Conditionnement, Martingales

1 Espérance conditionnelle

1.1 Définition

Définition-Proposition :
Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $\mathcal{B}$ une sous-tribu de $\mathcal{A}$, et X une variable aléatoire réelle (V.A.R.) sur $(\Omega, \mathcal{A}, P)$, *intégrable*.
Alors, il existe une unique variable aléatoire Z, définie à équivalence presque-sûre près, telle que
1. Z est $\mathcal{B}$-mesurable.
2. $\forall B \in \mathcal{B}, \quad \displaystyle\int_B Z\,dP = \int_B X\,dP$

Z est appelée **espérance conditionnelle de** X **sachant** B, notée $E\left(X|\mathcal{B}\right)$.

Ainsi, la condition (2) se traduit par :

$$\forall B \in \mathcal{B}, \quad E\left(X\mathbf{1}_B\right) = E\left(E\left(X|\mathcal{B}\right)\mathbf{1}_B\right)$$

ou

$$\boxed{\text{Pour toute } Y \text{ V.A.R. bornée, } \mathcal{B}\text{-mesurable,} \quad E\left(XY\right) = E\left(\,E(X|\mathcal{B})\,Y\right)}$$

Idées de la preuve :

- *Unicité.*
  Soient Z_1 et Z_2 $\mathcal{B}$-mesurables, telles que
  $$\forall B \in \mathcal{B}, \quad \int_B Z_1\,dP = \int_B Z_2\,dP = \int_B X\,dP$$
  Alors, puisque $B_0 = \{Z_1 \geq Z_2\}$ et $B_0' = \{Z_2 \geq Z_1\}$ sont $\mathcal{B}$-mesurables et :
  $$0 = \int_{B_0} \left(Z_1 - Z_2\right)\,dP = \int_{B_0} \left(Z_1 - Z_2\right)^+\,dP$$
  $$0 = \int_{B_0'} \left(Z_2 - Z_1\right)\,dP = \int_{B_0'} \left(Z_2 - Z_1\right)^+\,dP$$
  Donc $Z_1 = Z_2$ *p.s.* d'où l'unicité.
- *Existence.*
  → On commence par montrer l'existence dans L^2 puis dans L^{1+}, et enfin L^1.
  (Idée : $L^2(\Omega, \mathcal{B}, P)$ est un sev fermé du Hilbert $L^2(\Omega, \mathcal{A}, P)$. On utilise alors la projection QX de X sur $L^2(\Omega, \mathcal{B}, P)$).
  → Autre méthode : utilisation du théorème de Radon-Nikodym. En effet, la mesure $\mu(B) = \displaystyle\int_B X dP$, $B \in \mathcal{B}$, est absolument continue par rapport à $P_{|\mathcal{B}}$. Donc il existe $Z \in L^1(\Omega, \mathcal{B}, P)$ telle que $\mu(B) = \displaystyle\int_B Z dP$ et on montre que $Z = E(X|\mathcal{B})$.

1.2 Propriétés

Soient X, Y deux V.A.R. intégrables.

Prop. 1 : $X \mapsto E(X|\mathcal{B})$ est linéaire et idempotente, *i.e.*

$$\boxed{E(aX + bY + c \mid \mathcal{B}) = aE(X|\mathcal{B}) + bE(Y|\mathcal{B}) + c \quad p.s.}$$

$$\boxed{E\left(E(X|\mathcal{B})\,|\mathcal{B}\right) = E(X|\mathcal{B})}$$

Prop. 2 : $X \mapsto E(X|\mathcal{B})$ est croissante, *i.e.*

$$\boxed{X \leq Y \implies E(X|\mathcal{B}) \leq E(Y|\mathcal{B}) \quad p.s.}$$

Prop. 3 : Si $\varphi : \mathbb{R} \to \mathbb{R}$ est convexe, telle que $\varphi(X)$ soit intégrable, alors :

$$\boxed{\varphi\left(E(X|\mathcal{B})\right) \leq E\left(\varphi(X)|\mathcal{B}\right) \quad p.s.} \qquad \textit{(Inégalité de Jensen)}$$

En particulier :

$$|E(X|\mathcal{B})| \leq E\left(|X|\,|\,\mathcal{B}\right) \quad p.s.$$
$$\left(E(X|\mathcal{B})\right)^2 \leq E\left(X^2\,|\,\mathcal{B}\right) \quad p.s.$$
$$\forall p \geq 1, \quad |E(X|\mathcal{B})|^p \leq E\left(|X|^p\,|\,\mathcal{B}\right) \quad p.s.$$

Prop. 4 : Si $\mathcal{C}$ est une sous-tribu de $\mathcal{B}$, $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}$,

$$\boxed{E\left(E(X|\mathcal{B})|\mathcal{C}\right) = E\left(X|\mathcal{C}\right)}$$

Prop. 5 : Les variables aléatoires X et $E(X|\mathcal{B})$ ont même espérance :

$$\boxed{E\left(E(X|\mathcal{B})\right) = E(X)}$$

Prop. 6 : Si X est indépendante de $\mathcal{B}$, $E(X|\mathcal{B})$ est une variable aléatoire constante *p.s.* et

$$\boxed{E(X|\mathcal{B}) = E(X) \quad p.s.}$$

Prop. 7 : Si Y est $\mathcal{B}$-mesurable et XY est intégrable,

$$\boxed{E\left(XY|\mathcal{B}\right) = Y\,E(X|\mathcal{B})}$$

1.3 Interprétation géométrique

L'espace des variables $\mathcal{B}$-mesurables est un sev de $L^2(\Omega, \mathcal{A}, P)$. $E(X|\mathcal{B})$ est en fait la projection orthogonale de la variable X sur l'espace $L^2(\Omega, \mathcal{B}, P)$ des variables $\mathcal{B}$-mesurables.

Interprétation de la Prop. 7 :
Si Y est déjà $\mathcal{B}$-mesurable, sa projection sur $L^2(\Omega, \mathcal{B}, P)$ est la variable Y elle-même.

Interprétation de la Prop. 4 :
On a $L^2(\Omega, \mathcal{C}, P) \subset L^2(\Omega, \mathcal{B}, P)$. Par conséquent, projeter d'abord sur $L^2(\Omega, \mathcal{B}, P)$ puis projeter le résultat obtenu sur $L^2(\Omega, \mathcal{C}, P)$ est équivalent à réaliser directement une projection sur $L^2(\Omega, \mathcal{C}, P)$. C'est le *Théorème des trois perpendiculaires.*

1.4 Exemples importants

1.4.1 Conditionnement par une V.A. discrète

Soit X une V.A. discrète : $X(\Omega) = \{x_k, k \in \mathbb{N}\}$.

Si $\forall k \in \mathbb{N}, P(X = x_k) > 0$, on sait que

$$P\left(A \,|\, (X = x_k)\right) = \frac{P\left(A \cap (X = x_k)\right)}{P(X = x_k)}$$

On a donc $P\left(A \,|\, (X = x_k)\right) = E\left(\mathbf{1}_A \,|\, (X = x_k)\right)$.

On a donc pour $Y \in L^1(\Omega, \mathcal{A}, P)$,

$$\boxed{E(Y|X) = \sum_{k=0}^{+\infty} \mathbf{1}_{(X=x_k)} \int_{(X=x_k)} Y \cdot \frac{1}{P(X = x_k)}\, dP}$$

En particulier, dans le cas où Y est également une V.A. discrète,

$$\boxed{\forall \omega \in \{X = x_k\}, \quad E(Y|X)(\omega) = \sum_{i=0}^{+\infty} y_i\, P(Y = y_i | X = x_k)}$$

1.4.2 Cas d'un couple à densité

Soient X, Y des V.A.R. telles que le couple (X, Y) admette une densité $f(x, y)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$. On peut alors définir une loi de probabilité conditionnelle de Y sachant que $X = x$:

$$f_Y^x(y) = \frac{f(x, y)}{f_X(x)}$$

$\left(\text{avec } f_X(x) = \displaystyle\int_{\mathbb{R}} f(x, y) dy \text{ la densité marginale de } X\right)$

On a alors, pour toute fonction $\varphi : \mathbb{R}^2 \to \mathbb{R}$ mesurable bornée,

$$\boxed{E\left(\varphi(X, Y) \,|\, X = x\right) = \int_{\mathbb{R}} \varphi(x, y)\, f_Y^x(y) dy}$$

En particulier, on a

$$\boxed{\forall \omega \in \{X = x\}, \quad E(Y|X)(\omega) = \int_{-\infty}^{+\infty} y \,.\, \frac{f(x, y)}{f_X(x)}\, dy}$$

1.5 Exercice

1. Soient X, Y, deux V.A.R. Montrer que

$$X \text{ indépendante de } Y \iff \begin{array}{c} \forall g : \mathbb{R} \to \mathbb{R} \text{ borélienne bornée,} \\ E\left(g(Y) \,|\, X\right) = E\left(g(Y)\right) \quad p.s. \end{array}$$

2. Soit (X, Y) un couple aléatoire de densité

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

Calculer la loi conditionnelle de $(Y|X = x)$. En déduire que X et $Y - X$ sont indépendantes. Par une méthode similaire, montrer également que Y et $\dfrac{X}{Y}$ sont indépendantes.

2 Martingales (à temps discret)

2.1 Définitions

Définition 1 :

Sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, on appelle **filtration** toute suite croissante $(\mathcal{F}_n)_{n\geq 0}$ de sous-tribus de $\mathcal{A}$.

Une suite de variables aléatoires réelles $(X_n)_{n\geq 0}$ définies sur $(\Omega, \mathcal{A}, \mathcal{P})$ est appelée un **processus**. On dit que le processus est **adapté** à la filtration $(\mathcal{F}_n)_{n\geq 0}$ si pour tout $n \in \mathbb{N}$, la variable aléatoire X_n est $\mathcal{F}_n$-mesurable.

En particulier, la suite de tribu $(\mathcal{A}_n)_{n\geq 0}$ définie par $\mathcal{A}_n = \sigma(X_0, X_1, \ldots, X_n)$ est appelée la **filtration naturelle** associée au processus X_n. Tout processus est bien entendu adapté à sa filtration naturelle associée.

Un processus $(X_n)_{n\geq 0}$ est dit **prévisible** si, pour tout $n \geq 1$, X_n est $\mathcal{F}_{n-1}$-mesurable.

Définition 2 :

Soit $(X_n, \mathcal{F}_n)_{n\geq 0}$ un processus adapté tel que pour tout $n \geq 0$, X_n soit intégrable.

On dit que le processus est une **martingale** si pour tout $n \geq 0$,

$$E\left(X_{n+1} \mid \mathcal{F}_n\right) = X_n$$

On dit que le processus est une **sous-martingale** si pour tout $n \geq 0$,

$$E\left(X_{n+1} \mid \mathcal{F}_n\right) \geq X_n$$

On dit que le processus est une **sur-martingale** si pour tout $n \geq 0$,

$$E\left(X_{n+1} \mid \mathcal{F}_n\right) \leq X_n$$

Quelques remarques :

- Il s'ensuit d'après la définition que si (X_n) est une martingale, alors

  $$\forall k \geq 1, \quad E\left(X_{n+k} \mid \mathcal{F}_n\right) = X_n$$

  De même, si (X_n) est une sous-martingale, alors $\forall k \geq 1, \quad E\left(X_{n+k} \mid \mathcal{F}_n\right) \geq X_n$.

  De même, si (X_n) est une sur-martingale, alors $\forall k \geq 1, \quad E\left(X_{n+k} \mid \mathcal{F}_n\right) \leq X_n$.

- Une martingale $(X_n)_{n\geq 0}$ est d'espérance constante, *i.e.*

  $$\forall n \geq 0, \quad E(X_n) = E(X_0)$$

  De même, une sous-martingale (resp. sur-martingale) est d'espérance croissante (resp. décroissante).

- Si $(X_n)_{n\geq 0}$ est une martingale (resp. sous-martingale, sur-martingale) pour une filtration $(\mathcal{F}_n)_{n\geq 0}$, alors c'est également une martingale (resp. sous-martingale, sur-martingale) pour la filtration naturelle associée aux X_n : $(\mathcal{A}_n)_{n\geq 0} = \left(\sigma(X_0, X_1, \ldots, X_n)\right)_{n\geq 0}$.

2.2 Exemples et constructions de martingales

- **Interprétation en termes de jeu.**

  Supposons que $(X_n)_{n\geq 1}$ représente la fortune d'un joueur à l'instant n. Ainsi, X_1 est sa fortune initiale, et pour tout $n \geq 1$, $Z_{n+1} = X_{n+1} - X_n$ représente son "gain" (éventuellement négatif) au n-ième coup. La tribu $\mathcal{A}_n = \sigma(X_1, \ldots, X_n) = \sigma(X_1, Z_2, \ldots, Z_n)$ renferme toute l'information jusqu'à l'instant n, et l'on a

  $$E(X_{n+1} \mid \mathcal{A}_n) - X_n = E(Z_{n+1} \mid \mathcal{A}_n)$$

  Ainsi, si $(X_n)_{n\geq 0}$ est une sous-martingale, $E(Z_{n+1} \mid \mathcal{A}_n) \geq 0$: le jeu est favorable au joueur, l'espérance du gain au n-ième coup, connaissant toute l'information apportée par ce qui s'est passé précédemment, est positive.

  Si $(X_n)_{n\geq 0}$ est une martingale, $E(Z_{n+1} \mid \mathcal{A}_n) = 0$: le jeu est équitable.

  Si $(X_n)_{n\geq 0}$ est une sur-martingale, $E(Z_{n+1} \mid \mathcal{A}_n) \geq 0$: le jeu est défavorable.

- **Construction de martingales comme espérances conditionnelles.**

  Soit Z une variable aléatoire intégrable sur $(\Omega, \mathcal{A}, P)$ et $(\mathcal{F}_n)_n$ une filtration de $\mathcal{A}$.

  Soit $X_n = E(Z \mid \mathcal{F}_n)$, $n \in \mathbb{N}$. Alors $(X_n, \mathcal{F}_n)$ est une martingale.

- **Lien avec la convexité.**

  Si $(X_n)_n$ est une martingale, $\varphi : \mathbb{R} \to \mathbb{R}$ est convexe telle que $\varphi(X_n)$ soit intégrable pour tout n, alors $(\varphi(X_n))_n$ est une sous-martingale.

  Ex : avec $\varphi(x) = |x|$ ou $\varphi(x) = x^2$.

  Le résultat reste également vrai si $(X_n)_n$ est une sous-martingale et si φ est croissante.

- **Mode additif.**

  Soient $(X_n)_{n\geq 1}$, des V.A. indépendantes sur $(\Omega, \mathcal{A}, P)$ intégrables, de moyenne M. On note toujours $A_n = \sigma(X_1, X_2, \ldots, X_n)$ et $S_n = X_1 + X_2 + \ldots + X_n$, $n \geq 1$. Alors $(S_n, \mathcal{A}_n)$ est une martingale (resp. sur-martingale, resp. sous-martingale) si $M = 0$ (resp. $M < 0$, resp. $M > 0$).

- **Mode multiplicatif.**

  Soient $(X_n)_{n\geq 1}$ des variables aléatoires indépendantes intégrables de moyenne $m > 0$. On note $P_n = \prod_{k=1}^{n} X_k$. Alors $\dfrac{P_n}{m^n}$ est une martingale.

- **Mode multiplicatif et additif.**

  Soit $a > 0$ et (X_n) des V.A. indépendantes. Soit $Y_n = A^{X_n}$. Alors

  $$\prod Y_n = a^{S_n} \text{ martingale } \iff E(a^{X_n}) = 1$$

Exercice : Soit X_n des V.A. i.i.d. : $P(X_n = 1) = p$, $P(X_n = -1) = 1 - p$.

On définit $S_n = \sum X_n$. Déterminer $a > 0$ pour que a^{S_n} soit une martingale.

2.3 Décomposition de Doob

Théorème : *Soit $X = (X_n)_{n\geq 0}$ une sous-martingale. Alors il existe une martingale $M = (M_n)_{n\geq 0}$ avec $M_0 = 0$ et un processus $A = (A_n)_{n\geq 0}$ prévisible et croissant avec $A_0 = 0$, tels que*

$$X_n = X_0 + M_n + A_n$$

De plus, cette décomposition, appelée ***Décomposition de Doob****, est unique p.s..*

Dans le cas où X est une martingale, la décomposition est évidente. Si X est une sur-martingale, alors il existera de la même manière une unique (*p.s.*) décomposition

$$X_n = X_0 + M_n - A_n$$

avec (M_n) et (A_n) définis comme dans le théorème précédent.

Preuve : On pose $A_0 = 0$ et

$$A_n = \sum_{k=1}^{n} E(X_k - X_{k-1} \mid \mathcal{F}_{k-1})$$

X étant une sous-martingale, on a $E(X_k - X_{k-1} \mid \mathcal{F}_{k-1}) \geq 0$ *p.s.*. donc $A_{k+1} \geq A_k$ *p.s.*, tandis que par construction A_{k+1} est $\mathcal{F}_k$-mesurable. Par ailleurs, on a

$$E(X_n \mid \mathcal{F}_{n-1}) - X_{n-1} = E(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = A_n - A_{n-1}$$

et donc

$$E(X_n \mid \mathcal{F}_{n-1}) - A_n = X_{n-1} - A_{n-1}$$

mais $A_n \in \mathcal{F}_{n-1}$, donc

$$E(X_n - A_n \mid \mathcal{F}_{n-1}) = X_{n-1} - A_{n-1}$$

Si on pose $M_n = X_n - A_n - A_0$, il suit que $M = (M_n)_{n\geq 0}$ est une martingale avec $M_0 = 0$, et on a la décomposition souhaitée.

Quant à l'unicité, supposons que

$$X_n = X_0 + M_n + A_n = X_0 + L_n + C_n$$

soient deux décompositions. En soustrayant l'une de l'autre, on arrive à

$$L_n - M_n = A_n - C_n$$

Comme A_n et C_n sont $\mathcal{F}_{n-1}$-mesurables, il en est de même de $L_n - M_n$, et donc

$$L_n - M_n = E\left(L_n - M_n \mid \mathcal{F}_{n-1}\right) = L_{n-1} - M_{n-1} = A_{n-1} - C_{n-1} \text{ } p.s.$$

Par une récurrence évidente, on obtient alors que $L_n = M_n = L_0 - M_0 = 0$ *p.s.* (car $L_0 = M_0 = 0$). Par suite, $L_n = M_n$ *p.s.*, donc aussi $A_n = C_n$ *p.s.*, et l'unicité est démontrée. □

Interprétation de la décompositon de Doob.

Si (X_n) représente l'évolution du processus au cours du temps, si à la date n on estime la valeur de X_{n+1}, on a une partie "sûre", dans le sens où A_{n+1} est connu dès la date n (puisque A est prévisible), et une partie aléatoire centrée qui correspond au M_{n+1} (une perturbation en quelque sorte).

2.4 Temps d'arrêt

Définition :
Une variable aléatoire $T : \Omega \to \overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$ est appelée un **temps d'arrêt** adapté à la filtration $(\mathcal{F}_n)_n$ si pour chaque n, $\{T \leq n\} \in \mathcal{F}_n$.

Toute constante entière, ou égale à $+\infty$, est évidemment un temps d'arrêt.

$\mathbb{N}$ est donc un ensemble de "temps" tandis que $\mathcal{F}_n$ représente la famille des événements qu'on peut observer avant ou à l'instant n. On peut donc imaginer un temps d'arrêt comme le premier instant où une certaine suite de V.A. (X_n) avec chaque X_n observable à l'instant n (*i.e.* $\mathcal{F}_n$-mesurable), vérifie une certaine propriété, avec la convention que le temps d'arrêt vaut $+\infty$ si cela n'arrive jamais.

Par exemple, si (X_n) est une martingale, et T est le premier temps auquel elle atteint ou dépasse $N \in \mathbb{N}$ fixé. On peut écrire

$$T = \begin{cases} \inf\limits_{n \geq 0}\{n \ / \ X_n \geq N\} & \text{si } X_n \geq N \text{ pour un } n \in \mathbb{N} \\ +\infty & \text{sinon} \end{cases}$$

La terminologie "temps d'arrêt" vient de la modélisation des jeux de hasard : un joueur peut arrêter de jouer à un temps aléatoire (dépendant de ses pertes antérieures par exemple), mais uniquement sur la base des événements effectivement observables avant l'instant ou à l'instant où il s'arrête.

Si T est un temps d'arrêt <u>fini</u>, on note X_T la variable

$$X_T(\omega) = \sum_{n=0}^{+\infty} X_n(\omega) \mathbf{1}_{T(\omega)=n}$$

Si T est un temps d'arrêt (non nécessairement fini), on définit la **tribu antérieure à** T comme la tribu :

$$\mathcal{F}_T = \{A \in \mathcal{A} \ / \ \forall n, A \cap (T \leq n) \in \mathcal{F}_n\}$$

Propriétés :

- Si T est un temps d'arrêt borné et si $(X_n)_n$ est une martingale, on a

$$E(X_T) = E(X_0)$$

  De même, si $(X_n)_n$ est une sur-martingale, on a $E(X_T) \geq E(X_0)$, et si $(X_n)_n$ est une sous-martingale, on a $E(X_T) \leq E(X_0)$.

- Si S et T sont des temps d'arrêt tels que $S \leq T$, on a $\mathcal{F}_S \subset \mathcal{F}_T$.

- X_T est $\mathcal{F}_T$-mesurable.

Théorème d'arrêt de Doob :
Soit $(X_n)_{n\geq 0}$ une martingale. Soit S et T deux temps d'arrêt bornés, avec $S \leq T$ *p.s.*. On a alors

$$E(X_T \mid \mathcal{F}_S) = X_S \quad p.s.$$

Autrement dit, si (X_n) une martingale (resp. sur-martingale) par rapport à $(\mathcal{F}_n)$, et T un temps d'arrêt de $(\mathcal{F}_n)_n$, la suite arrêtée à T, $Y_n = Y_{min(T,n)}$ définie par $Y_n = X_{min(k,n)}$ sur $\{T = k\}$ est une martingale (resp. une sur-martingale).

On a alors une sorte de réciproque à la première des propriétés des temps d'arrêt :

Si $(X_n)_n$ est un processus, X_n intégrables et $\mathcal{F}_n$-mesurables. Si $E(X_T) = E(X_0)$ pour tout temps d'arrêt T borné, alors (X_n) est une martingale.

Corollaire : Identité de Doob

Si $(X_n, \mathcal{F}_n)$ est une sous-martingale, pour tout $t > 0$,

$$P\left(\max_{1\leq n\leq k} X_n \geq t\right) \leq \frac{E(X_k^+)}{t}$$