n$, $1_{(T=n)} Z$ est $\mathcal{F}_n$ mesurable.*

■

On peut même construire une nouveau processus :

Définition 12 *Soit (X_n) un processus adapté et T un temps d'arrêt. Le processus $(X_{n \wedge T})$ est défini par*

$$X_{n \wedge T}(\omega) = \begin{cases} X_n(\omega) & si \quad T(\omega) \geq n \\ X_{T(\omega)}(\omega) & si \quad T(\omega) < n \end{cases}$$

On l'appelle processus arrêté *à l'instant T. C'est encore un processus adapté.*

4.5 Exercices

1. **Minimum de deux temps d'arrêt.** Soit $(\mathcal{F}_n)_n$ une filtration et S, T deux temps d'arrêt. Montrer que

$$\mathcal{F}_{S \wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$$

2. **Connaissance progressive d'une v.a.** Soit $(\mathcal{F}_n)_n$ une filtration, X une v.a.intégrable ou positive et T un temps d'arrêt. Posons pour tout $n \in \bar{\mathbb{N}}$, $X_n = E(X|\mathcal{F}_n)$. Montrer que

$$X_T = E(X|\mathcal{F}_T)$$

Chapitre 5

Martingales

Dans la suite de ce cours, on se limite aux processus à temps discret. $(\mathcal{B}_n)_{n\in\mathbb{N}}$ désigne une filtration. $\mathcal{B}_\infty$ est la tribu $\sigma(\bigcup_{n\in\mathbb{N}}\mathcal{B}_n)$.

5.1 Définition et interprétations

Définition 13 *Un processus $(X_n)_n$ adapté à une filtration $(\mathcal{B}_n)_n$ est appelé une $\mathcal{B}_n$- martingale* [1] *(resp. surmartingale, sous-martingale) si*
(i) $\forall n$ X_n est intégrable (ou positive)
(ii) $\mathbb{E}^{\mathcal{B}_n}X_{n+1} = X_n$ (resp. $\leq X_n$, resp. $\geq X_n$)

Remarque. Si (X_n) est une martingale (resp. surmartingale , sous-martingale) pour une filtration $(\mathcal{B}_n)_{n\in\mathbb{N}}$, elle l'est aussi pour sa filtration naturelle. La réciproque n'est pas vraie. □

Remarque. La suite (X_n) est une surmartingale ssi $(-X_n)$ est une sous-martingale. □

La terminologie sur ou sous-martingale est trompeuse : une sous-martingale est *croissante* en moyenne conditionnelle. Autrement dit, pour gagner en moyenne, il faut trouver une *sous*-martingale.

Proposition 8 *Si $(X_n)_n$ est une $\mathcal{B}_n$-martingale (resp. surmartingale , sous-martingale) alors*
(i) $\mathbb{E}^{\mathcal{B}_n}(X_{n+1} - X_n) = 0$ (resp. ≤ 0, resp. ≥ 0)
(ii) la suite $\mathbb{E}X_n$ est constante (resp. décroissante, croissante).

Preuve. (i) Par linéarité $\mathbb{E}^{\mathcal{B}_n}(X_{n+1} - X_n) = \mathbb{E}^{\mathcal{B}_n}X_{n+1} - \mathbb{E}^{\mathcal{B}_n}X_n$. Or X_n est adapté donc $\mathbb{E}^{\mathcal{B}_n}X_n = X_n$.
(ii) Il suffit d'utiliser que $\mathbb{E}(\mathbb{E}^{\mathcal{B}}X) = \mathbb{E}X$. ■

Exemple. Soit X une v.a. intégrable (resp. positive) et $X_n = \mathbb{E}^{\mathcal{B}_n}X$. On vérifie aisément que $(X_n)_n$ est une $\mathcal{B}_n$-martingale intégrable (resp.positive).

[1] du provençal *martegalo* Martigues. *jouga a la martegalo* : jouer à la martingale. Système de jeu fondé sur des considérations découlant du calcul des probabilités et qui prétend assurer un bénéfice certain dans les jeux de hasard.

Exemple. Stratégie de placement. Soit (X_n) une $\mathcal{B}_n$-martingale intégrable. Soit (Y_n) une processus *prévisible*, i.e. pour tout $n \geq 1$, Y_n est $\mathcal{B}_{n-1}$-mesurable et Y_0 est $\mathcal{B}_0$-mesurable. On suppose en outre que pour tout n, Y_n est borné. On pose

$$Z_n = Y_0\, X_0 + \sum_{1 \leq k \leq n} Y_k\, (X_k - X_{k-1})$$

Le processus (Z_n) une $\mathcal{B}_n$-martingale.
★ A méditer : L'exemple précédent s'interprète en terme de finance. Si X_k désigne le cours d'une action donnée au temps k, et Y_k la quantité d'actions détenue au temps k, qui est prévisible puisque le financier décide en fonction des cours *précédents* combien d'actions il va détenir, le bénéfice (algèbrique!) réalisé entre le jour $k-1$ et le jour k est $Y_k\, (X_k - X_{k-1}$. La fortune totale accumulée au temps n est donc précisément :

$$Z_n = Y_0\, X_0 + \sum_{1 \leq k \leq n} Y_k\, (X_k - X_{k-1})$$

Exemple. Marche aléatoire. Soit (X_n) une suite de v.a. indépendantes et centrées. Soit

$$S_n = \sum_{1 \leq k \leq n} X_k.$$

S_n est une martingale pour la filtration $\mathcal{F}_n = \sigma(X_k,\ k \leq n)$
La propriété suivante généralise la définition.

Proposition 9 *Si $(X_n)_n$ est une $\mathcal{B}_n$-martingale (resp. surmartingale , sous-martingale) alors pour $m \geq n$ $\mathbb{E}^{\mathcal{B}_n} X_m = X_n$ (resp. $\leq X_n$, resp. $\geq X_n$.)*

Preuve. Comme les tribus $\mathcal{B}_n$ sont emboîtées, le théorème 6 donne

$$\mathbb{E}^{\mathcal{B}_n} X_m = \mathbb{E}^{\mathcal{B}_n} \mathbb{E}^{\mathcal{B}_{n+1}} \ldots \mathbb{E}^{\mathcal{B}_{m-1}} X_m$$

en appliquant la définition à répétition, on obtient le résultat ■

La propriété fondamentale des martingales est que ce résultat s'étend aux instants aléatoires, pourvu qu'ils soient des temps d'arrêts *bornés* et fait l'objet de la prochaine section.

5.2 Martingales et temps d'arrêts

Les martingales se comportent très bien vis à vis des temps d'arrêts.

Théorème 10 *Soit (X_n) une (resp. sur, sous-) martingale et T un temps d'arrêt. Alors le processus arrêté $(X_{n \wedge T})$ est une (resp. sur, sous-) martingale.*

Preuve. On a $X_{(n+1)\wedge T} - X_{n \wedge T} = 1_{(T \geq n+1)}\, (X_{n+1} - X_n)$. Or $(T \geq n+1) = (T \leq n)^c \in \mathcal{B}_n$ d'où le résultat en conditionnant par $\mathcal{B}_n$. ■

Théorème 11 (d'arrêt borné) *Soit $(X_n)_n$ une martingale (resp. surmartingale , sous-martingale), $S \leq T$ deux temps d'arrêts* bornés. *Alors*

$$\mathbb{E}^{\mathcal{B}_S} X_T = X_S \ (resp. \leq X_S,\ resp. \geq X_S) \tag{5.1}$$

Preuve. On suppose que $S \leq T \leq k$. Dans le cas où S et T sont constants, le résultat a déjà été prouvé dans la proposition 9. Supposons maintenant que $T = n$ est constant. Montrons

$$\mathbb{E}^{\mathcal{B}_S} X_n = X_S \tag{5.2}$$

Tout d'abord, X_S est bien $\mathcal{B}_S$-mesurable, d'après la proposition 7. Ensuite, soit $B \in \mathcal{B}_S$, comme $S \leq T = n$ on a

$$\int_B X_S \, d\mathbb{P} = \int_B \sum_{m \leq n} 1_{(S=m)} X_m \, d\mathbb{P}$$

$$\int_B \sum_{m \leq n} 1_{(S=m)} X_m \, d\mathbb{P} = \int \sum_{m \leq n} 1_{B \cap (S=m)} X_m \, d\mathbb{P}$$

Utilisons que $X_m = \mathbb{E}^{\mathcal{B}_m} X_n$

$$\int \sum_{m \leq n} 1_{B \cap (S=m)} X_m \, d\mathbb{P} = \sum_{m \leq n} \int 1_{B \cap (S=m)} \mathbb{E}^{\mathcal{B}_m} X_n \, d\mathbb{P}$$

Or $B \cap (S = m) \in \mathcal{B}_m$ donc

$$\sum_{m \leq n} \int 1_{B \cap (S=m)} \mathbb{E}^{\mathcal{B}_m} X_n \, d\mathbb{P} = \sum_{m \leq n} \int 1_{B \cap (S=m)} X_n \, d\mathbb{P}$$

$$\sum_{m \leq n} \int 1_{B \cap (S=m)} X_n \, d\mathbb{P} = \int_B (\sum_{m \leq n} 1_{(S=m)}) X_n \, d\mathbb{P}$$

$$\int_B (\sum_{m \leq n} 1_{(S=m)}) X_n \, d\mathbb{P} = \int_B X_n \, d\mathbb{P}$$

ce qui établit (5.2). Montrons maintenant le théorème dans le cas le plus général. Pour cela considérons le processus arrêté $X'_n = X_{n \wedge T}$, c'est encore une martingale, d'après le théorème précédent. Comme $S \leq k$, nous pouvons appliquer la relation (5.2) au processus (X'_n) à l'instant $n = k$

$$\mathbb{E}^{\mathcal{B}_S} X'_k = X'_S$$

Or $X'_k = X_{k \wedge T} = X_T$ et $X'_S = X_{S \wedge T} = X_S$ d'où le résultat. ■

En prenant l'espérance dans (5.1) et $S = 0$.

Corollaire 1 *Soit $(X_n)_n$ une martingale (resp. surmartingale , sous-martingale), et T un temps d'arrêt borné alors*

$$\mathbb{E}X_T = \mathbb{E}X_0 \quad (resp. \leq, \geq)$$

Remarque. Si les temps d'arrêts ne sont pas bornés, le résultat est faux. Voici un contre-exemple. Considérons à nouveau le jeu de pile ou face, qui est une martingale d'après l'exemple précédent. Notons $S_n = \sum_k X_k$ et $T = \inf\{n; S_n \geq 1\}$. On peut démontrer (cf exercice) que $T < \infty$ p.s.. Alors il est clair que $S_T \equiv 1$. Supposons la fortune initiale nulle $X_0 = 0$. Appliquons le corollaire,

$$\mathbb{E}S_T = \mathbb{E}S_0$$

Or $S_0 \equiv 0$ et $S_T \equiv 1$ d'où la contradiction. Ici T n'est pas borné, bien que fini p.s. ! Ce phénomène est bien connu des joueurs, un jour ou l'autre on finit par gagner, mais l'attente peut-être très, très longue ... On peut démontrer (cf exercice) que $\mathbb{E}T = +\infty$ □

5.3 Convergence presque sûre

Nous commençons par traiter le cas des surmartingales positives. Une surmartingale décroît en moyenne conditionnelle. Nous allons exploiter à fond cette propriété, conjointement à la positivité.

5.3.1 Le cas des surmartingales positives

Le corollaire du théorème d'arrêt, dans le cas des surmartingales positives seulement, s'étend aux temps d'arrêt quelconques.

Théorème 12 (théorème d'arrêt fini) *Soit une surmartingale positive* $(X_n)_n$*, et* T *un temps d'arrêt* quelconque*, alors*

$$\mathbb{E}X_T 1_{(T<\infty)} \leq \mathbb{E}X_0$$

Preuve. On a

$$X_T 1_{(T<\infty)} = \lim_k X_{T\wedge k} 1_{(T<\infty)}$$

Donc le lemme de Fatou donne

$$\mathbb{E}X_T 1_{(T<\infty)} \leq \liminf \mathbb{E}X_{T\wedge k} 1_{(T<\infty)}$$

D'autre part le théorème d'arrêt appliqué au temps d'arrêt $T \wedge k$ donne

$$\mathbb{E}X_{T\wedge k} \leq \mathbb{E}X_0$$

Comme $\mathbb{E}X_{T\wedge k} 1_{(T<\infty)} \leq \mathbb{E}X_{T\wedge k}$, a fortiori $\mathbb{E}X_{T\wedge k} 1_{(T<\infty)} \leq \mathbb{E}X_0$, d'où l'on tire

$$\mathbb{E}X_T 1_{(T<\infty)} \leq \mathbb{E}X_0$$

■

Corollaire 2 *Soit une surmartingale positive* $(X_n)_n$*,* S *et* T *deux temps d'arrêt* quelconques*, vérifiant* $S \leq T$ *alors*

$$\mathbb{E}X_T 1_{(T<\infty)} \leq \mathbb{E}X_S 1_{(S<\infty)}$$

Preuve. On effectue une translation de l'origine des temps, i.e. on considère

$$Y_n = X_{S+n} 1_{(S<\infty)}$$

On vérifie aisément que le processus $(Y_n)_n$ est une $\mathcal{B}_{S+n}$-surmartingale. Soit

$$\tau = \begin{cases} T - S & sur \quad (T < \infty) \\ +\infty & sur \quad (T = \infty) \end{cases}$$

On vérifie aisément que τ est un temps d'arrêt pour la filtration $(\mathcal{B}_{S+n})_n$ Le théorème précédent appliqué à $(Y_n)_n$ et τ donne

$$\mathbb{E}Y_\tau 1_{(\tau<\infty)} \leq \mathbb{E}Y_0$$

or $Y_\tau 1_{(\tau<\infty)} = X_T 1_{(T<\infty)}$ et $Y_0 = X_S 1_{(S<\infty)}$ d'où le résultat ■

Moralement, une surmartingale tend à décroître en moyenne. Il semble naturel qu'elle soit bornée supérieurement. Plus précisément, on a le théorème.

Proposition 10 (inégalité maximale) *Soit $(X_n)_n$ une surmartingale positive et $a > 0$, alors*

$$a\mathbb{P}(\sup_n X_n > a) \leq \mathbb{E}\min(X_0, a)$$

De plus si X_0 est finie p.s. alors la v.a. $X^\star = \sup_n X_n$ est finie p.s.

Preuve. Soit $T_a = \inf\{n \geq 0;\ X_n > a\}$, avec la convention $\inf \emptyset = \infty$. Soit

$$Y_n = \begin{cases} X_n(\omega) & si \quad n < T(\omega) \\ a & si \quad n \geq T(\omega) \end{cases}$$

Montrons que $(Y_n)_n$ est une surmartingale. On peut exprimer

$$Y_n = 1_{(T_a>n)}X_n + 1_{(T_a\leq n)}a$$

donc $(Y_n)_n$ est adaptée. D'autre part

$$Y_{n+1} = 1_{(T_a>n+1)}X_{n+1} + 1_{(T_a\leq n+1)}a$$

Comme sur $(T_a = n+1)$, $X_{n+1} > a$, on peut majorer

$$Y_{n+1} \leq 1_{(T_a\geq n+1)}X_{n+1} + 1_{(T_a\leq n)}a$$

$$\mathbb{E}^{\mathcal{B}_n}Y_{n+1} \leq \mathbb{E}^{\mathcal{B}_n}(1_{(T_a\geq n+1)}X_{n+1} + 1_{(T_a\leq n)}a)$$

Les propriétés de l'espérance conditionnelle et le fait que T_a est un temps d'arrêt donnent alors

$$\mathbb{E}^{\mathcal{B}_n}Y_{n+1} = 1_{(T_a\geq n+1)}\mathbb{E}^{\mathcal{B}_n}X_{n+1} + a1_{(T_a\leq n)}$$

Comme $(X_n)_n$ est une surmartingale on a

$$\mathbb{E}^{\mathcal{B}_n}Y_{n+1} \leq 1_{(T_a>n)}X_n + a1_{(T_a\leq n)} = Y_n$$

Utilisons le fait que Y_n est une surmartingale et appliquons le corollaire du théorème d'arrêt au temps d'arrêt borné $T_a \wedge n$

$$\mathbb{E}Y_{n\wedge T_a} \leq \mathbb{E}Y_0$$

Or $Y_{n\wedge T_a} = X_n 1_{(T_a>n)} + a1_{(T_a\leq n)}$ donc $\mathbb{E}Y_{n\wedge T_a} \geq a\mathbb{P}(T_a \leq n)$. De plus l'évènement

$$(T_a \leq n) = (\sup_{k\leq n} X_k > a)$$

donc

$$a\mathbb{P}(\sup_{k\leq n} X_k > a) \leq \mathbb{E}min(X_0, a)$$

en faisant tendre $n \to \infty$ on obtient l'inégalité annoncée. Montrons que $\sup_k X_k$ est finie p.s.

$$\mathbb{P}(\sup_k X_k > a) \leq \mathbb{E}min(X_0/a, 1)$$

Faisons tendre $a \to \infty$. La v.a. X_0 est finie donc $X_0/a \to 0$ Donc par le théorème de convergence dominée, $\mathbb{E}min(X_0/a, 1) \to 0$ donc

$$\mathbb{P}(\sup_k X_k = \infty) = \lim_{N\to\infty} \mathbb{P}(\sup_k X_k > N) = 0$$

ainsi $\sup_k X_k$ est p.s. finie ■

Remarque. La démonstration précédente est instructive, cependant on peut donner une démonstration plus courte en appliquant le théorème d'arrêt fini à la surmartingale $Y_n := X_n \wedge a$ et $T = \inf\{n \geq 0;\ X_n > a\}$. □

Non seulement les surmartingales positives sont bornées p.s. mais le théorème fondamental affirme que les surmartingales positives convergent p.s. La démonstration est subtile mais est basée sur une idée simple : bien qu'une surmartingale ne soit pas décroissante p.s., mais seulement en moyenne, ses "montées" à répétition sont très peu probables.

Théorème 13 *Toute surmartingale positive $(X_n)_n$ converge presque sûrement. De plus la limite $X_\infty = \lim X_n$ satisfait les inégalités*

$$\mathbb{E}^{\mathcal{B}_n} X_\infty \leq X_n$$

Preuve. Elle repose sur le lemme suivant

Lemma 3 (des montées) *Soit $0 < a < b$. La probabilité que X_n remonte au moins k fois d'une valeur inférieure à a à une valeur supérieure à b est moindre que $(a/b)^k$*

Définissons par récurrence la suite croissante de temps d'arrêt $(\tau_n)_n$.

$$\tau_0 = 0, \quad \tau_1 = \min\{n \geq \tau_0,\ X_n < a\}, \quad \tau_2 = \min\{n \geq \tau_1,\ X_n > b\}$$

$$\tau_{2k-1} = \min\{n \geq \tau_{2k-2},\ X_n < a\}, \quad \tau_{2k} = \min\{n \geq \tau_{2k-1},\ X_n > b\}, \ \ldots$$

L'évènement "X_n remonte au moins k fois d'une valeur inférieure à a à une valeur supérieure à b" est précisément $(\tau_{2k} < \infty)$. Utilisons le corollaire du théorème d'arrêt généralisé.

$$\mathbb{E}X_{\tau_{2k}} 1_{(\tau_{2k}<\infty)} \leq \mathbb{E}X_{\tau_{2k-1}} 1_{(\tau_{2k-1}<\infty)}$$

Il suffit alors de remarquer que

$$X_{\tau_{2k}} 1_{(\tau_{2k}<\infty)} \geq b 1_{(\tau_{2k}<\infty)}$$

et que

$$X_{\tau_{2k-1}} 1_{(\tau_{2k-1}<\infty)} \leq a 1_{(\tau_{2k-1}<\infty)}$$

pour obtenir

$$b\mathbb{P}(\tau_{2k} < \infty) \leq a\mathbb{P}(\tau_{2k-1} < \infty)$$

or $(\tau_{2k-1} < \infty) \subset (\tau_{2k-2} < \infty)$ donc $b\mathbb{P}(\tau_{2k} < \infty) \leq a\mathbb{P}(\tau_{2k-2} < \infty)$ et par récurrence

$$\mathbb{P}(\tau_{2k} < \infty) \leq (a/b)^k$$

■

Considérons maintenant l'évènement $\bigcap_k (\tau_{2k} < \infty)$, comme il s'agit d'une intersection *décroissante*, on peut affirmer que

$$\mathbb{P}(\bigcap_k (\tau_{2k} < \infty)) = \lim_k (a/b)^k = 0$$

D'autre part

$$\bigcap_k (\tau_{2k} < \infty) = (\limsup X_n \geq b > a \geq \liminf X_n)$$

En effet l'évènement est réalisé ssi la suite traverse une infinité de fois la bande $[a, b]$. Enfin l'évènement

$$(\limsup X_n > \liminf X_n) = \bigcup_{(a,b)\in\mathbb{Q}^2 : a<b} (\limsup X_n \geq b > a \geq \liminf X_n)$$

est négligeable donc la suite converge p.s. vers une variable aléatoire positive, éventuellement infinie, notée X_∞. $(X_n)_n$ étant une surmartingale, on peut écrire

$$\mathbb{E}^{\mathcal{B}_n} X_{n+p} \leq X_n$$

Faisons tendre $p \to \infty$, le lemme de Fatou généralisé aux espérances conditionnelles (exercice!) donne

$$\mathbb{E}^{\mathcal{B}_n} X_\infty \leq \liminf_{p\to\infty} \mathbb{E}^{\mathcal{B}_n} X_{n+p} \leq X_n$$

■

Remarque. D'après l'inégalité maximale, X_∞ est finie p.s. si $X_0 < \infty$ p.s. On déduit de $\mathbb{E}^{\mathcal{B}_n} X_\infty \leq X_n$ que $X_\infty < \infty$ p.s. hors de $\bigcap_n (X_n = +\infty)$. En effet X_∞ est intégrable sur chacun des évènements $(\mathbb{E}^{\mathcal{B}_n} X_\infty \leq N)$ donc est finie p.s. sur leur réunion $(\mathbb{E}^{\mathcal{B}_n} X_\infty < \infty)$, enfin $\bigcup_n (X_n < \infty) \subset (\mathbb{E}^{\mathcal{B}_n} X_\infty < \infty)$ □

5.3.2 Le cas des sous-martingales bornées dans L^1.

Le théorème de convergence s'applique évidemment aux martingales *positives*, ainsi qu'aux sous-martingales négatives. Mais pour s'affranchir de l'hypothèse de signe, il faut supposer la (sous-)martingale bornée dans L^1. C'est l'objet du théorème suivant.

Théorème 14 *Toute sous-martingale intégrable $(X_n)_n$ telle que*

$$\sup_n \mathbb{E} X_n^+ < \infty$$

converge presque sûrement. De plus la limite $X_\infty = \lim X_n \in L^1$

Preuve. Comme (X_n) est une sous-martingale intégrable, la suite (X_n^+) est une sous-martingale intégrable. En effet, on a $\mathbb{E}^{\mathcal{B}_n} X_{n+1}^+ \geq 0$ et

$$\mathbb{E}^{\mathcal{B}_n} X_{n+1}^+ \geq \mathbb{E}^{\mathcal{B}_n} X_{n+1} \geq X_n$$

donc

$$\mathbb{E}^{\mathcal{B}_n} X_{n+1}^+ \geq X_n^+$$

Mais alors, pour n fixé la suite $(\mathbb{E}^{\mathcal{B}_n} X_p^+)_{p \geq n}$ est croissante. En effet,

$$\mathbb{E}^{\mathcal{B}_n} X_{p+1}^+ = \mathbb{E}^{\mathcal{B}_n} \mathbb{E}^{\mathcal{B}_p} X_{p+1}^+ \geq \mathbb{E}^{\mathcal{B}_n} X_p^+$$

Posons alors

$$M_n = \lim_{p \to \infty} \mathbb{E}^{\mathcal{B}_n} X_p^+$$

et montrons que cette suite de v.a. positives est une martingale intégrable, donc finie p.s. Tout d'abord, M_n est $\mathcal{B}_n$-mesurable par construction. D'autre part,

$$\mathbb{E}^{\mathcal{B}_n} M_{n+1} = \mathbb{E}^{\mathcal{B}_n} \lim_{p \to \infty} \mathbb{E}^{\mathcal{B}_{n+1}} X_p^+$$

En appliquant le théorème de convergence monotone des espérances conditionnelles,

$$\mathbb{E}^{\mathcal{B}_n} M_{n+1} = \lim_{p \to \infty} \mathbb{E}^{\mathcal{B}_n} \mathbb{E}^{\mathcal{B}_{n+1}} X_p^+$$

Or $\mathbb{E}^{\mathcal{B}_n} \mathbb{E}^{\mathcal{B}_{n+1}} X_p^+ = \mathbb{E}^{\mathcal{B}_n} X_p^+$ donc

$$\mathbb{E}^{\mathcal{B}_n} M_{n+1} = \lim_{p \to \infty} \mathbb{E}^{\mathcal{B}_n} X_p^+ = M_n$$

Enfin, par hypothèse

$$\sup_n \mathbb{E} X_n^+ < \infty$$

Donc, par le théorème de convergence monotone,

$$\mathbb{E} M_n = \lim_{p \to \infty} \mathbb{E} \mathbb{E}^{\mathcal{B}_n} X_p^+ = \lim_{p \to \infty} \mathbb{E} X_p^+ < \infty$$

Considérons alors la différence $Y_n = M_n - X_n$. Elle définit une surmartingale positive et intégrable. En effet, $M_n = \lim_{p \to \infty} \mathbb{E}^{\mathcal{B}_n} X_p^+ \geq X_n^+$ donc $Y_n \geq 0$. De plus, M_n et X_n sont intégrable et $\mathcal{B}_n$-mesurable donc Y_n aussi. Enfin M_n et $-X_n$ sont des surmartingales donc leur somme aussi.
Le théorème de convergence p.s. des surmartingales positives permet de définir $M_\infty = \lim_n M_n$ et $Y_\infty = \lim_n Y_n$. Ces v.a. sont positives et aussi intégrables, car $\mathbb{E} M_\infty \leq \mathbb{E} M_0$ et $\mathbb{E} Y_\infty \leq \mathbb{E} Y_0$. Elles sont donc *finies* p.s. et leur différence $M_\infty - Y_\infty$ a un un sens donc $\lim_n X_n = M_\infty - Y_\infty$ ■

Remarque. Pour une sous-martingale intégrable, la condition

$$\sup_n \mathbb{E} X_n^+ < \infty$$

est équivalente à la condition

$$\sup_n \mathbb{E} |X_n| < \infty$$

En effet $|X_n| + X_n = 2X_n^+$ et $\mathbb{E}X_n \geq \mathbb{E}X_0$ □

Remarque. La démonstration précédente s'appuie sur la décomposition $X_n = M_n - Y_n$ d'une sous-martingale intégrable comme différence d'une martingale positive intégrable et d'une surmartingale positive intégrable. Cette décomposition est appelée *décomposition de Krickeberg.* □

Remarque. La limite X_∞ ne satisfait pas forcément les inégalités

$$\mathbb{E}^{\mathcal{B}_n} X_\infty = X_n \quad (resp. \geq)$$

En particulier, pour une martingale bornée dans L^1 on n'a pas forcément $\mathbb{E}X_\infty = \mathbb{E}X_n$ et la suite ne converge pas nécessairement dans L^1. Pour cela, il faut des conditions plus restrictives, qui font l'objet de la section suivante. □

5.4 Martingales uniformément intégrables

5.4.1 Uniforme intégrabilité et convergence dans L^1

La notion d'uniforme intégrabilité est une notion de théorie de la mesure.

Définition 14 *Une famille de fonctions mesurables $\{f_j\}$ est uniformément intégrable pour une mesure bornée μ si*

$$\int_{(|f_j|>M)} |f_j| d\mu \to 0 \quad \text{quand } M \to +\infty$$

uniformément par rapport à j.

Remarque. Certains auteurs parlent d'équi-intégrabilité ou de suite tendue. □

Le théorème suivant généralise le théorème de convergence dominée.

Théorème 15 (Vitali) *Soit une suite $(f_n)_n$ de fonctions mesurables qui converge presque partout vers une fonction f. Alors la suite converge dans L^1 ssi la suite est uniformément intégrable. Dans ce cas f est intégrable et la suite converge vers f dans L^1.*

Remarque. si la suite est dominée par une fonction intégrable, i.e.

$$\sup_j |f_j| \leq g \in L^1$$

alors la suite est trivialement uniformément intégrable. Le théorème de Vitali généralise donc bien le théorème de Lebesgue. □

Une condition suffisante d'uniforme intégrabilité est la suivante

Théorème 16 (La Vallée-Poussin) *Soit ψ une fonction positive coercive i.e. telle que*

$$\lim_{x\to\infty} \frac{\psi(x)}{x} = +\infty$$

Si

$$\sup_j \int \psi(|f_j|) d\mathbb{P} < \infty$$

alors la famille $\{f_j\}$ est uniformément intégrable

Corollaire 3 *Une suite bornée dans L^p avec $p > 1$ est uniformément intégrable*

Preuve. Il suffit de prendre $\psi(x) = |x|^p$ ■

5.4.2 Martingales régulières

Commençons par montrer la

Proposition 11 *Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, l'ensemble $\left\{\mathbb{E}^{\mathcal{B}} X, \quad \mathcal{B} \subset \mathcal{A}\right\}$ est uniformément intégrable.*

Preuve. Tout d'abord, nous avons $|\mathbb{E}^{\mathcal{B}} X| \leq \mathbb{E}^{\mathcal{B}}|X|$ donc

$$\int_{(|\mathbb{E}^{\mathcal{B}} X|>a)} |\mathbb{E}^{\mathcal{B}} X| d\mathbb{P} \leq \int_{(\mathbb{E}^{\mathcal{B}}|X|>a)} \mathbb{E}^{\mathcal{B}}|X| d\mathbb{P} = \int_{(\mathbb{E}^{\mathcal{B}}|X|>a)} |X| d\mathbb{P}$$

car $(\mathbb{E}^{\mathcal{B}}|X| > a)$ est $\mathcal{B}$-mesurable. Partitionnons Ω en $(|X| > b)$ et $(|X| \leq b)$

$$\int_{(\mathbb{E}^{\mathcal{B}}|X|>a)} |X| d\mathbb{P} \leq b\mathbb{P}(\mathbb{E}^{\mathcal{B}}|X| > a) + \int_{(|X|>b)} |X| d\mathbb{P}$$

D'autre part

$$\mathbb{P}(\mathbb{E}^{\mathcal{B}}|X| > a) \leq a^{-1}\mathbb{E}(\mathbb{E}^{\mathcal{B}}|X|) = a^{-1}\mathbb{E}|X|$$

En rassemblant les inégalités précédentes, on obtient

$$\sup_{\mathcal{B}} \int_{(|\mathbb{E}^{\mathcal{B}} X|>a)} |\mathbb{E}^{\mathcal{B}} X| d\mathbb{P} \leq \frac{b}{a}\mathbb{E}|X| + \int_{(|X|>b)} |X| d\mathbb{P}$$

Il suffit alors de choisir $b = \sqrt{a}$ et de faire tendre $a \to \infty$. ■

Nous avons vu que même pour une martingale positive bornée dans L^1 la convergence presque sûre n'entraîne pas la convergence dans L^1. Le théorème suivant précise les conditions pour lesquelles une martingale converge dans L^1

Théorème 17 *Soit $(X_n)_n$ une martingale intégrable, les conditions suivantes sont équivalentes*

(i) La suite $(X_n)_n$ converge dans L^1

(ii) $\sup_n \mathbb{E}|X_n| < \infty$ et $X_\infty = p.s. \lim_n X_n$ vérifie pour tout n

$$\mathbb{E}^{\mathcal{B}_n} X_\infty = X_n$$

(iii) Il existe une v.a. intégrable X telle que pour tout n

$$X_n = \mathbb{E}^{\mathcal{B}_n} X$$

(iv) La suite $(X_n)_n$ satisfait la condition d'uniforme intégrabilité

$$\sup_n \int_{(|X_n|>a)} |X_n| d\mathbb{P} \to 0 \quad \text{quand } a \to +\infty$$

Preuve. (i) $\Rightarrow$ (ii) : La convergence de X_n dans L^1 entraîne que la suite $(X_n)_n$ est bornée dans L^1 donc, d'après la section précédente, $(X_n)_n$ converge p.s. vers une v.a. X_∞. Soit $X = L^1 - \lim X_n$. D'après le lemme de Fatou,

$$\int |X_\infty - X| = \int \liminf |X_n - X| \leq \liminf \int |X_n - X| = 0$$

donc $X = X_\infty$ *p.s.* Maintenant fixons n, pour $p \geq n$ on a

$$\mathbb{E}^{\mathcal{B}_n} X_p = X_n$$

Or pour $p \to \infty$, par L^1-continuité de l'opérateur espérance conditionnelle on a

$$\mathbb{E}^{\mathcal{B}_n} X_p \to \mathbb{E}^{\mathcal{B}_n} X = \mathbb{E}^{\mathcal{B}_n} X_\infty$$

Donc

$$\mathbb{E}^{\mathcal{B}_n} X_\infty = X_n$$

(ii) $\Rightarrow$ (iii) : C'est clair car par le lemme de Fatou X_∞ est intégrable

$$\mathbb{E}|X_\infty| \leq \liminf \mathbb{E}|X_n| \leq \sup_n \mathbb{E}|X_n| < \infty$$

(iii) $\Rightarrow$ (iv) : Il suffit d'appliquer la proposition 11
(iv) $\Rightarrow$ (i) : Commençons par prouver la convergence presque sûre de X_n. La condition d'uniforme intégrabilité donne

$$\sup_n \mathbb{E}|X_n| \leq a + \sup_n \int_{(|X_n|>a)} |X_n| d\mathbb{P} < \infty$$

Donc la martingale est bornée dans L^1. Elle converge donc p.s. Le théorème de Vitali fournit alors la convergence dans L^1 ■

Remarque. Attention, si (iii) est vérifié, on a $X_n = \mathbb{E}^{\mathcal{B}_n} X$, mais a priori $X \neq X_\infty$. On montre cependant que $X_\infty = \mathbb{E}^{\mathcal{B}_\infty} X$ □

En appliquant le corollaire 3, on obtient immédiatemment.

Corollaire 4 *Soit une martingale bornée dans L^p, avec $p > 1$. Alors elle converge p.s. et dans L^1 vers la même limite. De plus la convergence a également lieu dans L^p.*

Preuve. La seule chose a montrer est la convergence dans L^p. Tout d'abord, par le lemme de Fatou $\int |X_\infty|^p \leq \liminf_n |X_n|^p < \infty$ donc $X_\infty \in L^p$. Le reste de la preuve est laissée en exercice. ■

Terminons ce cours par le théorème suivant, dont la preuve est laissée en exercice.

Théorème 18 (d'arrêt régulier) *Soit $(X_n)_n$ une martingale (resp. surmartingale, sous-martingale) uniformément intégrable, $S \leq T$ deux temps d'arrêts* quelconques, *Alors*

$$\mathbb{E}^{\mathcal{B}_S} X_T = X_S \ (resp. \leq X_S, \ resp. \geq X_S) \tag{5.3}$$

5.5 Exercices et problèmes.

1. **Somme et inf de surmartingales** Soit $(X_n)_n$ et $(Y_n)_n$ des surmartingales, montrer que $X_n + Y_n$ et $X_n \wedge Y_n$ aussi. En déduire que si $(X_n)_n$ est une martingale alors X_n^+ et X_n^- sont des sous-martingales positives.
2. **Fonction convexe de martingale.** Soit $(X_n)_n$ une martingale, et φ une fonction convexe telle que $\forall n,\ \varphi(X_n) \in L^1$. Montrer que $\varphi(X_n)$ est une sous-martingale. En déduire que $|X_n|$ et X_n^2 sont des sous-martingales positives.
3. **Réciproque du théorème d'arrêt** Soit $(X_n)_n$ un processus intégrable et adapté à une filtration. Montrer que (X_n) est une martingale ssi pour tout temps d'arrêt borné T on a
$$EX_T = EX_0$$
4. **Surmartingale conservative = martingale.** Soit $(X_n)_n$ une $\mathcal{F}_n$-surmartingale telle que
$$\forall n,\ EX_n = EX_0$$
Montrer que $(X_n)_n$ une $\mathcal{F}_n$-martingale.
5. **Compensateur d'une sous-martingale** Soit $(\mathcal{F}_n)_n$ une filtration, on dit qu'un processus $(A_n)_n$ est prévisible si $\forall n,\ A_{n+1}$ est $\mathcal{F}_n$-mesurable et A_0 est $\mathcal{F}_0$-mesurable. On dit que $(A_n)_n$ est croissant si $\forall n,\ A_n \leq A_{n+1}$. Montrer que toute sous-martinagle $(X_n)_n$ se décompose de façon unique
$$X_n = M_n + A_n$$
où $(M_n)_n$ est une martingale et $(A_n)_n$ un processus prévisible croissant de terme initial $A_0 = 0$. $(A_n)_n$ est appelé le compensateur de la sous-martingale (X_n) et cette décomposition est appelée décomposition de Doob. Dans le cas particulier où (X_n) est le carré d'une martingale (Y_n) (Cf ex 2), le compensateur de (Y_n^2) est noté $< Y_n >$.
6. **Martingales exponentielles** Soit $(Y_n)_{n\geq 1}$ une suite de v.a. indépendantes de même loi normale $\mathcal{N}(0, \sigma^2)$. On pose $\mathcal{F}_n = \sigma(Y_1, \ldots Y_n)$ et $X_n = Y_1 + \ldots + Y_n$. Soit $Z_n^u = \exp(uX_n - nu^2\sigma^2/2)$. Montrer que $(Z_n^u)_n$ est une $\mathcal{F}_n$-martingale.
7. **Martingales de carrés intégrable.** Soit $(M_n)_n$ une martingale telle que $\forall n,\ M_n \in L^2$. Montrer que
$$E(M_{n+p} - M_n)^2 = E(M_{n+p}^2 - M_n^2) \tag{5.4}$$
On suppose en outre que $\sup_n E(M_n^2) < \infty$. Montrer, simplement à l'aide de (5.4) que $(M_n^2)_n$ est une sous-martingale et que la suite $(EM_n^2)_n$ est croissante. Soit $m^* = \lim_n E(M_n^2)$. Montrer que la suite $(M_n)_n$ est convergente dans L^2.
8. **Loi du 0-1 de Kolmogorov.** Soit $(Y_n)_n$ une suite de v.a. indépendantes. On définit $\mathcal{F}_n = \sigma(Y_1, \ldots Y_n)$, $\mathcal{F}^n = \sigma(Y_n, Y_{n+1} \ldots)$, $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$, et la tribu de queue $\mathcal{F}^\infty = \bigcap_n \mathcal{F}^n$ Soit $A \in \mathcal{F}^\infty$. Montrer que $P(A) = 0$ *ou* 1. *indication* : considérer $Z_n = E^{\mathcal{F}_n} 1_A$.
En déduire que si X est une v.a. $\mathcal{F}^\infty$-mesurable, elle est constante p.s.

9. **Martingales exponentielles (suite).** Soit $(Y_n)_{n\geq 1}$ une suite de v.a. indépendantes de même loi normale $\mathcal{N}(0,\sigma^2)$. On pose $\mathcal{F}_n = \sigma(Y_1, \ldots Y_n)$ et $X_n = Y_1 + \ldots + Y_n$. Soit $Z_n^u = \exp(uX_n - nu^2\sigma^2/2)$. Montrer que $\forall u \in \mathbb{R}$, $(Z_n^u)_n$ converge p.s. vers une variable Z_∞^u finie. Que vaut cette limite ? Pour quelles valeurs de u, Z_n^u est -elle une martingale régulière ?

10. **Un jour ou l'autre** Considérer le jeu de pile ou face, i.e. une suite de v.a. $(X_n)_n$ indépendantes de même loi $P(X_i = 1) = P(X_i = -1) = 1/2$. Soit a un entier positif. Noter $S_n = \sum_{k\leq n} X_k$ et $T = \inf\{n; S_n \geq a\}$. Supposer la fortune initiale nulle $X_0 = 0$.

    Le but du problème est de montrer que $T < \infty$ p.s

    Soit $\mathcal{F}_n = \sigma(X_k,\, k \leq n)$ et

    $$Z_n^\theta = \frac{\exp \theta S_n}{\cosh \theta^n}$$

    (a) Montrer que Z_n^θ est une $\mathcal{F}_n$-martingale.

    (b) Montrer que, pour tout $\theta > 0$, $Z_{n\wedge T}^\theta$ converge p.s. et dans L^2 vers la v.a.

    $$W_\theta = \frac{\exp \theta a}{(\cosh \theta)^T} 1_{T<\infty}$$

    (c) En déduire que $P(T < \infty) = 1$ (faire tendre $\theta \to 0$)

    (d) En déduire que, pour tout θ positif

    $$E(\cosh \theta)^{-T} = \exp(-\theta a)$$

    (e) En déduire que $\mathbb{E}T = \infty$, expliquer comment calculer les $\mathbb{P}(T = k)$ à partir de $\psi(\theta) = \mathbb{E}(\cosh \theta)^{-T}$. (penser à la transformée de Laplace).

11. Soit une suite de v.a. $(X_n)_n$ indépendantes de même loi normale $\mathcal{N}(m, \sigma^2)$, avec $m < 0$. On pose $S_0 = 0$, $S_n = \sum_{k\leq n} X_k$, $\mathcal{F}_n = \sigma(X_k,\, k \leq n)$ et $W = \sup_n S_n$. Le but du problème est d'étudier W.

    (a) Montrer que $P(W < \infty) = 1$

    (b) Calculer $E^{\mathcal{F}_n}(e^{\lambda S_{n+1}})$

    (c) Montrer qu'il existe un λ_0 tel que $e^{\lambda_0 S_n}$ soit une martingale

    (d) Montrer que pour tout $a > 1$

    $$P(e^{\lambda_0 W} > a) \leq 1/a$$

    et que $P(W > t) \leq e^{-\lambda_0 t}$

    (e) En déduire que pour tout $\lambda < \lambda_0$, $E(e^{\lambda W}) < \infty$ et qu'en particulier W a des moments de tous ordres.

12. **Marche aléatoire et temps d'arrêt.** Soit $(X_n)_{n\geq 1}$ une sous-martingale et T un temps d'arrêt adaptés à une filtration $(\mathcal{B}_n)_n$ tel que $T \geq 1$ et $ET < \infty$. On suppose de plus qu'il existe $K > 0$ tel que

    $$|X_n(\omega) - X_{n-1}(\omega)| \leq K \quad p.s.$$

    (a) Montrer que $X_T = X_1 + \sum_{n\geq 1}(X_{n+1} - X_n) \cdot 1_{(T\geq n+1)}$

    (b) Montrer que $E^{\mathcal{B}_n}[(X_{n+1} - X_n) \cdot 1_{(T\geq n+1)}] \geq 0$

(c) En déduire que $E(X_T|X_1) \geq X_1$ puis que $EX_T \geq EX_1$

(d) Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes, de même loi d'espérance finie. Pour tout $n \geq 1$, on pose $S_n = X_1 + \ldots + X_n$. Montrer que $Y_n = S_n - nEX_1$ est une martingale.

(e) Soit $T \geq 1$ un temps d'arrêt adapté à la suite $(Y_n)_n$ tel que $ET < \infty$. Montrer que $ES_T = ET \cdot EX_1$

(f) On suppose que $P(X_n = 1) = P(X_n = -1) = 1/2$ et on prend $T = \inf\{n \geq 1 |\ S_n > 0\}$. Montrer par l'absurde que $ET = +\infty$

13. **Loi du log-itéré.** Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes, de même loi telle que $EX_1 = 0$ et $EX_1^2 = 1$. On suppose de plus que $\psi(\lambda) = E(e^{\lambda X_1}) < \infty$, $\forall \lambda$. Pour tout $n \geq 1$, on pose $S_n = X_1 + \ldots + X_n$. Pour tout $t > 1$, on définit $h(t) = \sqrt{2t \ln\ln t}$

(a) Montrer que $Y_n = e^{\lambda S_n}/\psi(\lambda)^n$ est une martingale

(b) Montrer en utilisant l'inégalité maximale que $\forall a > 0$ et pour tout $\forall \lambda$

$$P\left(\bigcup_{n\geq 1}(S_n > a + n\frac{\ln\psi(\lambda)}{\lambda}\right) \leq e^{-\lambda a}$$

(c) Soit $t > 1$ et $\alpha > 1$. Pour $k \geq 1$, on pose

$$a_k = \frac{\alpha}{2}h(t^k), \quad \lambda_k = \frac{h(t^k)}{t^k}, \quad c_k = \frac{\alpha}{2} + t\frac{\ln\psi(\lambda_k)}{\lambda_k^2}.$$

Démontrer que

$$P\left(\bigcup_{t^k < n \leq t^{k+1}}(S_n > h(n)c_k)\right) \leq (k\ln t)^{-\alpha}$$

(d) Montrer que p.s. ω, il existe $N \geq 1$ tel que

$$\forall k \geq N, \quad \forall n \in]t^k, t^{k+1}], \ \frac{S_n(\omega)}{h(n)} \leq c_k$$

On utilisera pour cela le *lemme de Borel-Cantelli* :
Soit $(A_k)_k$ une suite d'évènements tels que la série $\sum_k P(A_k)$ converge. Alors $P(\bigcap_N \bigcup_{k\geq N} A_k) = 0$ autrement dit $P(\bigcup_N \bigcap_{k\geq N} A_k^c) = 1$, c'est-à-dire que p.s. ω appartient à tous les A_k^c à partir d'un certain rang.

(e) En *admettant* que $\psi(\lambda) = 1 + \lambda^2/2 + o(\lambda^2)$, démontrer que

$$c_k = \frac{\alpha + t}{2} + \epsilon(k),$$

avec $\lim_{k\to\infty} \epsilon(k) = 0$

(f) En déduire que presque-sûrement

$$\limsup_{n\to\infty} \frac{S_n(\omega)}{\sqrt{2n\ln\ln n}} \leq 1$$

(g) En utilisant le théorème de dérivation sous le signe somme de Lebesgue, prouver que $\psi(\lambda) = 1 + \lambda^2/2 + o(\lambda^2)$.

14. **Ensemble de convergence d'une martingale.**

(a) Soit $(X_n)_{n\geq 0}$ une sous-martingale telle que $E\sup_{n\geq 0}(X_{n+1}-X_n)^+ < +\infty$, avec $X_0 = 0$. Soit un réel $a > 0$ et $T = \inf\{n \geq 0, \quad X_n \geq a\}$ (avec la convention $\inf(\emptyset) = +\infty$). Montrer que sur $(T = n)$ on a : $X_n \leq a + (X_n - X_{n-1})^+$

(b) En déduire que $E|X_{T\wedge n}|$ est bornée.

(c) Soit $(X_n)_{n\geq 0}$ un processus aléatoire quelconque. On pose $F = (\limsup_n X_n < +\infty)$ et $F_a = (\sup_n X_n < a)$. Montrer que $F = \bigcup_{a>0} F_a$.

(d) Soit $(X_n)_{n\geq 0}$ une martingale telle que $E\sup_{n\geq 0}|X_{n+1}-X_n|^+ < +\infty$, avec $X_0 = 0$. Soit $a > 0$.

i. Montrer que p.s. si $\omega \in F_a$ alors la suite $(X_n(\omega))$ converge vers une limite finie.

ii. En déduire que, p.s. ω, ou bien la suite $(X_n(\omega))$ converge vers une limite finie, ou bien elle oscille indéfiniment entre $-\infty$ et $+\infty$. *On pourra noter* $A_1 = \{\omega \in \Omega, \quad \lim_n X_n(\omega)$ existe et est finie.$\}$, $A_2 = \{\omega \in \Omega, \quad \limsup X_n(\omega) = +\infty$ et $\liminf X_n(\omega) = -\infty\}$.

15. **Quelques inégalités de Doob.**

(a) Soit $(X_n)_n$ une sous-martingale positive associée à une filtration $(\mathcal{B}_n)_n$ et T un temps d'arrêt. Montrer que, sur $(T \leq n)$ on a $X_T \leq E^{\mathcal{B}_T} X_n$.

(b) Soit $T_a = \inf\{n \geq 0,\ X_n > a\}$. Démontrer que

$$aP(\sup_{k\leq n} X_k > a) \leq \int_{(\sup_{k\leq n} X_k > a)} X_n \, dP. \tag{5.5}$$

Pour cela, utiliser que $(T_a \leq n) = (\sup_{k\leq n} X_k > a)$

(c) Soit $p \geq 1$ et Z une v.a. positive, montrer que

$$E(Z^p) = \int_0^{+\infty} pa^{p-1} P(Z > a)\, da. \tag{5.6}$$

On pourra remarquer que $P(Z > a) = E1_{(a<Z)}$ *et utiliser le théorème de Fubini.*

(d) Soit $(X_n)_n$ une martingale intégrable.

i. Montrer que $|X_n|$ est une sous-martingale positive.

ii. On pose $S_n = \sup_{k\leq n}|X_k|$. Soit $p > 1$. Montrer à l'aide des inégalités (5.6), (5.5) et du théorème de Fubini que

$$E(S_n^p) \leq \frac{p}{p-1} E(|X_n| S_n^{p-1})$$

iii. A l'aide de l'inégalité de Hölder, en déduire que

$$E(S_n^p)^{1/p} \leq \frac{p}{p-1} E(|X_n|^p)^{1/p}.$$

iv. On suppose que $\sup_n E(|X_n|^p) < +\infty$. Prouver que

$$||\sup_n |X_n|||_{L^p} \leq \frac{p}{p-1} \sup_n ||X_n||_{L^p}.$$

16. **Une preuve de la loi forte des grands nombres.** Soit $(X_n)_{n\geq 0}$ une martingale telle que $X_0 = 0$ et

$$M = \sup_{n\geq 0} E(X_{n+1} - X_n)^2 < +\infty.$$

(a) Montrer que $EX_n^2 = \sum_{1\leq k\leq n} E(X_n - X_{n-1})^2$

(b) En déduire que $EX_n^2 \leq M \cdot n$, puis que

$$\frac{X_n}{n} \xrightarrow{L^2} 0.$$

(c) Soit $Y_n = \sum_{1\leq k\leq n}(X_k - X_{k-1})/k$. Montrer que Y_n est une martingale.

(d) Montrer que Y_n est bornée dans L^2.

(e) En déduire que Y_n converge p.s.

(f) On *admettra* le lemme suivant : Soit une suite de nombre réels $(x_n)_n$ telle que la série $\sum_k (x_k - x_{k-1})/k$ est convergente. Alors la suite x_n/n tend vers 0. En déduire que

$$\frac{X_n}{n} \xrightarrow{p.s.} 0.$$

(g) Retrouver la loi forte des grands nombres suivante : Soit $(X_n)_n$ une suite de v.a. appartenant à L^2, indépendantes, de même espérance μ, de variances bornées par un même nombre σ^2. Alors

$$\frac{X_1 + X_2 + \ldots X_n}{n} \xrightarrow{p.s.} \mu$$

Bibliographie

[1] N. Bouleau, Processus stochastiques et applications, Hermann, 1988

[2] C. Robert, G. Casella, Monte Carlo statistical methods, Springer, 99

[3] C. Robert, Méthodes de Monte Carlo par chaîne de Marko, Economica , 96

[4] L. Devroye, Non-uniform random variable generation, Springer, 85

[5] D. Knuth, The art of computer programming, vol. 2, seminumerical algorithms, 2nd edition, Addison-Wesley, 81

[6] J. R. Norris, Markov chains, Cambridge, 1997

[7] W. Feller, An introduction to Probability Theory and Its Applications, vol. I, 3rd edition, 1957

[8] Encyclopédie Universalis,"Martingales", article de P. Crépel, J. Memin et A. Raugi

[9] P. Deheuvels, La probabilité, le hasard, la certitude, coll. Que-sais-je, PUF, 1990

[10] W. Feller, An introduction to probability theory, vol. I et II, Wiley, 1966

[11] J. Neveu, Martingales à temps discret, Masson, 1972

[12] N. Bouleau, Processus stochastiques et applications, Hermann, 1988

[13] D. Revuz, Probabilités, Hermann, 1997

[14] L. Breiman, Probability, Addison-Wesley, 1968

[15] B. Oksendal, Stochastic differential equations, Springer,1998

[16] L. Mazliak, P. Priouret, P. Baldi, Martingales et chaînes de Markov, Hermann, 1998

[17] B. Ycart, Modèles et Algorithmes markoviens, Springer, 2002