

I Espace Conditionnelle

1 Rappels sur l'espace L^2 .

On rappelle que $L^2(\cdot, F, P)$ (que l'on note plus brièvement $L^2(F)$) est la completion par la norme $E[|\cdot|^2]$ de l'ensemble des fonctions mesurables. Les éléments de $L^2(F)$ sont les classes d'équivalence des fonctions mesurables selon la relation $X \sim Y$ ssi $P(X = Y) = 1$.

Théorème 1. $L^2(\cdot, F, P)$ est un espace vectoriel complet.

Démonstration. Vérifions que il est bien un espace vectoriel. On va montrer que il est complet pour la topologie induite par la norme. Soit $X_n \in L^2(F)$ une suite de Cauchy: $\sup_{m, k > n} E[X_m - X_k]^2 \rightarrow 0$ pour $k \rightarrow \infty$. On veut montrer qu'il existe $X \in L^2(F)$ (unique p.s.) tel que $\lim_n X_n = X$ p.s. Soit k_n une suite croissante tel que $k_n \rightarrow \infty$ et $E[X_{k_n} - X_m]^2 \leq 2^{-n}$ pour tout $s, m \geq k_n$. Alors

$$E\left[\sum_{n=1}^{\infty} |X_{k_{n+1}} - X_{k_n}|^2\right] \leq \sum_{n=1}^{\infty} E[X_{k_{n+1}} - X_{k_n}]^2 < \infty$$

et donc pour presque tout ω la série $S(\omega) = \sum_{n=1}^{\infty} (X_{k_{n+1}}(\omega) - X_{k_n}(\omega))$ est absolument convergente et donc $\lim_n X_{k_n}(\omega)$ existe p.s.. Soit $X(\omega) = \limsup_n X_{k_n}(\omega)$, on a que $X \in F$ et que $X_{k_n} \rightarrow X$ p.s. Maintenant on observe que si $E[|X_r - X_{k_l}|^2] \leq 2^{-2n}$ pour tout $r \geq k_n$, donc une application du Lemme de Fatou donne

$$E[|X_r - X|^2] = E\left[\liminf_l |X_r - X_{k_l}|^2\right] \leq \liminf_l E[|X_r - X_{k_l}|^2] \leq 2^{-2n}$$

pour tout $r \geq k_n$, qui montre que $X \in L^2(F)$ et que $X_n \rightarrow X$ dans $L^2(F)$.

Corollaire 2. Si $B \subset F$ est une sous-tribu de F alors $L^2(B)$ est un sous-espace vectoriel fermé de $L^2(F)$ et pour tout $X \in L^2(F)$ il existe une v.a. $Y \in L^2(B)$ (unique p.s.) qui satisfait une des deux propriétés équivalentes suivantes:

- a) $E[|X - Y|^2] = \inf_{Z \in L^2(B)} E[|X - Z|^2]$;
- b) $X - Y \perp L^2(B)$.

On appelle Y la projection orthogonale de X sur $L^2(B)$.

Démonstration. Par le théorème 1 l'ensemble $L^2(B)$ est complet par la norme L^2 et donc fermé dans $L^2(F)$. Soit $\|X\| = \inf_{Z \in L^2(B)} E[|X - Z|^2]$ et Y_n une suite minimisante: $E[|X - Y_n|^2] \rightarrow \|X\|$ quand $n \rightarrow \infty$. On a donc

$$E[|X - Y_n|^2] + E[|X - Y_m|^2] = 2E[|X - (Y_n + Y_m)/2|^2] + E[|Y_n - Y_m|^2]/2$$

(on utilise $E[|A + B|^2] = 2E[A^2] + 2E[B^2]$). Mais $(Y_n + Y_m)/2 \in L^2(B)$ ce qui donne que

$$E[|Y_n - Y_m|^2]/2 \leq E[|X - Y_n|^2] + E[|X - Y_m|^2] \rightarrow 0$$

pour $n, m \rightarrow \infty$. Donc la suite Y_n est Cauchy. Soit $Y = \lim_n Y_n \in L^2(B)$ (dans L^2). On a que $E[|X - Y|^2] \leq E[|X - Y_n|^2] + E[|Y_n - Y|^2] \rightarrow 0$ car $E[|Y_n - Y|^2] \rightarrow 0$.

Pour tout $t \in \mathbb{R}$ et $Z \in L^2(B)$ on a que $Y + tZ \in L^2(B)$ et

$$0 \leq E[|X - Y - tZ|^2] = E[|X - Y|^2] - 2tE[(X - Y)Z] + t^2E[Z^2].$$

Le polynôme $P(t) = at^2 + bt$ satisfait $P(t) = 0$ pour tout $t = 0$ donc on doit avoir $b = 0$ ou dans notre cas $E[(X \otimes Y)Z] = 0$ pour tout Z . L'implication réciproque est facile à établir. Pour montrer l'unicité presque sûre de Y on suppose que Y est une autre projection orthogonale. On a que $E[(Y \otimes Y)Z] = 0$ pour tout $Z \in L^2(G)$ et donc aussi pour $Z = Y \otimes Y$ mais alors $E[(Y \otimes Y)^2] = 0 \implies Y \otimes Y = 0$ p.s.

2 L'espérance conditionnelle

Soit (Ω, \mathcal{F}, P) un espace de probabilité et soit $\mathcal{B} \subset \mathcal{F}$ une sous-tribu de \mathcal{F} . Soit $X: \Omega \rightarrow \mathbb{R}$ une variable aléatoire telle que $E(|X|) < +\infty$ (i.e. $X \in L^1(\Omega, \mathcal{F}, P)$).

Définition 3. L'espérance conditionnelle de X sachant \mathcal{B} est une variable aléatoire $\hat{X}^{\mathcal{B}}$ telle que

$$E[1_A X] = E[1_A \hat{X}^{\mathcal{B}}] \quad \forall A \in \mathcal{B} \quad (1)$$

L'assertion (1) est en fait équivalente à

$$E[Z X] = E[Z \hat{X}^{\mathcal{B}}] \quad \forall Z \in \mathcal{B} \text{ bornée} \quad (2)$$

L'existence d'une variable aléatoire Y qui a ces propriétés n'est pas triviale, on va y revenir plus avant. Par ailleurs, cette variable aléatoire est unique (à égalité presque-sûre près (voir la preuve du 2. de la proposition suivante)).

On utilisera les notations $Y = E(X/\mathcal{B})$, ainsi que $E(X/Z) = E(X/(Z))$. La probabilité conditionnelle $P(\cdot/\mathcal{B})$ sachant \mathcal{B} (ou par rapport à \mathcal{B}) est définie par $P(A/\mathcal{B}) = E[1_A/\mathcal{B}]$. On remarque que $P(A/\mathcal{B})$ est une variable aléatoire.

Exercice 1. Soient $G \subset \mathcal{F}$, $X \in L^2(F)$, $Z \in L^2(G)$ et $Y = E[X/G]$, montrer que

$$E[|X \otimes Z|^2] = E[|X \otimes Y|^2] + E[|Y \otimes Z|^2]$$

et en déduire que

$$E[|X \otimes Y|^2] = \inf_{Z \in L^2(G)} E[|X \otimes Z|^2].$$

L'exercice précédent montre que l'espérance conditionnelle dans $L^2(F)$ est le meilleur estimateur G -mesurable de X selon le risque quadratique:

$$E[|X \otimes E[X/\mathcal{B}]|^2] \leq E[|X \otimes Z|^2] \quad \text{pour tout } Z \in L^2(\mathcal{B}).$$

En fait cette interprétation géométrique est la base d'une stratégie pour montrer l'existence de l'espérance conditionnelle.

Soit $X \in L^2(F)$ et \mathcal{B} une sous-tribu de \mathcal{F} . Alors la projection orthogonale Y de X sur $L^2(\mathcal{B})$ satisfait $E[XZ] = E[YZ]$ pour tout $Z \in L^2(\mathcal{B})$ et donc pour tout Z \mathcal{B} -mesurable et bornée. Donc $Y = E[X/\mathcal{B}]$ ce qui montre l'existence de l'espérance conditionnelle pour $X \in L^2(F)$.

Théorème 4. Pour tout $X \in L^1(F)$ il existe l'espérance conditionnelle $E[X/\mathcal{B}] \in L^1(\mathcal{B})$.

Démonstration. Pour prouver l'existence pour tout v.a. $X \in L^1(F)$ on procède par approximation. Soit $X \geq 0$ et dans L^1 . Soit $X_n(\omega) = \min(X(\omega), n)$ et Y_n la projection orthogonale correspondante sur $L^2(\mathcal{B})$. Alors pour $n < m$ on a que $0 \leq E[1_A(X_n \otimes X_m)] = E[1_A(Y_n \otimes Y_m)]$ pour tout $A \in \mathcal{B}$ ce qui implique que $Y_n \leq Y_m$ p.s. (voir 2.6) et qu'il existe un ensemble de mesure nulle $N \subset \Omega$ en dehors duquel la suite $\{Y_n(\omega)\}_n$ est croissante pour tout $\omega \in N^c$. Soit $Y = \sup_n Y_n$. On a que $E[1_A Y] = \sup_n E[1_A Y_n] = \sup_n E[1_A X_n] = E[1_A X]$ par convergence monotone et donc que $Y \in L^1(\mathcal{B})$ et que $Y = E[X/\mathcal{B}]$. Pour une g.v. $X \in L^1$ soit $X = X_+ \otimes X_-$ avec $X_+, X_- \geq 0$ et dans L^1 . On pose $Y_+ = E[X_+/\mathcal{B}]$ et $Y_- = E[X_-/\mathcal{B}]$. On obtient que $Y \in L^1(\mathcal{B})$ et que $E[1_A X] = E[1_A Y]$ pour tout $A \in \mathcal{B}$.

Proposition 5.

1. $E(X/B) \in L^1(\mathcal{B}, P)$.
2. Soient X, Y deux espérances conditionnelles de X sachant B , alors $Y = X$ p.s.. En particulier si $X \in \mathcal{B}$ alors $E(X/B) = X$ p.s.

Démonstration. (Voir le poly du cours de processus discrets)

Si on conditionne par rapport à une v.a. X donné on trouve bien que la probabilité conditionnelle est une fonction des valeurs de X :

Proposition 6. Il existe une fonction mesurable h_Z telle que $E[Z/X] = h_Z(X)$ p.s.

Démonstration. La v.a. $E[Z/X]$ est (X) -mesurable. On utilise donc le théorème 1.

Proposition 7. Pour tout $X, Y \in L^1(F)$ et tout sous-tribu $G, H \subset F$ on a les propriétés suivantes:

1. Linéarité: $E[X + Y/G] = E[X/G] + E[Y/G]$;
2. Positivité: $X \geq 0$ p.s. $\Rightarrow E[X/G] \geq 0$ p.s.
3. Convergence monotone: $0 \leq X_n \leq X$ p.s. $\Rightarrow E[X_n/G] \leq E[X/G]$ p.s.
4. Inégalité de Jensen: ϕ convexe et $(X) \subset L^1: E[\phi(X)/G] \geq \phi(E[X/G])$
5. Contractivité dans L^p : $E[X/G]_p \leq X_p$.
6. Emboîtement: Si H est une sous-tribu de G alors $E[E[X/G]/H] = E[X/H] = E[E[X/H]/G]$.
7. Si $Z \in G$, $E[X/Z] < +\infty$ et $E[Z/G] < +\infty$ alors $E[XZ/G] = Z E[X/G]$.

Démonstration.

1. Exercice.
2. on remarque que si $E[X/G] < 0$ sur $A \in G$ tel que $P(A) > 0$ alors $0 < E[X1_A] = E[E[X/G]1_A] = P(A) < 0$ ce qui est impossible.
3. Soit $Y_n = E[X_n/G]$. Par la positivité de l'esp. cond. on a que Y_n est une suite croissante. Soit $Y = \limsup_n Y_n$ alors $Y \in G$ et le théorème de convergence monotone nous permet de passer à la limite dans l'égalité $E[X_n 1_A] = E[Y_n 1_A]$ pour obtenir que $E[X 1_A] = E[Y 1_A]$ pour tout $A \in G$. Donc $Y = E[X/G]$ p.s.
4. Tout fonction convexe peut s'écrire dans la forme $\phi(x) = \sup_{n \geq 1} (a_n x + b_n)$ pour une suite dénombrable des couples $(a_n, b_n) \in \mathbb{R}^2$. Donc $E[\phi(X)/G] \geq a_n E[X/G] + b_n$ et on peut conclure.
5. On utilise la propriété 4). exercice.
6. Exercice.
7. Admis. (Facile pour des fonctions ≥ 0 , utiliser des limites monotones dans le cas $X, Z \geq 0$ et conclure).

On rappelle que un σ -système I sur C est une famille de parties de C stable pour intersection finie. Et que si deux mesures μ, ν de $\mathcal{M}_+(I)$ coïncident sur I alors elles sont égales.

Proposition 8. Si H et G sont indépendantes X est G -mesurable et $G \subset H$, alors

$$E[X/H, G] = E[X/G].$$

Démonstration. On suppose que $X \geq 0$ et L^1 . Soit $G \subset G$ et $H \subset H$. Par hypothèse $X 1_G \wedge G$ et $1_H \wedge H$ sont indépendantes, donc $E[X 1_G 1_H] = E[X 1_G] E[1_H]$ et si on note $Y = E[X/G]$ on a aussi que, $E[Y 1_G 1_H] = E[Y 1_G] E[1_H]$ ce qui nous dit que $E[X 1_G 1_H] = E[Y 1_G 1_H]$ donc pour les mesures

$$\mu_X(F) = E[X 1_F] \text{ et } \mu_Y(F) = E[Y 1_F]$$

de sur (G, H) ont la même masse et vérifient $\mathbb{E}_X(G \cdot H) = \mathbb{E}_G(G \cdot H)$ pour tout $G \in \mathcal{G}$ et $H \in \mathcal{H}$. Mais la classe des éléments de la forme $G \cdot H$ est un π -système et donc les mesures sont égales sur tout (G, H) .

Proposition 9. (Conditionnement et indépendance) Soit X_1, \dots, X_n une famille des v.a. indépendantes et $f(X_1, \dots, X_n) \in L^1$ alors

$$\mathbb{E}[f(X_1, \dots, X_n) | X_1] = f(X_1)$$

où $\mathbb{E}(X) = \mathbb{E}[f(X, X_2, \dots, X_n)]$.

Démonstration. Utiliser le théorème de Fubini sur la loi jointe de X_1, \dots, X_n .

Exercice 2. (Processus de branchement) Soit $\{X_{m,r} : m, r \in \mathbb{N}\}$ une double suite des v.a. iid. discrètes et valeurs ≥ 0 . On pose $Z_0 = 1$ et $Z_n = X_{n,1} + \dots + X_{n,Z_{n-1}}$ pour $n \geq 1$. Montrer que la fonction génératrice $f_n(\cdot) = \mathbb{E}[Z_n]$ pour tout $\cdot \in [0, 1]$ satisfait

$$f_0(\cdot) = 1 \quad f_n = f_{n-1}(f(\cdot)) \quad \text{pour } n \geq 1.$$

Solution.

$$f_n(\cdot) = \mathbb{E}[\mathbb{E}[Z_n | Z_{n-1}]]$$

Or

$$\mathbb{E}[Z_n | Z_{n-1} = k] = \mathbb{E}[X_{n,1} + \dots + X_{n,k} | Z_{n-1} = k] = \mathbb{E}[X_{n,1} + \dots + X_{n,k} | Z_{n-1} = k]$$

et

$$\mathbb{E}[X_{n,1} + \dots + X_{n,k} | Z_{n-1} = k] = \mathbb{E}[X_{n,1} + \dots + X_{n,k}] = (k \mathbb{E}[X_{1,1}]) = f(\cdot)^k$$

car $\{X_{n,k} : k \geq 1\}$ est indépendant de $Z_{n-1} = (X_{m,k} : 1 \leq m \leq n-1, k \geq 1)$. Donc

$$\mathbb{E}[Z_n | Z_{n-1} = k] = f(\cdot)^k |_{Z_{n-1} = k} = f(\cdot)^{Z_{n-1}}$$

ce qui nous permet de conclure que

$$f_n(\cdot) = \mathbb{E}[f(\cdot)^{Z_{n-1}}] = f_{n-1}(f(\cdot)).$$

Exercice 3. Soit $\mathbb{E}[Y|G] = X$ et $\mathbb{E}[X^2] = \mathbb{E}[Y^2]$ en déduire que $X = Y$ a.s.

Exercice 4. Prouver une inégalité de Chebishev conditionnelle.

Exercice 5. Prouver l'inégalité de Cauchy-Schwartz conditionnelle

$$\mathbb{E}[XY|G]^2 \leq \mathbb{E}[X^2|G] \mathbb{E}[Y^2|G].$$

Exercice 6. Soit (X_0, X_1, \dots, X_n) un vecteur Gaussien de moyenne nulle et matrice de covariance $\Sigma = (\sigma_{ij})_{i,j=1,\dots,n}$. Montrer que

$$\mathbb{E}[X_0 | X_1, \dots, X_n] = \sum_{i=1}^n \sigma_{0i} X_i \quad p.s.$$

et déterminer les poids σ_{0i} en fonction de Σ .

Exercice 7. Donner un exemple avec $\Sigma = \{a, b, c\}$ pour montrer que, en général,

$$\mathbb{E}[\mathbb{E}[X|F_1]|F_2] \neq \mathbb{E}[\mathbb{E}[X|F_2]|F_1].$$

Exercice 8. Montrer les implications suivantes

$$X, Y \text{ indépendantes} \iff \mathbb{E}[X|Y] = \mathbb{E}[X] \iff \mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$$

et trouver des v.a. $X, Y \in \{a, b, 0, 1\}$ pour montrer que les implications inverses sont fausses.