[M. Gubinelli - Contrôle des chaînes de Markov - M1 MMD 2010/2011 - 20100923 - poly 1 - v.2]

I Espérance Conditionnelle

1 Rappels sur l'espace L^2.

On rappelle que $L^2(\Omega, \mathcal{F}, \mathbb{P})$ (que l'on dénote plus brièvement $L^2(\mathcal{F})$) est la completion par la norme $\| \cdot \|_2 = (\mathbb{E}[| \cdot |^2])^{1/2}$ de l'ensemble des fonction étagées. Les elements de $L^2(\mathcal{F})$ sont des classes d'equivalence des fonctions mesurables selon la relation $X \sim Y \Leftrightarrow \mathbb{P}(X = Y) = 1$.

Théorème 1. $L^2(\Omega, \mathcal{F}, \mathbb{P})$ *est un espace vectoriel complet.*

Démonstration. Verifier que il est bien un espace vectoriel. On va montrer que il est complet pour la topologie induite par la norme. Soit $X_n \in L^2(\mathcal{F})$ une suite de Cauchy: $\sup_{m,k>n} \|X_m - X_k\|_2 \to 0$ pour $k \to \infty$. On veut montrer qu'il existe $X \in L^2(\mathcal{F})$ (unique p.s.) tel que $\lim \|X_n - X\|_2 = 0$. Soit k_n une suite croissante tel que $k_n \nearrow \infty$ et $\|X_s - X_m\|_2 \leqslant 2^{-n}$ pour tout $s, m \geqslant k_n$. Alors

$$\mathbb{E} \sum_{n \geqslant 1} |X_{k_{n+1}} - X_{k_n}| \leqslant \sum_{n \geqslant 1} \|X_{k_{n+1}} - X_{k_n}\|_2 < \infty$$

et donc pour presque tout $\omega \in \Omega$ la série $S(\omega) = \sum_{n \geqslant 1} (X_{k_{n+1}}(\omega) - X_{k_n}(\omega))$ est absolument convergente et donc $\lim_n X_{k_n}(\omega)$ existe p.s.. Soit $X(\omega) = \text{limsup}_n X_{k_n}(\omega)$, on a que $X \hat{\in} \mathcal{F}$ et que $X_{k_n} \to X$ p.s. Maintenant on observe que si $l \geqslant n$ $\mathbb{E}[|X_r - X_{k_l}|^2] \leqslant 2^{-2n}$ pour tout $r \geqslant k_n$, donc une application du Lemme de Fatou donne

$$\mathbb{E}[|X_r - X|^2] = \mathbb{E}[\lim_{l \to \infty} |X_r - X_{k_l}|^2] \leqslant \lim_{l \to \infty} \mathbb{E}[|X_r - X_{k_l}|^2] \leqslant 2^{-2n}$$

pour tout $r \geqslant k_n$, qui montre que $X \in L^2(\mathcal{F})$ et que $X_n \to X$ dans $L^2(\mathcal{F})$. $\square$

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

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

b) $X - Y \perp L^2(\mathcal{B})$.

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

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

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

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

$$\mathbb{E}[|Y_n - Y_m|^2]/2 \leqslant \mathbb{E}[|X - Y_n|^2] + \mathbb{E}[|X - Y_m|^2] - 2\Delta \to 0$$

pour $n, m \to \infty$. Donc la suite Y_n est Cauchy. Soit $Y = \lim_n Y_n \in L^2(\mathcal{B})$ (dans L^2). On a que $\|X - Y\|_2 \leqslant \|X - Y_n\|_2 + \|Y_n - Y\|_2$ et donc que $\|X - Y\|_2 = \sqrt{\Delta}$ car $\|Y_n - Y\|_2 \to 0$.

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

$$0 \leqslant \mathbb{E}[|X - Y - t Z|^2] - \mathbb{E}[|X - Y|^2] = -2 t\mathbb{E}[(X - Y)Z] + t^2\mathbb{E}[Z^2].$$

Le polynome $P(t) = a t^2 + b\, t$ satisfait $P(t) \geqslant 0$ pour tout $t \geqslant 0$ donc on doit avoir $b = 0$ ou dans notre cas $\mathbb{E}[(X - Y)Z] = 0$ 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 $\mathbb{E}[(Y - Y')Z] = 0$ pour tout $Z \in L^2(\mathcal{G})$ et donc aussi pour $Z = Y - Y'$ mais alors $\mathbb{E}[(Y - Y')^2] = 0 \Rightarrow Y - Y' = 0$ p.s. □

2 L'espérance conditionnelle

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

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

$$\mathbb{E}[1_A X] = \mathbb{E}[1_A Y] \qquad \forall A \in \mathcal{B} \tag{1}$$

L'assertion (1) est en fait équivalente à

$$\mathbb{E}[Z X] = \mathbb{E}[Z Y] \qquad \forall Z \hat{\in} \mathcal{B} \,\text{bornée} \tag{2}$$

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 l'égalité presque-sûre près (voir la preuve du 2. de la proposition suivante).

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

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

$$\mathbb{E}[|X - Z|^2] = \mathbb{E}[|X - Y|^2] + \mathbb{E}[|Y - Z|^2]$$

et en déduire que

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

L'exercice précèdent montre que l'espérance conditionnelle dans $L^2(\mathcal{F})$ est le meilleure estimateur $\mathcal{G}$-mesurable de X selon le risque quadratique:

$$\mathbb{E}[|X - \mathbb{E}[X|\mathcal{B}]|^2] \leqslant \mathbb{E}[|X - Z|^2] \quad \text{pour tout } Z \in L^2(\mathcal{B}).$$

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

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

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

Démonstration. Pour étendre l'existence à tout v.a. $X \in L^1(\mathcal{F})$ on procède par approximation. Soit $X \geqslant 0$ 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 $n \geqslant m$ on a que $0 \leqslant \mathbb{E}[1_A(X_n - X_m)] = \mathbb{E}[1_A(Y_n - Y_m)]$ pour tout $A \in \mathcal{B}$ ce qu'implique que $Y_n \geqslant Y_m$ p.s. (vérifier) et qu'il existe un ensemble de mesure nulle $N \in \mathcal{B}$ 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 $\mathbb{E}[1_A Y] = \sup_n \mathbb{E}[1_A Y_n] = \sup_n \mathbb{E}[1_A X_n] = \mathbb{E}[1_A X]$ par convergence monotone et donc que $Y \in L^1(\mathcal{B})$ et que $Y = \mathbb{E}[X|\mathcal{B}]$. Pour une générique $X \in L^1$ soit $X = X_+ - X_-$ avec $X_+, X_- \geqslant 0$ et dans L^1. On pose $Y_\pm = \mathbb{E}[X_\pm|\mathcal{B}]$ et $Y = Y_+ - Y_-$. On obtient que $Y \in L^1(\mathcal{B})$ et que $\mathbb{E}[1_A X] = \mathbb{E}[1_A Y]$ pour tout $A \in \mathcal{B}$. □

Proposition 5.

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

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

Si on conditionne par rapport à une v.a. X donnée 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 $\mathbb{E}[Z|X] = h_Z(X)$ p.s.*

Démonstration. La v.a. $\mathbb{E}[Z|X]$ est $\sigma(X)$-mesurable. Donc il existe une fonction Borelienne $h_Z: \mathbb{R} \to \mathbb{R}$ telle que $\mathbb{E}[Z|X](\omega) = h_Z(X(\omega))$ pour presque tout $\omega \in \Omega$. □

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

1. *Linéarité:* $\mathbb{E}[X + Y|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}]$;
2. *Positivité:* $X \geqslant 0\, p.s. \Rightarrow \mathbb{E}[X|\mathcal{G}] \geqslant 0\, p.s.$
3. *Convergence monotone:* $0 \leqslant X_n \nearrow X\, p.s. \Rightarrow \mathbb{E}[X_n|\mathcal{G}] \nearrow \mathbb{E}[X|\mathcal{G}]\, p.s.$
4. *Inégalité de Jensen: si φ est convexe et $\varphi(X) \in L^1$:* $\mathbb{E}[\varphi(X)|\mathcal{G}] \leqslant \varphi(\mathbb{E}[X|\mathcal{G}])$
5. *Contractivité dans L^p:* $\|\mathbb{E}[X|\mathcal{G}]\|_p \leqslant \|X\|_p$.
6. *Emboîtement: Si $\mathcal{H}$ est une sous-tribu de $\mathcal{G}$ alors* $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}]$.
7. *Si $Z \hat{\in} \mathcal{G}$, $\mathbb{E}[|X|] < \infty$ et $\mathbb{E}[|X Z|] < +\infty$ alors* $\mathbb{E}[X Z|\mathcal{G}] = Z\, \mathbb{E}[X|\mathcal{G}]$.

Démonstration.

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

On rappelle que un π-système $\mathcal{I}$ sur $\mathcal{C}$ est une famille de parties de $\mathcal{C}$ stable pour intersection finie. Et que si deux mesures μ, ν defines sur $\sigma(\mathcal{I})$ coincident sur $\mathcal{I}$ alors elle sont egales.

Proposition 8. *Si $\mathcal{H}$ et $\mathcal{G}$ sont indépendantes, X est $\mathcal{G}$-mesurable et $\mathcal{G}' \subseteq \mathcal{G}$, alors*

$$\mathbb{E}[X|\mathcal{H}, \mathcal{G}'] = \mathbb{E}[X|\mathcal{G}'].$$

Démonstration. On supposer que $X \geqslant 0$ et L^1. Soit $G \in \mathcal{G}'$ et $H \in \mathcal{H}$. Par hypothese $X 1_G \hat{\in} \mathcal{G}$ et $1_H \in \mathcal{H}$ sont indépendantes, donc $\mathbb{E}[X 1_G 1_H] = \mathbb{E}[X 1_G]\mathbb{E}[1_H]$ et si on note $Y = \mathbb{E}[X|\mathcal{G}']$ on a aussi que, $\mathbb{E}[Y 1_G 1_H] = \mathbb{E}[Y 1_G]\mathbb{E}[1_H]$ ce qui nous dit que $\mathbb{E}[X 1_G 1_H] = \mathbb{E}[Y 1_G 1_H]$ donc pour les mesures

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

définie sur $\sigma(\mathcal{G}', \mathcal{H})$ ont la même masse et vérifient $\mu_X(G \cap H) = \mu_Y(G \cap H)$ pour tout $G \in \mathcal{G}'$ et $H \in \mathcal{H}$. Mais la classe des événements de la forme $G \cap H$ est un π-système et donc les mesures sont egales sur tout $\sigma(\mathcal{G}', \mathcal{H})$. □

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

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

où $\varphi(x) = \mathbb{E}[f(x, X_2, \ldots, X_n)]$.

Démonstration. Utiliser le théorème de Fubini sur la loi jointe de $X_1, \ldots, X_n$. □

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

$$f_0(\theta) = 1 \qquad f_n = f_{n-1}(f(\theta)) \quad \text{pour } n \geqslant 1.$$

Solution.

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

Or

$$\mathbb{E}[\theta^{Z_n}|Z_{n-1}] = \sum_{k=0}^{\infty} \mathbb{E}[\theta^{X_{n,1}+\cdots+X_{n,k}} I_{Z_{n-1}=k}|Z_{n-1}] = \sum_{k=0}^{\infty} \mathbb{E}[\theta^{X_{n,1}+\cdots+X_{n,k}}|Z_{n-1}] I_{Z_{n-1}=k}$$

et

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

car $\{X_{n,k}: k \geqslant 1\}$ est indépendant de $Z_{n-1} \hat{\in} \sigma(X_{m,k}, 1 \leqslant m \leqslant n-1, k \geqslant 1)$. Donc

$$\mathbb{E}[\theta^{Z_n}|Z_{n-1}] = \sum_{k=0}^{\infty} f(\theta)^k I_{Z_{n-1}=k} = f(\theta)^{Z_{n-1}}$$

ce qui nous permet de conclure que

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

Exercice 3. Soit $\mathbb{E}[Y|\mathcal{G}] = X$ 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

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

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

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

et déterminer les poids λ_i en fonction de Γ.

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

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}_1]|\mathcal{F}_2] \neq \mathbb{E}[\mathbb{E}[X|\mathcal{F}_2]|\mathcal{F}_1].$$

Exercice 8. Montrer les implications suivantes

$$X, Y \text{ independantes} \Rightarrow \mathbb{E}[X|Y] = \mathbb{E}[X] \Rightarrow \mathbb{E}[XY] = \mathbb{E}[X]\, \mathbb{E}[Y]$$

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