ESPERANCE CONDITIONNELLE
INTRODUCTION AUX MARTINGALES [1]

Préparation à l'agrégation externe de Mathématiques de l'université Rennes 1
Année 2008/2009

1. Espérance conditionnelle

1.1 Introduction

Soit $X \in L^2(\Omega, \mathcal{F}, I\!P)$ une var et $\{A_1, \cdots, A_n\}$ une partition de Ω constituée d'événements de probabilité non nulle. Notons $\mathcal{G}$ la tribu engendrée par cette partition. Comme $\mathcal{G} = \{\cup_{i\in I} A_i,\ I \subset \{1, \cdots, n\}\}$, il est naturel de définir, dans ce cas, l'espérance conditionnelle de X sachant $\mathcal{G}$ par la relation

$$I\!E[X|\mathcal{G}] = \sum_{i=1}^n I\!E[X|A_i]\mathbf{1}_{A_i}.$$

Un calcul montre que cet objet est en fait la projection orthogonale de X sur $L^2(\Omega, \mathcal{G}, I\!P)$. On exploite ensuite cette observation pour exporter la notion d'espérance conditionnelle au cas d'une tribu quelconque.

Définition 1.1 *Soient $X \in L^2(\Omega, \mathcal{F}, I\!P)$ une var et $\mathcal{G}$ une sous-tribu. On appelle espérance conditionnelle de X sachant $\mathcal{G}$, que l'on note $I\!E[X|\mathcal{G}]$, la projection orthogonale de X sur $L^2(\Omega, \mathcal{G}, I\!P)$.*

$I\!E(X|\mathcal{G})$ est donc -au sens de la norme L^2- la meilleure approximation de X par une var $\mathcal{G}$-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 $\mathcal{G}$.

Cette notion se prolonge de manière isométrique aux variables seulement intégrables :

Définition-Théorème 1.2 *Soient $X \in L^1(\Omega, \mathcal{F}, I\!P)$ une var et $\mathcal{G}$ une sous-tribu. Il existe une unique var $Y \in L^1(\Omega, \mathcal{G}, I\!P)$ telle que $\forall A \in \mathcal{G}$: $I\!E(X\mathbf{1}_A) = I\!E(Y\mathbf{1}_A)$. Cette var Y est appelée espérance conditionnelle de X sachant $\mathcal{G}$, et notée $I\!E[X|\mathcal{G}]$.*

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

Exercice 1.1 [Formule de Bayes] Soit $\mathcal{G}$ une sous-tribu, et A, B deux événements tels que $I\!P(A) \neq 0$ et $B \in \mathcal{G}$. Montrer que

$$I\!P(B|A) = \frac{\int_B I\!P[A|\mathcal{G}]dI\!P}{\int_\Omega I\!P[A|\mathcal{G}]dI\!P}.$$

1.2 Comment calculer une espérance conditionnelle ?

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

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

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

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

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 :

[1] Benoît Cadre - ENS Cachan Bretagne

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

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

Astuce 1.2 [V.a. indépendantes] *Soient X, Y des va à valeurs dans $I\!R^d$ et $I\!R^p$ resp., et $\psi : I\!R^d \times I\!R^p \to I\!R$ telle que $I\!E|\psi(X,Y)| < \infty$. Si Y est indépendante de X, alors*

$$I\!E[\psi(X,Y)|X] = \int_{I\!R^p} \psi(X,y) I\!P_Y(dy), \text{ p.s.}$$

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

Exercice 1.2 Soient $X_1, \cdots, X_n$ des vaiid de loi $\mathcal{U}[0,1]$ et $T = \sup(X_1, \cdots, X_n)$. Calculer $I\!E[X_1|T]$.

1.3 Propriétés fondamentales de l'espérance conditionnelle

Propriétés 1.1 *Soient X, Y des var intégrables et $\mathcal{G}$ une sous-tribu.*
- *(1.) Si X est $\mathcal{G}$-mesurable, $X = I\!E[X|\mathcal{G}]$ p.s.*
- *(2.) $\forall a, b \in R$: $I\!E[aX + bY|\mathcal{G}] = aI\!E[X|\mathcal{G}] + bI\!E[Y|\mathcal{G}]$ p.s.*
- *(3.) Si $X \le Y$ p.s., $I\!E[X|\mathcal{G}] \le I\!E[Y|\mathcal{G}]$ p.s.*
- *(4.) Si X est indépendante de $\mathcal{G}$, $I\!E[X|\mathcal{G}] = I\!EX$ p.s.*
- *(5.) Si $I\!E|XY| < \infty$ et si X est $\mathcal{G}$-mesurable, $I\!E[XY|\mathcal{G}] = XI\!E[Y|\mathcal{G}]$ p.s.*

Propriétés 1.2 *$X \in L^1(\Omega, \mathcal{F}, I\!P)$ une var et $\mathcal{G}_1, \mathcal{G}_2$ des sous-tribus.*
- *(1.) Si $\mathcal{G}_1 \subset \mathcal{G}_2$, $I\!E[I\!E[X|\mathcal{G}_1]|\mathcal{G}_2] = I\!E[I\!E[X|\mathcal{G}_2]|\mathcal{G}_1] = I\!E[X|\mathcal{G}_1]$ p.s. En particulier, $I\!E[I\!E[X|\mathcal{G}_1]] = I\!EX$ p.s.*
- *(2.) Si X et $\mathcal{G}_1$ sont indépendantes de $\mathcal{G}_2$, $I\!E[X|\mathcal{G}_1 \vee \mathcal{G}_2] = I\!E[X|\mathcal{G}_1]$ p.s.*

Propriété 1.3 [Inégalité de Jensen] *Soient $\varphi : I\!R \to I\!R$ convexe, $X \in L^1(\Omega, \mathcal{F}, I\!P)$ une var et $\mathcal{G}$ une sous-tribu. Si $I\!E|\varphi(X)| < \infty$, alors $\varphi(I\!E[X|\mathcal{G}]) \le I\!E[\varphi(X)|\mathcal{G}]$ p.s.*

Propriétés 1.4 *Soient $(X_n)_n$ des var appartenant à $L^1(\Omega, \mathcal{F}, I\!P)$, X une var et $\mathcal{G}$ une sous-tribu.*
- *(1.)* Propriété de Beppo-Lévi *: si $0 \le X_n \nearrow X$ lorsque $n \nearrow \infty$ et $I\!E|X| < \infty$, alors $I\!E[X_n|\mathcal{G}] \nearrow I\!E[X|\mathcal{G}]$ p.s. lorsque $n \nearrow \infty$.*
- *(2.)* Propriété de Lebesgue *: si $X_n \to X$ lorsque $n \to \infty$ et $I\!E\sup_n |X_n| < \infty$, alors $I\!E[|X_n - X||\mathcal{G}] \to 0$ p.s. lorsque $n \to \infty$.*
- *(3.)* Propriété de Fatou *: si $X_n \ge 0$ p.s. et pour tout n, et si $I\!E|\liminf_n X_n| < \infty$, alors $I\!E[\liminf_n X_n|\mathcal{G}] \le \liminf_n I\!E[X_n|\mathcal{G}]$ p.s.*

Exercice 1.3 Soient $X_1, \cdots, X_n$ des variid de loi intégrable et S_n leur somme. Calculer $I\!E[X_1|S_n]$, puis sa limite lorsque $n \to \infty$.

2. Martingales

2.1 Introduction

Le terme *martingale* provient du mot provencal "martegalo" désignant 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 $(\mathcal{F}_n)_n$ une filtration. Le processus stochastique $(X_n)_n$ est dit $(\mathcal{F}_n)_n$-adapté si pour tout n, X_n est $\mathcal{F}_n$-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 $(\mathcal{F}_n)_n$ 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 $(\mathcal{F}_n)_{n\geq 0}$ 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 $(\mathcal{F}_n)_n$, 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 $(\mathcal{F}_n)_n$ une filtration. On dit que $(X_n)_n$ est une $(\mathcal{F}_n)_n$-(sur ; sous)-martingale si*

(i) $\forall n$, *X_n est intégrable ;*

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

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

Noter que l'on a alors $\mathbb{E}[X_{n+p}|\mathcal{F}_n](\leq;\ \geq) = X_n$ pour chaque n, p. De plus, si $(X_n)_n$ est une $(\mathcal{F}_n)_n$-(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 variid 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 var intégrable et $(\mathcal{F}_n)_n$ une filtration. Le processus $(\mathbb{E}[Z|\mathcal{F}_n])_n$ est une $(\mathcal{F}_n)_n$-martingale.

(c.) Soit $(X_n)_n$ une $(\mathcal{F}_n)_n$-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 $(\mathcal{F}_n)_n$-sous-martingale.

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

(e.) Processus de Galton-Watson. Soient $(\xi_i^n)_{i,n\geq 0}$ des vaiid, 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. Soit $(X_n)_n$ une $(\mathcal{F}_n)_n$-martingale, et $(Y_n)_n$ un processus prévisible (i.e. Y_0 est $\mathcal{F}_0$-mesurable et Y_n est $\mathcal{F}_{n-1}$-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 $(\mathcal{F}_n)_n$-martingale.

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

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

(b.) Z représente la taille d'un individu à l'âge de 20 ans. Cette taille est fonction d'un grand nombre de variables, dont seulement n peuvent être observées à l'âge de 10 ans. $\mathcal{F}_n$ représente l'information apportée par ces n variables. Dans ce contexte, $\mathbb{E}[Z|\mathcal{F}_n]$ est la meilleure approximation (pour le critère quadratique) de Z lorsque seule l'information $\mathcal{F}_n$ 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 va 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 var 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 var intégrables. En particulier, lorsque $m < 1$, la population de bactéries finira donc par s'éteindre.

Par ailleurs, une surmartingale positive étant bornée dans L^1, on a le résultat :

Corollaire 2.1 *Une sur-martingale positive converge p.s. vers une var 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 : considérons $\xi_1, \xi_2, \cdots$ des variid 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 L^1- est une martingale qui converge p.s. vers 0, mais qui ne converge pas dans L^1. La bonne condition pour s'assurer une convergence L^1 fera l'objet d'un prochain cours.

En revanche, pour ce qui concerne la convergence dans L^1, le traitement des martingales de carrés intégrables est plus simple. Soit $(X_n)_{n\geq 0}$ une martingale bornée dans L^2 (elle converge p.s. car elle est en particulier bornée dans 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 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.