ESPERANCE CONDITIONNELLE
INTRODUCTION AUX MARTINGALES [1]

PRÉPARATION À L'AGRÉGATION EXTERNE DE MATHÉMATIQUES DE L'UNIVERSITÉ RENNES 1
ANNÉE 2010/2011

1. ESPÉRANCE CONDITIONNELLE

1.1 PROJECTION ORTHOGONALE ET MOYENNE

Pour toute sous-tribu $\mathscr{G}$ de $\mathscr{F}$, $\mathbb{L}_2(\mathscr{G})$ est un s.e.v. fermé. Donc toute v.a. $X \in \mathbb{L}_2(\mathscr{F})$ se projette sur $\mathbb{L}_2(\mathscr{G})$, et sa projection orthogonale $p_{\mathscr{G}}(X)$ vérifie $\langle X - p_{\mathscr{G}}(X), Z\rangle = 0 \; \forall Z \in \mathbb{L}_2(\mathscr{G})$. Examinons sur 3 exemples ce que donne une telle opération.

Exemple 1 Soit $\mathscr{T}$ la tribu triviale. La projection orthogonale de X sur $\mathbb{L}_2(\mathscr{T})$ est $\mathbb{E}X$ car

$$\mathbb{E}[(X - \mathbb{E}X)^2] \leq \mathbb{E}[(X-a)^2], \quad \forall a \in \mathbb{R}.$$

Exemple 2 Soit $\{A_1, \cdots, A_n\}$ une partition de Ω constituée d'événements de probabilité non nulle et $\mathscr{G}$ la tribu engendrée par cette partition. Comme $\mathscr{G} = \{\cup_{i\in I} A_i, I \subset \{1, \cdots, n\}\}$, la projection orthogonale de X sur $\mathbb{L}_2(\mathscr{G})$ vaut

$$\sum_{i=1}^{n} \mathbb{E}[X|A_i]\mathbf{1}_{A_i}.$$

Exemple 3 Soit $(X,Y) \sim \mathscr{U}(T)$, où $T = \{(x,y) \in [0,1]^2 : x+y \leq 1\}$. La projection orthogonale de X sur $\mathbb{L}_2(\sigma(Y))$ vaut $(1-Y)/2$, ce qui correspond à la moyenne des valeurs prises par X, lorsque Y est fixée.

Dans ce cadre d'étude, il existe donc un rapport entre projection orthogonale et espérance -on parlera plutôt d'espérance conditionnelle-. On exploite ces observations pour exporter la notion d'espérance conditionnelle au cas d'une tribu quelconque.

Définition 1.1 *Soient $X \in \mathbb{L}_2(\mathscr{F})$ une v.a.r. et $\mathscr{G}$ une sous-tribu. On appelle espérance conditionnelle de X sachant $\mathscr{G}$, que l'on note $\mathbb{E}[X|\mathscr{G}]$, la projection orthogonale de X sur $\mathbb{L}_2(\mathscr{G})$.*

$\mathbb{E}(X|\mathscr{G})$ est donc, au sens de la norme $\mathbb{L}_2$, la meilleure approximation de X par une v.a.r. $\mathscr{G}$-mesurable, ou encore la meilleure prévision (au même sens) qu'on puisse faire de X si on ne dispose que de l'information apportée par $\mathscr{G}$. Cette observation est fondamentale en modélisation : supposons, par exemple, qu'un phénomène évoluant dans le temps (discrétisé) a été modélisé par le processus stochastique $(X_k)_{k\geq 1}$. On peut chercher à "prédire" la valeur prise par X_{n+1}, sachant que seules les v.a.r. $X_1, \cdots, X_n$ ont été observées. Une solution est de calculer $\mathbb{E}[X_{n+1}|\sigma(X_1, \cdots, X_n)]$, en tant que meilleure approximation de X_{n+1} sachant le passé.

1.2 CAS GÉNÉRAL

Le concept d'espérance conditionnelle se prolonge de manière isométrique aux valables seulement intégrables :

Définition-Théorème 1.2 *Soient $X \in \mathbb{L}_1(\mathscr{F})$ une v.a.r. et $\mathscr{G}$ une sous-tribu. Il existe une unique v.a.r. $Y \in \mathbb{L}_1(\mathscr{G})$ telle que $\forall A \in \mathscr{G}$: $\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(Y\mathbf{1}_A)$. Cette v.a.r. Y est appelée espérance conditionnelle de X sachant $\mathscr{G}$, et notée $\mathbb{E}[X|\mathscr{G}]$.*

Bien entendu, les deux définitions 1.1 et 1.2 coincident ! Lorsque A est un événement, on pose $\mathbb{P}[A|\mathscr{G}] := \mathbb{E}[\mathbf{1}_A|\mathscr{G}]$. Par ailleurs, dans le cas où la sous-tribu $\mathscr{G}$ est engendrée par une v.a. Y, on note habituellement $\mathbb{E}[X|Y] := \mathbb{E}[X|\mathscr{G}]$.

Exercice 1.1 [FORMULE DE BAYES] Soit $\mathscr{G}$ une sous-tribu, et A, B deux événements tels que $\mathbb{P}(A) \neq 0$ et $B \in \mathscr{G}$. Montrer que

$$\mathbb{P}(B|A) = \frac{\int_B \mathbb{P}[A|\mathscr{G}]d\mathbb{P}}{\int_\Omega \mathbb{P}[A|\mathscr{G}]d\mathbb{P}}.$$

1. Benoît Cadre - ENS Cachan Bretagne

Lorsque X est une v.a.r. intégrable et Y est une v.a. à valeurs dans $\mathbb{Z}^d$, on peut toujours écrire pour tout $y \in \mathbb{Z}^d$ tel que $\mathbb{P}(Y = y) \neq 0$:

$$\mathbb{E}[X|Y = y] = \frac{1}{\mathbb{P}(Y = y)} \mathbb{E}[X\mathbf{1}_{\{Y=y\}}].$$

Ainsi, si on note ψ la fonction telle que $\psi(y) = \mathbb{E}[X|Y = y]$, on a $\mathbb{E}[X|Y] = \psi(Y)$.

La formule donnant $\mathbb{E}[X|Y = y]$ ne s'exporte pas en général dans le cas où Y est à valeurs dans $\mathbb{R}^d$ et pourtant, elle est abondamment utilisée ... Pourquoi ? D'après le lemme de factorisation de Doob, il existe une fonction borélienne φ telle que $\mathbb{E}[X|Y] = \varphi(Y)$. La notation $\mathbb{E}[X|Y = y]$ désigne alors $\varphi(y)$.

1.3 Propriétés fondamentales de l'espérance conditionnelle

Propriétés 1.1 *Soient X, Y des v.a.r. intégrables et $\mathscr{G}$ une sous-tribu.*

(1.) Si X est $\mathscr{G}$-mesurable, $X = \mathbb{E}[X|\mathscr{G}]$ p.s.
(2.) $\forall a, b \in R$: $\mathbb{E}[aX + bY|\mathscr{G}] = a\mathbb{E}[X|\mathscr{G}] + b\mathbb{E}[Y|\mathscr{G}]$ p.s.
(3.) Si $X \leq Y$ p.s., $\mathbb{E}[X|\mathscr{G}] \leq \mathbb{E}[Y|\mathscr{G}]$ p.s.
(4.) Si X est indépendante de $\mathscr{G}$, $\mathbb{E}[X|\mathscr{G}] = \mathbb{E}X$ p.s.
(5.) Si $\mathbb{E}|XY| < \infty$ et si X est $\mathscr{G}$-mesurable, $\mathbb{E}[XY|\mathscr{G}] = X\mathbb{E}[Y|\mathscr{G}]$ p.s.

Exercice 1.2 Soient $X_1, \cdots, X_n$ des v.a.r.iid de loi intégrable et S_n leur somme. Calculer $\mathbb{E}[X_1|S_n]$.

Exercice 1.3 Soit (X, Y) un vecteur gaussien centré de $\mathbb{R}^2$ tel que $\mathrm{var}(X) = \mathrm{var}(Y) = \mathrm{cov}(X, Y) = 1$. Calculer $\mathbb{E}[X|Y]$.

Propriétés 1.2 *$X \in \mathbb{L}_1(\mathscr{F})$ une v.a.r. et $\mathscr{G}_1, \mathscr{G}_2$ des sous-tribus.*

(1.) Si $\mathscr{G}_1 \subset \mathscr{G}_2$, $\mathbb{E}[\mathbb{E}[X|\mathscr{G}_1]|\mathscr{G}_2] = \mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] = \mathbb{E}[X|\mathscr{G}_1]$ p.s. En particulier, $\mathbb{E}[\mathbb{E}[X|\mathscr{G}_1]] = \mathbb{E}X$ p.s.
(2.) Si X et $\mathscr{G}_1$ sont indépendantes de $\mathscr{G}_2$, $\mathbb{E}[X|\mathscr{G}_1 \vee \mathscr{G}_2] = \mathbb{E}[X|\mathscr{G}_1]$ p.s.

Propriété 1.3 [Inégalité de Jensen] *Soient $\varphi : \mathbb{R} \to \mathbb{R}$ convexe, $X \in \mathbb{L}_1(\mathscr{F})$ une v.a.r. et $\mathscr{G}$ une sous-tribu. Si $\mathbb{E}|\varphi(X)| < \infty$, alors $\varphi(\mathbb{E}[X|\mathscr{G}]) \leq \mathbb{E}[\varphi(X)|\mathscr{G}]$ p.s.*

Propriétés 1.4 *Soient $(X_n)_n$ des v.a.r. appartenant à $\mathbb{L}_1(\mathscr{F})$, X une v.a.r. et $\mathscr{G}$ une sous-tribu.*

(1.) Propriété de Beppo-Lévi *: si $0 \leq X_n \nearrow X$ lorsque $n \nearrow \infty$ et $\mathbb{E}|X| < \infty$, alors $\mathbb{E}[X_n|\mathscr{G}] \nearrow \mathbb{E}[X|\mathscr{G}]$ p.s. lorsque $n \nearrow \infty$.*
(2.) Propriété de Lebesgue *: si $X_n \to X$ lorsque $n \to \infty$ et $\mathbb{E}\sup_n |X_n| < \infty$, alors $\mathbb{E}[|X_n - X||\mathscr{G}] \to 0$ p.s. lorsque $n \to \infty$.*
(3.) Propriété de Fatou *: si $X_n \geq 0$ p.s. et pour tout n, et si $\mathbb{E}|\liminf_n X_n| < \infty$, alors $\mathbb{E}[\liminf_n X_n|\mathscr{G}] \leq \liminf_n \mathbb{E}[X_n|\mathscr{G}]$ p.s.*

1.4 Comment calculer une espérance conditionnelle ?

Il n'existe pas de méthode générale pour calculer une espérance conditionnelle. Selon les cas, on peut néanmoins utiliser l'une des 3 astuces suivantes :

Astuce 1.1 [Vecteur à densité] *Soit (X, Y) un vecteur aléatoire à valeurs dans $\mathbb{R}^p \times \mathbb{R}^q$ qui possède une densité f, et $\psi : \mathbb{R}^p \to \mathbb{R}$ borélienne telle que $\psi(X)$ est intégrable. Pour $\mathbb{P}_Y$-p.t. y :*

$$\mathbb{E}\big[\psi(X)|Y = y\big] = \int_{\mathbb{R}^p} \psi(x) \frac{f(x, y)}{f_Y(y)} dx.$$

Astuce 1.2 [V.A. indépendantes] *Soient X, Y des v.a. à valeurs dans $\mathbb{R}^d$ et $\mathbb{R}^p$ resp., et $\psi : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}$ telle que $\mathbb{E}|\psi(X, Y)| < \infty$. Si Y est indépendante de X, alors pour $\mathbb{P}_X$-p.t. x :*

$$\mathbb{E}[\psi(X, Y)|X = x] = \int_{\mathbb{R}^p} \psi(x, y) \mathbb{P}_Y(dy), \text{ p.s.}$$

Astuce 1.3 [π-système] *Soit X une v.a.r., $\mathscr{G}$ une sous-tribu et $\mathscr{P}$ un π-système qui engendre $\mathscr{G}$. Si il existe une v.a.r. intégrable Z qui est $\mathscr{G}$-mesurable et telle que $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A]$ pour tout $A \in \mathscr{P}$, alors $\mathbb{E}[X|\mathscr{G}] = Z$ p.s.*

Exercice 1.4 Soient $X_1, \cdots, X_n$ des v.a.iid de loi $\mathscr{U}[0, 1]$ et $T = \sup(X_1, \cdots, X_n)$. Calculer $\mathbb{E}[X_1|T]$.

2. Martingales

2.1 Introduction

Le terme *martingale* provient du mot provencal "martegalo" . Il s'agit d'une courroie qui, placée sous le ventre d'un cheval, relie la sangle à la muserolle pour empêcher l'animal de trop lever la tête. Cette terminologie fut introduite par de Moivre, peut-être dans l'espoir que les martingales permettraient de "brider" le hasard ?

Définitions 2.1

(1.) Une suite croissante de sous-tribus est appelée filtration ;

(2.) Soit $(\mathscr{F}_n)_n$ une filtration. Le processus stochastique $(X_n)_n$ est dit $(\mathscr{F}_n)_n$-adapté si pour tout n, X_n est $\mathscr{F}_n$-mesurable ;

(3.) La filtration naturelle du processus $(X_n)_n$ est $\big(\sigma(X_0,\cdots,X_n)\big)_n$: on dit que $(X_n)_n$ engendre cette filtration.

$(X_n)_n$ est donc adapté à $(\mathscr{F}_n)_n$ si, à chaque instant n, l'information disponible permet de connaître la valeur prise par X_n. De plus, $(X_n)_{n\geq 0}$ engendre $(\mathscr{F}_n)_{n\geq 0}$ si, à chaque instant n, l'information disponible est exactement celle qui résulte des valeurs prises par $X_0,\cdots,X_n$. Lorsque $(X_n)_n$ engendre $(\mathscr{F}_n)_n$, cette filtration est la plus petite filtration qui adapte $(X_n)_n$.

Définition 2.2 [Ville et Lévy]*Soit $(X_n)_n$ un processus stochastique à valeurs réelles et $(\mathscr{F}_n)_n$ une filtration. On dit que $(X_n)_n$ est une $(\mathscr{F}_n)_n$-(sur ; sous)-martingale si*

(i) $\forall n$, X_n est intégrable ;

(ii) $(X_n)_n$ est $(\mathscr{F}_n)_n$-adapté ;

(iii) $\forall n$, $\mathbb{E}[X_{n+1}|\mathscr{F}_n](\leq;\ \geq) = X_n$ p.s.

Noter que l'on a alors $\mathbb{E}[X_{n+p}|\mathscr{F}_n](\leq;\ \geq) = X_n$ pour chaque n,p. De plus, si $(X_n)_n$ est une $(\mathscr{F}_n)_n$-(sur ; sous)-martingale, c'est aussi une (sur ; sous)-martingale pour sa filtration naturelle.

2.2 Quelques martingales classiques

(a.) Marche aléatoire. Soient $(\xi_n)_n$ des v.a.r.iid de moyenne p, et $S_n = \xi_0+\cdots+\xi_n$ $\forall n$. Le processsus $(S_n)_n$ est une sur-martingale, sous-martingale ou martingale, selon le signe de p.

(b.) Martingale de Doob. Soit Z une v.a.r. intégrable et $(\mathscr{F}_n)_n$ une filtration. Le processus $(\mathbb{E}[Z|\mathscr{F}_n])_n$ est une $(\mathscr{F}_n)_n$-martingale.

(c.) Soit $(X_n)_n$ une $(\mathscr{F}_n)_n$-martingale et $\varphi:\mathbb{R}\to\mathbb{R}$ une fonction convexe telle que $\varphi(X_n)$ est intégrable $\forall n$. Le processus $\big(\varphi(X_n)\big)_n$ est une $(\mathscr{F}_n)_n$-sous-martingale.

(d.) Soit $(X_n)_n$ une chaîne de Markov homogène à valeurs dans $\mathscr{E}$ fini, de matrice de transition $(p_{ij})_{i,j\in\mathscr{E}}$, et $\psi:\mathscr{E}\to\mathscr{E}$ une fonction bornée et *harmonique*, i.e. $\forall i\in\mathscr{E}$: $\psi(i)=\sum_{j\in\mathscr{E}} p_{ij}\psi(j)$. Alors $(\psi(X_n))_n$ est une martingale.

(e.) Processus de Galton-Watson. Soient $(\xi_i^n)_{i,n\geq 0}$ des v.a.iid, intégrables de moyenne m, et à valeurs dans $\mathbb{N}$. On note $Z_0=1$ et pour $n\geq 0$:

$$Z_{n+1}=\sum_{i=1}^{Z_n}\xi_i^{n+1},$$

avec la convention $\sum_{i=1}^{0}=0$. Le processus $(Z_n/m^n)_n$ est une martingale.

(f.) Intégrale stochastique discrète. Soit $(X_n)_n$ une $(\mathscr{F}_n)_n$-martingale, et $(Y_n)_n$ un processus prévisible (i.e. Y_0 est $\mathscr{F}_0$-mesurable et Y_n est $\mathscr{F}_{n-1}$-mesurable). On suppose en outre que Y_n est borné. Le processus $(Z_n)_n$ défini pour tout n par

$$Z_n = Y_0X_0+\sum_{k=1}^{n} Y_k(X_k-X_{k-1})$$

est une $(\mathscr{F}_n)_n$-martingale.

Les martingales qui apparaissent ci-dessus peuvent modéliser les phénomènes suivants :

(a.) Marche de l'ivrogne ...

(b.) Supposons que les v.a. $X_1,X_2,\cdots$ représentent les positions successives d'une fourmi en déplacement, et que l'on veuille prédire la valeur de $Z=X_p$ en ayant seulement observé $X_1,\cdots,X_n$, avec $n\leq p$. Dans ce contexte, $\mathbb{E}[Z|X_1,\cdots,X_n]$ représente la meilleure approximation de Z (pour le critère quadratique) lorsque seule l'information apportée par $X_1,\cdots,X_n$ est connue. Cependant, cela ne nous dit pas comment calculer cette espérance conditionnelle.

(e.) On considère une population de bactéries hermaphrodites. Alors, ξ_i^{n+1} représente le nombre de "descendants" de la i-ème bactérie de la génération n, m est le taux de fécondité de l'espèce et Z_n recense le nombre d'individus de la génération n.

(f.) Considérons que, à l'instant k, X_k est le cours d'une action donnée et Y_k est la quantité d'actions détenues. Le processus $(Y_n)_n$ est prévisible car le financier décide en fonction des cours précédents combien d'actions il v.a. détenir. Le bénéfice algébrique réalisé entre les instants k-1 et k est $Y_k(X_k - X_{k-1})$, et la fortune totale accumulée à l'instant n est Z_n.

2.3 Convergence p.s. des martingales

Théorème 2.1 *Soit $(X_n)_n$ une sous-martingale bornée dans L_1 (i.e., $\sup_n \mathbb{E}|X_n| < \infty$). Alors, il existe une v.a.r. intégrable X_∞ telle que $X_n \to X_\infty$ p.s.*

La martingale de Doob et le processus $(Z_n/m^n)_n$ de l'exemple **(e.)** convergent p.s. vers des v.a.r. intégrables. En particulier, lorsque $m < 1$, la population de bactéries finira donc par s'éteindre.

Par ailleurs, une surmartingale positive étant bornée dans $\mathbb{L}_1$, on a le résultat :

Corollaire 2.1 *Une sur-martingale positive converge p.s. vers une v.a.r. intégrable.*

La convergence p.s. n'entraîne pas la convergence des espérances, ce qui est fâcheux en général, et tout particulièrement pour les martingales qui sont définies par des espérances conditionnelles. Pour illustrer cette affirmation, considérons des v.a.r.iid $\xi_1, \xi_2, \cdots$ telles que $\mathbb{P}(\xi_1 = 0) = \mathbb{P}(\xi_1 = 2) = 1/2$, et $X_n = \xi_1 \cdots \xi_n$. Le processus $(X_n)_n$ -qui est borné dans $\mathbb{L}_1$- est une martingale qui converge p.s. vers 0, mais qui ne converge pas dans $\mathbb{L}_1$. La bonne condition pour s'assurer une convergence $\mathbb{L}_1$ fera l'objet d'un prochain cours.

En revanche, le comportement asymptotique moyen des martingales de carrés intégrables est plus simple à traiter. Soit $(X_n)_{n\geq 0}$ une martingale bornée dans $\mathbb{L}_2$ (elle converge p.s. car elle est en particulier bornée dans $\mathbb{L}_1$). Comme $(X_n^2)_{n\geq 0}$ est une sous-martingale, la suite $(\mathbb{E}X_n^2)_{n\geq 0}$ est croissante, et elle est aussi bornée, donc elle converge. Or, pour chaque $n \geq p \geq 0$:

$$\mathbb{E}(X_n - X_p)^2 = \mathbb{E}X_n^2 - \mathbb{E}X_p^2,$$

ce qui montre que $(X_n)_{n\geq 0}$ est une suite de Cauchy dans $\mathbb{L}_2$, et donc que $(X_n)_{n\geq 0}$ converge dans $\mathbb{L}_2$.

REFERENCES

• D. Dacunha-Castelle et M. Duflo, *Probabilités et statistiques - Tome 2 : Problèmes à temps mobile (Cours et Exercices)*, Masson, 1983.
• D. Foata et A. Fuchs, *Processus stochastiques - Processus de Poisson, chaînes de Markov et martingales*, Dunod, 2002.
• J. Neveu, *Martingales à temps discret*, Masson, 1972.
• J.-Y. Ouvrard, *Probabilités 2 : maîtrise agrégation*, Cassini, 2001.
• D. Williams, *Probability with martingales*, Cambridge University Press, 1991.