Séance XII : Espérance conditionnelle et introduction aux processus stochastiques

A) Objectifs de la séance

A la fin de cette séance,

- je maîtrise la caractérisation de l'espérance conditionnelle d'une v.a. dans L^1 par rapport à une sous-tribu;
- je suis capable d'exprimer l'espérance conditionnelle dans L^2 comme une projection orthogonale;
- je suis capable de déterminer l'espérance conditionnelle d'une variable aléatoire;
- je connais les propriétés de l'espérance conditionnelle et je suis capable de manipuler cet objet.

B) Pour se familiariser avec les concepts (à traiter avant les séances de TD)

Les questions XII.1 et XII.2 sont à traiter avant la séance de TD. Les corrigés sont disponibles sur internet.

Question XII.1
Soient X une variable aléatoire réelle dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On pose $\mathbb{V}\mathrm{ar}(X \mid \mathcal{G}) = \mathbb{E}\big[(X - \mathbb{E}[X \mid \mathcal{G}])^2 \mid \mathcal{G}\big]$.
Q. XII.1.1 Montrer que

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\big[\mathbb{V}\mathrm{ar}(X \mid \mathcal{G})\big] + \mathbb{V}\mathrm{ar}\big(\mathbb{E}[X \mid \mathcal{G}]\big).$$

Question XII.2 (Conditionnement par rapport à une variable discrète)
Dans l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, soit X une variable aléatoire réelle dans $L^1(\Omega, \mathcal{F}, \mathbb{P})$ et soit Y une variable aléatoire prenant ses valeurs dans l'ensemble dénombrable $\{y_n;\ n \in \mathbb{N}\}$ telle que $\mathbb{P}(Y = y_n) > 0$ pour tout n.
Q. XII.2.1 Déterminer $\mathbb{E}[X \mid Y]$.

C) Exercices

Dans le cas de variables aléatoires gaussiennes, l'espérance conditionnelle peut s'exprimer directement à partir des matrices de covariances. Cette relation est importante en statistiques pour les problèmes de régression linéaire.

Exercice XII.1 (Cas gaussien)
Soient X et Y des vecteurs aléatoires à valeurs dans $\mathbb{R}^n$ et $\mathbb{R}^p$ respectivement, tels que (X, Y) soit un vecteur gaussien. On suppose le vecteur gaussien X non-dégénéré.

E. XII.1.1 Dans cette question, on suppose $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. On pose

$$U = Y - \Sigma_{YX} K_X^{-1} X,$$

où $K_X = [\mathrm{Cov}(X_i, X_j)]_{i,j}$ et $\Sigma_{YX} = [\mathrm{Cov}(Y_i, X_j)]_{i,j}$.

(a) Montrer que (U, X) est gaussien.

(b) Montrer que X et U sont indépendants.

(c) En déduire que

$$\mathbb{E}[Y \mid X] = \Sigma_{YX} K_X^{-1} X.$$

E. XII.1.2 Montrer que $\mathbb{E}[Y \mid X]$ peut s'écrire sous la forme

$$\mathbb{E}[Y \mid X] = \mathbb{E}[Y] + \Sigma_{YX} K_X^{-1} (X - \mathbb{E}[X]).$$

E. XII.1.3 Montrer que $Y - \mathbb{E}[Y \mid X]$ et X sont indépendants.

L'espérance conditionnelle permet de caractériser l'indépendance entre variables aléatoires.

Exercice XII.2

E. XII.2.1 Soient X et Y deux variables aléatoires réelles. Montrer que X et Y sont indépendantes si et seulement si pour toute application $g : \mathbb{R} \to \mathbb{R}$ borélienne bornée, on a

$$\mathbb{E}[g(Y) \mid X] = \mathbb{E}[g(Y)] \quad \mathbb{P} - \text{p.s.}$$

E. XII.2.2 Application : soit (X, Y) un couple de variables aléatoires admettant pour densité $p : (x, y) \mapsto e^{-y} \mathbb{1}_{\{0<x<y\}}(x, y)$. Calculer la loi conditionnelle de Y sachant $X = x$. En déduire que X et $Y - X$ sont indépendantes.

Exercice XII.3 (Somme d'un nombre aléatoire de v.a.)
Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a. réelles i.i.d. et soit N une v. a. à valeurs dans $\mathbb{N}$ indépendante des X_n. On suppose que les variables N et les X_n possèdent des moments d'ordre 1.

On définit

$$Y(\omega) = \sum_{i=1}^{N(\omega)} X_i(\omega).$$

E. XII.3.1 Montrer que Y est une variable aléatoire. Calculer $\mathbb{E}[Y]$.

D) Approfondissement

Exercice XII.4 (Dérivée de Radon-Nikodym)
Dans l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, soit X une varable aléatoire réelle admettant une densité f continue. Soit $A \in \mathcal{F}$ un événement fixé.

On suppose qu'il existe une fonction continue g telle que pour tout $x \in \mathbb{R}$,

$$\frac{\mathbb{P}(A \cap \{X \in [x, x+h[\})}{\mathbb{P}(X \in [x, x+h[)} \underset{h \to 0}{\longrightarrow} g(x).$$

E. XII.4.1 Montrer que $\mathbb{P}(A \mid X) = g(X)$ $\mathbb{P}$-p.s.

Exercice XII.5
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. et $(\mathcal{F}_n)_{n\in\mathbb{N}}$ une filtration. On suppose que pour tout n, X_n est $\mathcal{F}_n$-mesurable et $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = 0$. On pose $S_n = X_0 + \cdots + X_n$ pour tout $n \in \mathbb{N}$.
E. XII.5.1 Montrer que $\{S_n;\ n \in \mathbb{N}\}$ est une $\mathcal{F}_n$-martingale.

Exercice XII.6 (Décomposition de Doob)
Soit $\{X_n;\ n \in \mathbb{N}\}$ une martingale pour la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, de carré intégrable.
E. XII.6.1 Montrer qu'il existe un processus croissant prévisible $\{A_n;\ n \in \mathbb{N}\}$ et une $\mathcal{F}_n$-martingale $\{Y_n;\ n \in \mathbb{N}\}$ tels que $\mathbb{E}(X_0^2) = A_0$ et

$$\forall n \in \mathbb{N}; \quad X_n^2 = A_n + Y_n.$$

E. XII.6.2 Application. Soit $(T_n)_{n\in\mathbb{N}^*}$ une suite de v.a. réelles i.i.d. de carré intégrable et telle que $\mathbb{E}[T_n] = 0$ et $\mathbb{E}[T_n^2] = \sigma^2$. On pose $S_0 = 0$ et $S_n = T_1 + \cdots + T_n$ pour $n \geq 1$.

Montrer que $\{S_n^2 - n\sigma^2;\ n \in \mathbb{N}\}$ est une martingale.

Exercice XII.7
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires i.i.d. définies sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, et à valeurs dans l'espace mesurable $(E, \mathcal{E})$. On note μ la loi des X_n.

Soient N_1 et N_2 deux v.a. à valeurs dans $\mathbb{N}^*$ telles que $1 \leq N_1 < N_2$. On suppose que pour tout $m \in \mathbb{N}^*$, l'événement $\{N_i = m\}$ ne dépend que de $X_0, X_1, \ldots, X_{m-1}$.
E. XII.7.1 Montrer que les v.a. X_{N_1} et X_{N_2} sont i.i.d.

Séance 12 : Eléments de correction des exercices

Solution de Q. XII.1.1 On écrit

$$X - \mathbb{E}[X] = X - \mathbb{E}[X \mid \mathcal{G}] + \mathbb{E}[X \mid \mathcal{G}] - \mathbb{E}[X].$$

Or $\mathbb{E}[X \mid \mathcal{G}] - \mathbb{E}[X]$ est une variable aléatoire $\mathcal{G}$-mesurable (et dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$) et, par définition de l'espérance conditionnelle dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$, la variable aléatoire $X - \mathbb{E}[X \mid \mathcal{G}]$ est orthogonale à toute variable dans $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

On en déduit que

$$\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[\left(X - \mathbb{E}[X \mid \mathcal{G}]\right)^2\right] + \mathbb{E}\left[\left(\mathbb{E}[X \mid \mathcal{G}] - \mathbb{E}[X]\right)^2\right].$$

On identifie alors

$$\mathbb{E}\left[\mathbb{V}\mathrm{ar}(X \mid \mathcal{G})\right] = \mathbb{E}\left[\left(X - \mathbb{E}[X \mid \mathcal{G}]\right)^2\right]$$

et

$$\mathbb{V}\mathrm{ar}(\mathbb{E}[X \mid \mathcal{G}]) = \mathbb{E}\left[\left(\mathbb{E}[X \mid \mathcal{G}] - \mathbb{E}[X]\right)^2\right],$$

ce qui implique le résultat.

Solution de Q. XII.2.1 On a vu dans l'Exercice VI.5 que toute variable aléatoire réelle qui est $\sigma(Y)$-mesurable peut s'écrire sous la forme $\Phi(Y)$ où $\Phi : \mathbb{R} \to \mathbb{R}$ est borélienne.

Ainsi, l'espérance conditionnelle $\mathbb{E}[X \mid Y]$ étant $\sigma(Y)$-mesurable où Y ne prend que les valeurs discrètes $\{y_n;\ n \in \mathbb{N}\}$, on peut écrire

$$\mathbb{E}[X \mid Y] = \sum_{n=1}^{\infty} b_n\, \mathbb{1}_{\{y_n\}}(Y) = \sum_{n=1}^{\infty} b_n\, \mathbb{1}_{\{Y=y_n\}} \quad \text{p.s.}$$

où $(b_n)_{n\in\mathbb{N}}$ est une suite de réels.

On détermine la suite $(b_n)_{n\in\mathbb{N}}$ grâce à la définition de $\mathbb{E}[X \mid Y]$

$$\forall n \in \mathbb{N}, \quad \int_{\{Y=y_n\}} \mathbb{E}[X \mid Y]\, d\mathbb{P} = \int_{\{Y=y_n\}} X\, d\mathbb{P}.$$

D'où

$$b_n\, \mathbb{P}(Y = y_n) = \int_{\{Y=y_n\}} X\, d\mathbb{P},$$

ce qui entraîne

$$\mathbb{E}[X \mid Y] = \sum_{n=1}^{\infty} \frac{\int_{\{Y=y_n\}} X\, d\mathbb{P}}{\mathbb{P}(Y = y_n)}\, \mathbb{1}_{\{Y=y_n\}} \quad \text{p.s.}$$

Solution de E. XII.1.1

(a) Toute combinaison linéaire des composantes du vecteur (U, X) est combinaison linéaire des composantes du vecteur (X, Y) qui est gaussien, donc est une v.a. réelle gaussienne.

On peut également utiliser le formalisme de la question Q. X.4.1 pour voir le vecteur (U, X) comme image de (X, Y) par une transformation linéaire. Ainsi, le vecteur $(U, X)^t$ peut s'écrire

$$\begin{pmatrix} U \\ X \end{pmatrix} = \begin{pmatrix} -\Sigma_{YX}K_X^{-1} & I_p \\ I_n & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix},$$

où les vecteurs U et X sont notés en colonne.

(b) Dans la question Q. X.4.1, on a montré que U et X sont indépendantes si et seulement si $\Sigma_{UX} = 0$. Les vecteurs U et X étant centrés, cette condition s'écrit $\mathbb{E}[UX^t] = 0$.

En utilisant l'expression de U, on calcule

$$\begin{aligned} \mathbb{E}[UX^t] &= \mathbb{E}[YX^t] - \Sigma_{YX}K_X^{-1}\,\mathbb{E}[XX^t] \\ &= \Sigma_{YX} - \Sigma_{YX}K_X^{-1}K_X = 0. \end{aligned}$$

(c) Comme U et X sont indépendants, on a $\mathbb{E}[U \mid X] = \mathbb{E}[U] = 0$. Ainsi,

$$\begin{aligned} \mathbb{E}[U \mid X] &= \mathbb{E}[Y \mid X] - \Sigma_{YX}K_X^{-1}\,\mathbb{E}[X \mid X] \\ &= \mathbb{E}[Y \mid X] - \Sigma_{YX}K_X^{-1}X = 0, \end{aligned}$$

ce qui entraîne

$$\mathbb{E}[Y \mid X] = \Sigma_{YX}K_X^{-1}X.$$

Solution de E. XII.1.2 Dans le cas général où $\mathbb{E}[X]$ et $\mathbb{E}[Y]$ ne sont pas nécessairement nulles, on applique le résultat précédent à $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$.

On remarque d'abord que la covariance est invariante par translation (par une constante), donc les matrices de covariances Σ_{YX} et K_X sont invariantes par translation. Par conséquent, $\Sigma_{(Y-\mathbb{E}[Y])(X-\mathbb{E}[X])} = \Sigma_{YX}$ et $K_{X-\mathbb{E}[X]} = K_X$. On obtient

$$\mathbb{E}\left(Y - \mathbb{E}[Y] \mid X - \mathbb{E}[X]\right) = \Sigma_{YX}K_X^{-1}(X - \mathbb{E}[X]).$$

Or $\mathbb{E}\left(Y - \mathbb{E}[Y] \mid X - \mathbb{E}[X]\right) = \mathbb{E}\left(Y \mid X - \mathbb{E}[X]\right) - \mathbb{E}[Y]$. Donc

$$\mathbb{E}\left(Y \mid X - \mathbb{E}[X]\right) = \mathbb{E}[Y] + \Sigma_{YX}K_X^{-1}(X - \mathbb{E}[X]).$$

Enfin, comme la tribu engendrée par $X - \mathbb{E}[X]$ est la tribu $\sigma(X)$ engendrée par X (car $\mathbb{E}[X]$ est une constante), on a $\mathbb{E}\left(Y \mid X - \mathbb{E}[X]\right) = \mathbb{E}[Y \mid X]$. Il vient

$$\mathbb{E}[Y \mid X] = \mathbb{E}[Y] + \Sigma_{YX}K_X^{-1}(X - \mathbb{E}[X]).$$

Solution de E. XII.1.3 Dans le cas où $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, on a $Y - \mathbb{E}[Y \mid X] = U$ donc l'indépendance de $Y - \mathbb{E}[Y \mid X]$ et X résulte de celle de U et X.

En appliquant ce résultat à $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$, on vérifie qu'il reste valable lorsque $\mathbb{E}[X] \neq 0$ ou $\mathbb{E}[Y] \neq 0$.

Solution de E. XII.2.1 Si X et Y sont indépendantes, alors X et $g(Y)$ le sont également. On en déduit

$$\mathbb{E}[g(Y) \mid X] = \mathbb{E}[g(Y)] \quad \mathbb{P} - \text{p.s.}$$

Réciproquement, si cette égalité est vérifiée pour toute fonction g alors pour toute application $f : \mathbb{R} \to \mathbb{R}$ borélienne bornée, on a

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}\left(f(X)\mathbb{E}[g(Y) \mid X]\right) = \mathbb{E}\left(f(X)\mathbb{E}[g(Y)]\right) = \mathbb{E}[f(X)]\,\mathbb{E}[g(Y)].$$

Cette relation étant vraie pour toutes applications boréliennes f et g, elle implique l'indépendance entre X et Y (prendre $f = \mathbb{1}_A$ et $g = \mathbb{1}_B$ pour $A, B \in \mathcal{B}(\mathbb{R})$).

Solution de E. XII.2.2 Vérifions que l'application p est bien une densité de probabilité de $\mathbb{R}^2$. Elle est visiblement borélienne positive et on calcule

$$\begin{aligned}\int_{\mathbb{R}^2} p(x,y)\,\lambda(dx,dy) &= \int_0^{+\infty} e^{-y} \left(\int_0^y dx\right) dy \\ &= \int_0^{+\infty} ye^{-y}\,dy = [-ye^{-y}]_0^{+\infty} + \int_0^{+\infty} e^{-y}\,dy = 1,\end{aligned}$$

en utilisant le théorème de Tonelli et en intégrant par parties.

La densité marginale de X est alors

$$p_X(x) = \int_{\mathbb{R}} p(x,y)\,\lambda(dy) = \mathbb{1}_{\mathbb{R}_+^*}(x) \int_x^{+\infty} e^{-y}\,dy = \mathbb{1}_{\mathbb{R}_+^*}(x)\,e^{-x}.$$

La v.a. X suit donc une loi exponentielle de paramètre 1.

La densité conditionnelle de $Y \mid X = x$ (pour $x > 0$) vaut

$$f_{Y|X=x}(y) = \frac{p(x,y)}{p_X(x)} = e^{-(y-x)}\mathbb{1}_{x<y}(y).$$

Pour toute application $g : \mathbb{R} \to \mathbb{R}$ borélienne bornée, on a

$$\begin{aligned}\mathbb{E}[g(Y-X) \mid X = x] &= \int_{\mathbb{R}} g(y-x) f_{Y|X=x}(y)\,\lambda(dy) \\ &= \int_{\mathbb{R}} g(y-x)\,e^{-(y-x)}\mathbb{1}_{x<y}(y)\,\lambda(dy).\end{aligned}$$

D'où par un changement de variable,

$$\mathbb{E}[g(Y-X) \mid X = x] = \int_{\mathbb{R}_+} g(z)\,e^{-z}\,\lambda(dz).$$

Comme $\mathbb{E}[g(Y-X) \mid X=x]$ est indépendante de x, la variable aléatoire $\mathbb{E}[g(Y-X) \mid X]$ est constante presque surement. Elle est donc égale presque sûrement à son espérance

$$\mathbb{E}[g(Y-X) \mid X] = \mathbb{E}[g(Y-X)] \quad \text{p.s.}$$

La question 1) s'applique donc, ce qui montre l'indépendance entre $Y-X$ et X.

Solution de E. XII.3.1 Soit A un borélien de $\mathbb{R}$. Alors

$$\begin{aligned} \{Y \in A\} &= \bigcup_{n \in \mathbb{N}} \left\{ \sum_{i=1}^{n} X_i \in A,\ N=n \right\} \\ &= \bigcup_{n \in \mathbb{N}} \left\{ \sum_{i=1}^{n} X_i \in A \right\} \cap \{N=n\}. \end{aligned}$$

Or pour tout n fixé, on sait que $\sum_{i=1}^{n} X_i$ est une v.a. Donc l'expression précédente est une union dénombrable d'évènements, c'est donc un évènement. Ainsi Y est une v.a.

De plus, en écrivant Y sous la forme

$$Y(\omega) = \sum_{n=1}^{+\infty} \sum_{i=1}^{n} X_i(\omega) \mathbb{1}_{\{N(\omega)=n\}} \; p.s.,$$

on a $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ car

$$\begin{aligned} \mathbb{E}[|Y|] \leq \mathbb{E}\left(\sum_{n=1}^{+\infty} \sum_{i=1}^{n} |X_i| \mathbb{1}_{\{N=n\}} \right) &= \sum_{n=1}^{+\infty} \sum_{i=1}^{n} \mathbb{E}\left(|X_i| \mathbb{1}_{\{N=n\}} \right) \\ &= \sum_{n=1}^{+\infty} \sum_{i=1}^{n} \mathbb{E}\left(|X_i| \right) \, \mathbb{P}(N=n) \\ &= \mathbb{E}[|X_1|] \sum_{n=1}^{+\infty} n \, \mathbb{P}(N=n) = \mathbb{E}[|X_1|] \, \mathbb{E}[N] < +\infty. \end{aligned}$$

Par le même calcul, on montre que $\mathbb{E}[Y] = \mathbb{E}[X_1] \, \mathbb{E}[N]$.

On peut également calculer $\mathbb{E}[Y]$ par

$$\begin{aligned} \mathbb{E}[Y] = \mathbb{E}\left(\mathbb{E}[Y \mid N] \right) &= \sum_{n} \mathbb{E}[Y \mid N=n] \, \mathbb{P}(N=n) \\ &= \sum_{n} n \, \mathbb{E}[X_1] \, \mathbb{P}(N=n) = \mathbb{E}[X_1] \, \mathbb{E}[N]. \end{aligned}$$

Remarquons que $\mathbb{E}[Y \mid N=n] = n \, \mathbb{E}[X_1]$ résulte de l'Exercice XII.2.