Université d'Artois
Faculté des Sciences Jean Perrin
Probabilités (Master 1 Mathématiques-Informatique)
Daniel Li

Chapitre 4

Espérances conditionnelles et martingales

1 Espérances conditionnelles

Cette notion sert à modéliser la réponse à la question suivante : si X est une *v.a.r.* liée à une certaine expérience, que sait-on d'elle si l'on n'a pas toute l'information (donnée par la tribu $\mathscr{A}$ des événements, mais seulement une information partielle (donnée par une sous-tribu $\mathscr{B}$?

1.1 Définition

Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace de probabilité, et soit $\mathscr{B}$ une sous-tribu de $\mathscr{A}$; on veut définir, à partir d'une variable aléatoire réelle $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$, une autre *v.a.r.*, qui va "oublier" tout se qui se passe en dehors de $\mathscr{B}$.

Notons $\mathbb{P}_{|\mathscr{B}}$ la restriction à la sous-tribu $\mathscr{B}$ de la probabilité $\mathbb{P}$, et considérons l'espace $L^1(\Omega, \mathscr{B}, \mathbb{P}_{|\mathscr{B}})$.

Pour tout $B \in \mathscr{B}$, on pose :

$$\nu_X(B) = \mathbb{E}\,(X \mathbb{1}_B) = \int_B X \, d\mathbb{P} \, . \tag{1.1}$$

On obtient une mesure (réelle) sur $(\Omega, \mathscr{B})$, qui est visiblement absolument continue par rapport à $\mathbb{P}_{|\mathscr{B}}$:

$$\mathbb{P}_{|\mathscr{B}}(B) = \mathbb{P}\,(B) = 0 \quad \Longrightarrow \quad \nu_X(B) = 0 \, .$$

Le Théorème de Radon-Nikodym assure donc l'existence d'une variable aléatoire <u>$\mathscr{B}$-mesurable</u> *unique* $Y_X \in L^1(\Omega, \mathscr{B}, \mathbb{P}_{|\mathscr{B}})$ telle que $\nu_X = Y_X.\mathbb{P}_{|\mathscr{B}}$. On dit que Y_X est l'*espérance conditionnelle* de X par rapport à $\mathscr{B}$.

Dans toute la suite, on écrira simplement $\mathbb{P}$ au lieu de $\mathbb{P}_{|\mathscr{B}}$; on a donc :

Définition 1.1 *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace de probabilité, et soit $\mathscr{B}$ une sous-tribu de $\mathscr{A}$. Pour toute v.a.r. $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$, on appelle* **espérance conditionnelle** de X par rapport à $\mathscr{B}$, *ou* sachant $\mathscr{B}$, *l'unique v.a.r., notée $\mathbb{E}^{\mathscr{B}}(X)$, qui est* $\mathscr{B}$-mesurable *et qui vérifie :*

$$\boxed{\int_B \mathbb{E}^{\mathscr{B}}(X)\, d\mathbb{P} = \int_B X\, d\mathbb{P}}\,, \quad \boxed{\forall B \in \mathscr{B}}\,.$$

Ecrit autrement : $\boxed{\mathbb{E}\left[\mathbb{E}^{\mathscr{B}}(X)\mathbb{1}_B\right] = \mathbb{E}\left(X\mathbb{1}_B\right)}$.

On note aussi l'espérance conditionnelle $\mathbb{E}\left(X \mid \mathscr{B}\right)$.

Il résulte de la définition, en prenant $B = \Omega$ que l'on a :

Proposition 1.2 *Pour toute v.a.r. intégrable X, on a* $\boxed{\mathbb{E}\left[\mathbb{E}^{\mathscr{B}}(X)\right] = \mathbb{E}\left(X\right)}$.

1.2 Exemples.

a) Si $\mathscr{B} = \mathscr{A}$, il est clair que $\mathbb{E}^{\mathscr{A}}(X) = X$. Cela s'interprète en disant que, puisque l'on a toute l'information, on sait tout sur X.

Plus généralement, *si X est $\mathcal{B}$-mesurable, on a* : $\boxed{\mathbb{E}^{\mathscr{B}}(X) = X}$.

b) Si $\mathscr{B} = \{\emptyset, \Omega\}$ est la tribu grossière, alors les seules *v.a.r.* qui sont $\{\emptyset, \Omega\}$-mesurables sont les constantes. De par l'égalité $\mathbb{E}\left[\mathbb{E}^{\mathscr{B}}(X)\right] = \mathbb{E}\left(X\right)$, cette constante ne peut être que $\mathbb{E}\left(X\right)$; donc $\mathbb{E}^{\{\emptyset,\Omega\}}(X) = \mathbb{E}\left(X\right)\mathbb{1}$.

Cela s'interprète en disant que, puisque l'on ne dispose d'aucune information, tout ce que l'on peut savoir sur X est sa valeur moyenne.

c) Si $\mathscr{B}$ est la *tribu engendrée par une partition* $(B_n)_{n\geqslant 1}$ (finie ou infinie) de Ω, formée de parties $\mathscr{A}$-mesurables deux-à-deux disjointes et telles que $\mathbb{P}\left(B_n\right) \neq 0$, comme les *v.a.r.* $\mathscr{B}$-mesurables sont celles qui sont *constantes sur chaque* B_n, on a, si a_n désigne cette constante :

$$\mathbb{E}^{\mathscr{B}}(X) = \sum_{n\geqslant 1} a_n\, \mathbb{1}_{B_n}\,;$$

d'où :

$$\int_{B_n} X\, d\mathbb{P} = \int_{B_n} \mathbb{E}^{\mathscr{B}}(X)\ , d\mathbb{P} = \int_{B_n} a_n\, d\mathbb{P} = a_n \mathbb{P}\left(B_n\right).$$

Donc :

$$\boxed{\mathbb{E}^{\mathscr{B}}(X) = \sum_n \left(\frac{1}{\mathbb{P}\left(B_n\right)} \int_{B_n} X\, d\mathbb{P}\right) \mathbb{1}_{B_n}}\,;$$

ainsi, sur chaque B_n, X n'est connu que par sa moyenne sur B_n. En particulier, si $\mathscr{B} = \{\emptyset, B, B^c, \Omega\}$, pour un $B \in \mathscr{A}$ tel que $0 < \mathbb{P}(B) < 1$, alors :

$$\begin{aligned}\mathbb{E}^{\mathscr{B}}(\mathbb{1}_A) &= \frac{1}{\mathbb{P}(B)}\mathbb{P}(A \cap B)\mathbb{1}_B + \frac{1}{\mathbb{P}(B^c)}\mathbb{P}(A \cap B^c)\mathbb{1}_{B^c} \\ &= \mathbb{P}(A \mid B)\mathbb{1}_B + \mathbb{P}(A \mid B^c)\mathbb{1}_{B^c}\,,\end{aligned}$$

où $\mathbb{P}(A \mid B)$ est la *probabilité conditionnelle* de A sachant B.

1.3 Propriétés

Proposition 1.3 *Soit* $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$. *Pour toute v.a.r.* Z $\mathscr{B}$-mesurable et bornée *(*$Z \in L^\infty(\Omega, \mathscr{B}, \mathbb{P})$*), on a :*

$$\boxed{\mathbb{E}^{\mathscr{B}}(ZX) = Z\,\mathbb{E}^{\mathscr{B}}(X)} \qquad \textit{(propriété d'idéal).}$$

Corollaire 1.4 *Pour toute* $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$ *et toute* $Z \in L^\infty(\Omega, \mathscr{B}, \mathbb{P})$*, on a :*

$$\boxed{\int_\Omega Z\,\mathbb{E}^{\mathscr{B}}(X)\,d\mathbb{P} = \int_\Omega ZX\,d\mathbb{P}}\,.$$

En effet, ce n'est autre que l'égalité $\mathbb{E}\left[\mathbb{E}^{\mathscr{B}}(ZX)\right] = \mathbb{E}(ZX)$.

Preuve de la Proposition 1.3. a) Il suffit en fait de montrer l'égalité du Corollaire 1.4, car, en remplaçant dedans Z par $Z\mathbb{1}_B$, avec $B \in \mathscr{B}$, qui est encore $\mathscr{B}$-mesurable, on aura :

$$\int_B ZX\,d\mathbb{P} = \int_B Z\,\mathbb{E}^{\mathscr{B}}(X)\,d\mathbb{P}\;;$$

mais comme $Z\,\mathbb{E}^{\mathscr{B}}(X)$ est $\mathscr{B}$-mesurable, la définition (et l'unicité) de $\mathbb{E}^{\mathscr{B}}(ZX)$ donne bien :

$$\mathbb{E}^{\mathscr{B}}(ZX) = Z\,\mathbb{E}^{\mathscr{B}}(X)\,.$$

b) Or l'égalité :

$$\int_\Omega ZX\,d\mathbb{P} = \int_\Omega Z\,\mathbb{E}^{\mathscr{B}}(X)\,d\mathbb{P}$$

est valable, par définition, pour les $Z = \mathbb{1}_B$, avec $B \in \mathscr{B}$; elle est donc aussi valable pour les *v.a.r.* étagées $\mathscr{B}$-mesurables. Par convergence dominée, puisque $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$, elle est ensuite valable pour toutes les Z $\mathscr{B}$-mesurables bornées. □

Proposition 1.5 *L'espérence conditionnelle :*

$$\mathbb{E}^{\mathscr{B}}\colon L^1(\Omega,\mathscr{A},\mathbb{P}) \to L^1(\Omega,\mathscr{B},\mathbb{P})$$

est une application linéaire, continue*, de norme 1,* positive*, et idempotente (projecteur).*

Remarque. L'utilisation du terme "idempotent" laisse sous-entendre que l'espace $L^1(\Omega,\mathscr{B},\mathbb{P})$ est contenu dans $L^1(\Omega,\mathscr{A},\mathbb{P})$; ce n'est en fait pas le cas.

En effet, il y a là une petite *difficulté*. Rappelons éléments de $L^1(\Omega,\mathscr{A},\mathbb{P})$ ne sont pas réellement des fonctions, mais des *classes de fonctions* (modulo l'égalité presque sûre), même si d'habitude on ne fait pas de distinction entre la fonction et sa classe d'équivalence.

Maintenant, si X est une variable aléatoire $\mathscr{B}$-mesurable et si X' est une variable aléatoire ($\mathscr{A}$-mesurable), qui est presque sûrement égale à X : $\mathbb{P}(X' \neq X) = 0$, alors il n'y a aucune raison que X' soit elle aussi $\mathscr{B}$-mesurable (sauf si la tribu $\mathscr{B}$ est $\mathbb{P}$-complète). En notant $\mathbb{P}_{|\mathscr{B}}$ la restriction à $\mathscr{B}$ de la probabilité $\mathbb{P}$ (définie sur $\mathscr{A}$), on doit donc distinguer :

a) la classe d'équivalence $\mathbb{P}$-*p.s.* sur $(\Omega,\mathscr{A},\mathbb{P})$ de X

et :

b) sa classe d'équivalence $\mathbb{P}_{|\mathscr{B}}$-*p.s.* sur $(\Omega,\mathscr{B},\mathbb{P}_{|\mathscr{B}})$,

la première étant en général strictement plus grande que la seconde.

En d'autres termes, bien que, pour les espaces de *fonctions*, on ait :

$$\mathscr{L}^r(\Omega,\mathscr{B},\mathbb{P}_{|\mathscr{B}}) \subseteq \mathscr{L}^r(\Omega,\mathscr{A},\mathbb{P}),$$

pour $1 \leqslant r \leqslant +\infty$, par contre l'espace des *classes* de fonctions $\mathscr{B}$-mesurables $L^r(\Omega,\mathscr{B},\mathbb{P}_{|\mathscr{B}})$ n'est **pas** contenu dans l'espace des classes de fonctions $\mathscr{A}$-mesurables $L^r(\Omega,\mathscr{A},\mathbb{P})$.

Néanmoins, l'application qui à la $\mathbb{P}_{|\mathscr{B}}$-classe de X fait correspondre sa $\mathbb{P}$-classe est injective, et définit une **isométrie** :

$$J\colon L^r(\Omega,\mathscr{B},\mathbb{P}_{|\mathscr{B}}) \longrightarrow L^r(\Omega,\mathscr{A},\mathbb{P}).$$

Par cette isométrie, on peut donc *identifier* isométriquement $L^r(\Omega,\mathscr{B},\mathbb{P}_{|\mathscr{B}})$ à un sous-espace (fermé) de $L^r(\Omega,\mathscr{A},\mathbb{P})$, *ce que l'on fera toujours par la suite.*

Preuve de la Proposition 1.5. a) La linéarité est facile, par unicité.

b) Montrons que $\boxed{X \geqslant 0 \ \Rightarrow\ \mathbb{E}^{\mathscr{B}}(X) \geqslant 0}$.

En effet, si $X \geqslant 0$, la mesure ν_X définie en 1.1 est positive ; donc $\mathbb{E}^{\mathscr{B}}(X) \geqslant 0$, puisque $\nu_X = \mathbb{E}^{\mathscr{B}}(X).\mathbb{P}$

c) On a alors : $\boxed{|\mathbb{E}^{\mathscr{B}}(X)| \leqslant \mathbb{E}^{\mathscr{B}}(|X|)}$, car $|X| \pm X \geqslant 0$.

Donc, puisque $\mathbb{E}\left[\mathbb{E}^{\mathscr{B}}(|X|)\right] = \mathbb{E}(|X|) = \|X\|_1$, on obtient :

$$\|\mathbb{E}^{\mathscr{B}}(X)\|_1 \leqslant \|X\|_1 \,.$$

d) Mais $\boxed{\mathbb{E}^{\mathscr{B}}(\mathbb{1}) = \mathbb{1}}$ et donc la norme est exactement 1.

e) Pour finir : $\mathbb{E}^{\mathscr{B}}\big[\mathbb{E}^{\mathscr{B}}(X)\big] = \mathbb{E}^{\mathscr{B}}(X)$ car $\mathbb{E}^{\mathscr{B}}(X) \in L^1(\mathscr{B}) \subseteq L^1(\mathscr{A})$ et que $\mathbb{E}^{\mathscr{B}}(X)$ est $\mathscr{B}$-mesurable. $\square$

Proposition 1.6

1) *Pour* $1 \leqslant r \leqslant \infty$ *:*

$$X \in L^r(\Omega, \mathscr{A}, \mathbb{P}) \quad \Rightarrow \quad \mathbb{E}^{\mathscr{B}}(X) \in L^r(\Omega, \mathscr{B}, \mathbb{P}).$$

2) *Pour* $r = 2$, $\mathbb{E}^{\mathscr{B}}$ *est la* **projection orthogonale** *de* $L^2(\Omega, \mathscr{A}, \mathbb{P})$ *sur* $L^2(\Omega, \mathscr{B}, \mathbb{P})$.

On utilisera le lemme suivant.

Lemme 1.7 *Si* $X \in L^r(\Omega, \mathscr{A}, \mathbb{P})$*, avec* $1 \leqslant r < +\infty$*, alors :*

$$\boxed{|\mathbb{E}^{\mathscr{B}}(X)|^r \leqslant \mathbb{E}^{\mathscr{B}}(|X|^r)}.$$

Preuve de la Proposition 1.6. 1) Pour $1 \leqslant r < \infty$, cela résulte du lemme. Pour $r = \infty$:

$$|X| \leqslant \|X\|_\infty \mathbb{1} \quad \Rightarrow \quad |\mathbb{E}^{\mathscr{B}}(X)| \leqslant \mathbb{E}^{\mathscr{B}}(|X|) \leqslant \mathbb{E}^{\mathscr{B}}(\|X\|_\infty \mathbb{1}) = \|X\|_\infty .$$

2) Pour $r = 2$, on doit vérifier que $X - \mathbb{E}^{\mathscr{B}}(X) \perp Z$ pour toute $Z \in L^2(\Omega, \mathscr{B}, \mathbb{P})$. Mais :

$$\int_\Omega \big[X - \mathbb{E}^{\mathscr{B}}(X)\big] Z \, d\mathbb{P} = \int_\Omega XZ \, d\mathbb{P} - \int_\Omega \mathbb{E}^{\mathscr{B}}(X) Z \, d\mathbb{P} = 0$$

pour toute $Z \in L^\infty(\Omega, \mathscr{B}, \mathbb{P})$, donc pour toute $Z \in L^2(\Omega, \mathscr{B}, \mathbb{P})$ par densité. $\square$

Remarque. Certains auteurs définissent directement, pour $X \in L^2(\Omega, \mathscr{A}, \mathbb{P})$, l'espérance conditionnelle de X sachant $\mathscr{B}$ comme la projection orthogonale de X sur $L^2(\Omega, \mathscr{B}, \mathbb{P})$, puis prolongent l'application $\mathbb{E}^{\mathscr{B}}$ à $L^1(\Omega, \mathscr{A}, \mathbb{P})$, par densité. C'est plus élémentaire, mais cela fait complètement perdre de vue le sens réel de l'espérance conditionnelle.

Preuve du Lemme 1.7. On peut supposer $r > 1$.

Pour toute $Z \in L^\infty(\Omega, \mathscr{B}, \mathbb{P})$ et tout $B \in \mathscr{B}$, on a :

$$\begin{aligned} \int_\Omega |Z \mathbb{1}_B \mathbb{E}^{\mathscr{B}}(X)| \, d\mathbb{P} &= \int_\Omega |\mathbb{E}^{\mathscr{B}}(Z \mathbb{1}_B X)| \, d\mathbb{P} \leqslant \int_\Omega \mathbb{E}^{\mathscr{B}}(|Z \mathbb{1}_B X|) \, d\mathbb{P} \\ &= \int_\Omega |Z \mathbb{1}_B X| \, d\mathbb{P} \leqslant \|Z\|_s \|\mathbb{1}_B X\|_r < +\infty , \end{aligned}$$

avec $\frac{1}{r} + \frac{1}{s} = 1$. Comme $L^\infty(\Omega, \mathscr{B}, \mathbb{P})$ est dense dans $L^s(\Omega, \mathscr{B}, \mathbb{P})$, cela signifie que :

$$Z \longmapsto \int_\Omega Z \mathbb{1}_B \mathbb{E}^{\mathscr{B}}(X) \, d\mathbb{P}$$

est une forme linéaire continue sur $L^s(\Omega, \mathscr{B}, \mathbb{P})$, de norme $\leqslant \|\mathbb{1}_B X\|_r$. Donc $\mathbb{1}_B \mathbb{E}^{\mathscr{B}}(X) \in L^r(\Omega, \mathscr{B}, \mathbb{P})$, et :

$$\|\mathbb{1}_B \mathbb{E}^{\mathscr{B}}(X)\|_r \leqslant \|\mathbb{1}_B X\|_r\,;$$

cela s'écrit aussi :

$$\int_B |\mathbb{E}^{\mathscr{B}}(X)|^r\, d\mathbb{P} \leqslant \int_B |X|^r\, d\mathbb{P} = \int_B \mathbb{E}^{\mathscr{B}}(|X|^r)\, d\mathbb{P}\,.$$

Mais, puisque c'est vrai pour tout $B \in \mathscr{B}$, cela entraîne :

$$|\mathbb{E}^{\mathscr{B}}(X)|^r \leqslant \mathbb{E}^{\mathscr{B}}(|X|^r)\,. \qquad \square$$

Voyons maintenant une propriété "d'emboîtement", que l'on a déjà vue dans le cas particulier de $\mathscr{B}_2 = \{\emptyset, \Omega\}$; elle s'écrit alors $\mathbb{E}\left[\mathbb{E}^{\mathscr{B}_1}(X)\right] = \mathbb{E}(X)\mathbb{1}$.

Proposition 1.8 (transitivité) *Si $\mathscr{B}_2 \subseteq \mathscr{B}_1 \subseteq \mathscr{A}$, on a :*

$$\boxed{\mathbb{E}^{\mathscr{B}_2}\left[\mathbb{E}^{\mathscr{B}_1}(X)\right] = \mathbb{E}^{\mathscr{B}_2}(X)}\,.$$

Preuve. Si $B \in \mathscr{B}_2$:

$$\int_B \mathbb{E}^{\mathscr{B}_2}\left[\mathbb{E}^{\mathscr{B}_1}(X)\right] d\mathbb{P} = \int_B \mathbb{E}^{\mathscr{B}_1}(X)\, d\mathbb{P}\,;$$

mais comme on a aussi $B \in \mathscr{B}_1$:

$$\int_B \mathbb{E}^{\mathscr{B}_1}(X)\, d\mathbb{P} = \int_B X\, d\mathbb{P}\,. \qquad \square$$

1.4 Cas d'une tribu engendrée par une variable aléatoire

Si $\mathscr{B} = \mathscr{B}_Y$ est la tribu engendrée par une *v.a.* Y, on notera :

$$\boxed{\mathbb{E}(X \mid Y)}$$

au lieu de $\mathbb{E}^{\mathscr{B}_Y}(X) = \mathbb{E}(X \mid \mathscr{B}_Y)$. On dit que c'est l'*espérance conditionnelle de X sachant Y*.

Rappelons que l'on a :

$$\mathscr{B}_Y = \{Y^{-1}(D)\,;\ D \in \mathscr{B}or(\mathbb{R}^d)\}\,.$$

Rappelons aussi que toute *v.a.* $\mathscr{B}_Y$-mesurable s'écrit comme la composée de Y et d'une fonction borélienne sur $\mathbb{R}^d$. Donc :

Proposition 1.9 *Pour toute v.a.r. $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$ et toute v.a. Y, il existe une fonction borélienne $h\colon \mathbb{R}^d \to \mathbb{R}$, $\mathbb{P}_Y$-intégrable, telle que :*

$$\mathbb{E}(X \mid Y) = h(Y)\,.$$

Par définition, pour tout borélien $D \in \mathscr{B}or(\mathbb{R}^d)$, on a donc :

$$\int_{Y^{-1}(D)} X \, d\mathbb{P} = \int_{Y^{-1}(D)} \mathbb{E}\,(X \mid Y)\, d\mathbb{P} = \int_{Y^{-1}(D)} h(Y)\, d\mathbb{P}\,,$$

ce qui s'écrit :

$$\boxed{\int_{Y\in D} X \, d\mathbb{P} = \int_D h(y)\, d\mathbb{P}_Y(y)}\,, \quad \forall D \in \mathscr{B}or(\mathbb{R}^d)\,. \tag{1.2}$$

1.5 Exemples de calcul d'espérances conditionnelles par rapport à une v.a.r.

1.5.1 Cas où Y est une v.a.r. discrète

On a :

$$\mathbb{P}_Y = \sum_{n\geqslant 1} a_n \delta_{x_n}\,, \quad \text{avec } a_n \neq 0\,,$$

la somme étant finie ou infinie.

Puisque $\mathbb{P}\,(Y \in D) = 0$ si $D \cap \{x_n\,;\ n \geqslant 1\} = \emptyset$, la formule (1.2) ci-dessus montre que l'on peut prendre $h(y) = 0$ si $y \notin \{x_n\,;\ n \geqslant 1\}$, et :

$$h(x_n) = \frac{1}{a_n} \int_{\{Y=x_n\}} X\, d\mathbb{P} = \frac{1}{\mathbb{P}\,(Y = x_n)} \int_{\{Y=x_n\}} X\, d\mathbb{P}\,.$$

On obtient :

$$\mathbb{E}\,(X \mid Y) = \sum_{n\geqslant 1} \left(\frac{1}{\mathbb{P}\,(Y = x_n)} \int_{\{Y=x_n\}} X\, d\mathbb{P} \right) \mathbb{1}_{\{Y=x_n\}}\,,$$

ce que l'on savait déjà, puisque $\mathscr{B}_Y$ est engendrée par la partition de Ω formée par les $\{Y = x_n\}$, $n \geqslant 1$.

1.5.2 Cas de variables à densité

Supposons que le **couple** (X, Y) possède une densité $f_{(X,Y)}$ sur $\mathbb{R}^{d+1}$.

La loi $\mathbb{P}_Y$ de Y a alors une densité (*densité marginale*), donnée par :

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dx\,.$$

Comme $f_Y \in L^1(\mathbb{R}^d)$, f_Y est finie λ_d-presque partout ; par conséquent, Y est presque sûrement à valeurs dans :

$$Q = \{y \in \mathbb{R}^d\,;\ 0 < f_Y(y) < +\infty\}\,.$$

Pour $y \in Q$, on peut poser :

$$\boxed{f_X(x \mid y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}\,.}$$

La fonction :

$$f_X(\,.\mid y)\colon x \longmapsto f_X(x \mid y)$$

est une *densité de probabilité* sur $\mathbb{R}$: on a $f_X(x \mid y) \geqslant 0$, et :

$$\int_{\mathbb{R}} f_X(x \mid y)\, dx = \frac{1}{f_Y(y)} \int_{\mathbb{R}} f_{(X,Y)}(x,y)\, dx = 1\,.$$

Définition 1.10 *On dit que $f_X(\,.\mid y)$ est la* **densité conditionnelle** *de X sachant $Y = y$.*

Il faut faire attention que $Y = y$ n'est qu'une notation ; en effet, $\mathbb{P}\,(Y = y) = 0$ pour tout $y \in \mathbb{R}^d$ puisque Y possède une densité.

On note parfois aussi :

$$f_X(x \mid y) = f_X(x \mid Y = y)$$

et :

$$f_X(x \mid y)\, dx = \mathbb{P}\,(x \mid Y = y)\,.$$

Proposition 1.11 *Si le vecteur aléatoire (X,Y) a une densité sur $\mathbb{R}^{d+1}$, alors pour toute $g \in L^1(\mathbb{R}, \mathbb{P}_X)$, on a :*

$$\boxed{\mathbb{E}\big(g(X) \mid Y\big) = h(Y)}\,,$$

avec, pour tout pour $y \in Q$:

$$\boxed{h(y) = \int_{\mathbb{R}} g(x) f_X(x \mid y)\, dx = \frac{1}{f_Y(y)} \int_{\mathbb{R}} g(x) f_{(X,Y)}(x,y)\, dx}\,.$$

Preuve. Nous avons vu que :

$$\mathbb{E}\big(g(X) \mid Y\big) = h(Y)\,,$$

avec :

$$\int_D h(y)\, d\mathbb{P}_Y(y) = \int_{Y \in D} g(X)\, d\mathbb{P}\,, \quad \forall D \in \mathscr{B}or(\mathbb{R}^d)\,.$$

Comme $\mathbb{P}_Y(Q) = 1$, on peut poser $f_X(x \mid y) = 0$ pour $y \in \mathbb{R}^d \setminus Q$. On a alors, par le Théorème de Fubini :

$$\begin{aligned}
\int_D \left[\int_R g(x) f_X(x \mid y)\, dx\right] d\mathbb{P}_Y(y) &= \int_D \left[\int_{\mathbb{R}} g(x) f_X(x \mid y)\, dx\right] f_Y(y)\, dy \\
&= \iint_{\mathbb{R}\times D} g(x) f_X(x \mid y) f_Y(y)\, dx dy \\
&= \iint_{\mathbb{R}\times D} g(x) f_{(X,Y)}(x,y)\, dx dy \\
&= \iint_{\mathbb{R}\times D} g(x)\, d\mathbb{P}_{(X,Y)}(x,y) \\
&= \iint_{\mathbb{R}^{d+1}} g(x) \mathbb{1}_D(y)\, d\mathbb{P}_{(X,Y)}(x,y) \\
&= \int_\Omega g(X) \mathbb{1}_D(Y)\, d\mathbb{P} = \int_{Y\in D} g(x)\, d\mathbb{P} \quad \square
\end{aligned}$$

1.5.3 Cas gaussien

Nous allons voir qu'alors $\mathbb{E}(X \mid Y)$ est une fonction *affine* de Y, et pas seulement une fonction borélienne.

Théorème 1.12 *Si $(X,Y)\colon \Omega \to \mathbb{R}^{n+1}$ est un* **vecteur gaussien***, avec $Y = (Y_1,\dots,Y_n)$, alors $\mathbb{E}(X \mid Y)$ est une v.a.r.* **gaussienne** *et il existe des nombres $a_1,\dots,a_n, b \in \mathbb{R}$ tels que :*

$$\boxed{\mathbb{E}(X \mid Y) = a_1 Y_1 + \cdots + a_n Y_n + b}\,.$$

On utilisera le lemme suivant, qui a d'ailleurs son intérêt propre.

Lemme 1.13 *Les tribus $\mathscr{B}_1$ et $\mathscr{B}_2$ sont indépendantes si et seulement si :*

$$\mathbb{E}^{\mathscr{B}_2}(U) = \mathbb{E}(U)\mathbb{1}$$

pour toute v.a.r. U $\mathscr{B}_1$-mesurable.

Preuve. S'il y a indépendance, on a, pour tout $B \in \mathscr{B}_2$:

$$\int_B U\, d\mathbb{P} = \mathbb{E}(\mathbb{1}_B U) = \mathbb{E}(\mathbb{1}_B)\mathbb{E}(U) = \int_B \mathbb{E}(U)\mathbb{1}\, d\mathbb{P}\,,$$

d'où $\mathbb{E}^{\mathscr{B}_2}(U) = \mathbb{E}(U)\mathbb{1}$.

Inversement, si $B_1 \in \mathscr{B}_1$ et $B_2 \in \mathscr{B}_2$:

$$\begin{aligned}
\mathbb{P}(B_1 \cap B_2) = \int_{B_2} \mathbb{1}_{B_1}\, d\mathbb{P} &= \int_{B_2} \mathbb{E}^{\mathscr{B}_2}(\mathbb{1}_{B_1})\, d\mathbb{P} \\
&= \mathbb{E}(\mathbb{1}_{B_1}) \int_{B_2} \mathbb{1}\, d\mathbb{P} = \mathbb{P}(B_1)\,\mathbb{P}(B_2)\,,
\end{aligned}$$

d'où l'indépendance de B_1 et B_2, et donc de $\mathscr{B}_1$ et $\mathscr{B}_2$. $\square$

Preuve du théorème. Supposons d'abord $\mathbb{E}(X) = \mathbb{E}(Y) = 0$.

Soit $\mathscr{G}$ l'espace vectoriel engendré par $Y_1, \ldots, Y_n$. C'est un sous-espace (de dimension finie) de $L^2(\Omega, \mathscr{A}, \mathbb{P})$. De plus, comme le vecteur $Y = (Y_1, \ldots, Y_n)$ est gaussien, $\mathscr{G}$ est entièment composé de gaussiennes (c'est un espace gaussien).

Rappelons que, puisque l'on est dans $L^2(\Omega, \mathscr{A}, \mathbb{P})$, l'espérance conditionnelle est la projection orthogonale de $L^2(\Omega, \mathscr{A}, \mathbb{P})$ sur $L^2(\Omega, \mathscr{B}_Y, \mathbb{P})$. Remarquons par ailleurs que $\mathscr{G} \subseteq L^2(\Omega, \mathscr{B}_Y, \mathbb{P})$.

Soit $P(X)$ la projection orthogonale de X sur $\mathscr{G}$.

On a $P(X) = a_1 Y_1 + \cdots + a_n Y_n$ pour des nombres réels $a_1, \ldots, a_n \in \mathbb{R}$.

Posons $Z = X - P(X)$.

Le vecteur (Z, Y) est gaussien, comme image linéaire de (X, Y), et $\mathbb{E}(ZY_k) = 0$ pour tout $k = 1, \ldots, n$, car Z est orthogonal à $\mathscr{G}$; mais si les composantes d'un vecteur gaussien sont non corrélées, elles sont indépendantes. Donc Z est indépendant de $Y_1, \ldots, Y_n$. Alors, par le lemme, $\mathbb{E}(Z \mid Y) = \mathbb{E}(Z)\,\mathbb{1} = 0$, car $\mathbb{E}(Z) = a_1 \mathbb{E}(Y_1) + \cdots + a_n \mathbb{E}(Y_n) = 0$, ce qui donne $\mathbb{E}\big(X \mid Y\big) = P(X)$, car $\mathbb{E}\big(P(X) \mid Y\big) = P(X)$ puisque $P(X) \in \mathscr{G} \subseteq L^2(\Omega, \mathscr{B}_Y, \mathbb{P})$. On peut donc bien écrire $\mathbb{E}(X \mid Y) = a_1 Y_1 + \cdots + a_n Y_n$.

Lorsque X et Y ne sont plus centrés, il suffit de poser $X' = X - \mathbb{E}(X)$ et $Y' = Y - \mathbb{E}(Y)$; comme $\mathscr{B}_{Y'} = \mathscr{B}_Y$, on a :

$$\begin{aligned}
\mathbb{E}(X \mid Y) &= \mathbb{E}(X' + \mathbb{E}(X) \mid Y') = \mathbb{E}(X' \mid Y') + \mathbb{E}(X) \\
&= \sum_{k=1}^{n} a_k Y'_k + \mathbb{E}(X) = \sum_{k=1}^{n} a_k (Y_k - \mathbb{E}(Y_k)) + \mathbb{E}(X) \\
&= \sum_{k=1}^{n} a_k Y_k + \Big(\sum_{k=1}^{n} a_k \mathbb{E}(Y_k) + \mathbb{E}(X) \Big).
\end{aligned}$$
$\square$

Calcul pratique. Pour calculer les coefficients $a_1, \ldots, a_n, b$, on part de :

$$\mathbb{E}(X \mid Y) = a_1 Y_1 + \cdots + a_n Y_n + b\,.$$

En prenant l'espérance, on obtient, puisque $\mathbb{E}[\mathbb{E}(X \mid Y)] = \mathbb{E}(X)$:

$$\mathbb{E}(X) = a_1 \mathbb{E}(Y_1) + \cdots + a_n \mathbb{E}(Y_n) + b\,.$$

De même, si on multiplie par Y_k, en utilisant $\mathbb{E}(X \mid Y) Y_k = \mathbb{E}(X Y_k \mid Y)$, on obtient :

$$\mathbb{E}(X Y_k) = \sum_{j=1}^{n} a_j \mathbb{E}(Y_j Y_k) + b \mathbb{E}(Y_k)\,.$$

Il ne reste plus qu'à résoudre ce système linéaire de $n+1$ équations à $n+1$ inconnues. Il n'y a évidemment pas unicité si $Y_1, \ldots, Y_n$ ne sont pas linéairement indépendants.

2 Martingales

Nous n'en dirons que quelques mots d'introduction. C'est une partie importante des Probabilités. Jusqu'à présent, nous n'avons parlé que de Probabilités "statiques", au sens où l'on ne s'intéressait qu'à une seule tribu d'événements (ou à deux dans le cas des espérances conditionnelles). Les martingales vont faire intervenir une notion d'"évolution" avec le temps : au fur et à mesure que le temps passe, le nombre d'informations que l'on connaît augmente ; autrement dit, on doit tenir compte de plus en plus d'événements, et l'on est amené à considérer des familles croissantes de sous-tribus d'événements ; pour mesurer une "expérience", on aura donc,à chaque instant, une variable aléatoire pour laquelle l'information que l'on a sur elle dans le passé, c'est-à-dire son espérance conditionnelle par rapport à la tribu correspondante, est la variable aléatoire connue à cet instant passé.

Nous ne parlerons par ailleurs que des *martingales à temps discret*, c'est-à-dire celle indexées par les entiers. Cela revient à regarder ce qui se passe à intervalles réguliers. Si l'on veut savoir se qui se passe à tout moment, on utilisera des *martingales à temps continu*, indexées par $\mathbb{R}_+$, par exemple.

Définition 2.1 *On appelle* **filtration** *toute suite* croissante *:*

$$\mathscr{B}_1 \subseteq \mathscr{B}_2 \subseteq \cdots \subseteq \mathscr{B}_n \subseteq \cdots \subseteq \mathscr{A}$$

de sous-tribus de $\mathscr{A}$*.*

Définition 2.2 *On appelle* **martingale** *par rapport à, ou* **adaptée à, la filtration** $(\mathscr{B}_n)_{n\geqslant 1}$*, toute suite de v.a.r.* $M_n \in L^1(\Omega, \mathscr{A}, \mathbb{P})$*,* $n \geqslant 1$*, telle que, pour tout* $n \geqslant 1$ *:*

1) M_n *est* $\mathscr{B}_n$*-mesurable ;*
2) $\mathbb{E}^{\mathscr{B}_n}(M_{n+1}) = M_n$ *.*

Voici deux exemples particulièrement importants.

Exemple 1. Soit $M \in L^1(\Omega, \mathscr{A}, \mathbb{P})$. Si l'on pose $M_n = \mathbb{E}^{\mathscr{B}_n}(M)$, alors $(M_n)_{n\geqslant 1}$ est une martingale par rapport à la filtration $(\mathscr{B}_n)_{n\geqslant 1}$, par la propriété de transitivité des espérances conditionnelles. On dit qu'une **martingale** $(M_n)_{n\geqslant 1}$ est **terminée** s'il existe $M \in L^1(\Omega, \mathscr{A}, \mathbb{P})$ telle que $M_n = \mathbb{E}^{\mathscr{B}_n}(M)$ pour tout $n \geqslant 1$.

Exemple 2. Soit $(X_n)_{n\geqslant 1}$ une suite de *v.a.r.* **indépendantes et centrées**. Si l'on pose :

$$S_n = X_1 + \cdots + X_n \,,$$

alors $(S_n)_{n\geqslant 1}$ est une martingale, adaptée à la filtration donnée par $\mathscr{B}_n = \sigma(X_1, \ldots, X_n)$.

On complète habituellement la filtration en introduisant $\mathscr{B}_0 = \{\emptyset, \Omega\}$. Alors, si l'on pose $M_0 = 0$, on définit les **accroissements de la martingale**, que l'on appelle aussi les **différences de la martingale**, par :

$$\boxed{d_n = M_n - M_{n-1} \quad n \geqslant 1}.$$

On a :

$$M_n = d_1 + \cdots + d_n .$$

Ainsi on a une situation semblable à celle de l'Exemple 2, sauf que les différences d_n, $n \geqslant 1$, ne sont plus indépendantes (en général). Néanmoins, on garde une version très affaiblie d'indépendance. En effet, dire que la suite $(X_n)_{n\geqslant 1}$ est indépendante, peut s'exprimer par :

Pour tout $n \geqslant 1$, la var X_{n+1} est indépendante de la tribu $\sigma(X_1, \ldots, X_n)$.

Pour la suite des différences d'une martingale, on a :

$$\boxed{\mathbb{E}^{\mathscr{B}_n}(d_{n+1}) = 0}$$

(puisque $\mathbb{E}^{\mathscr{B}_n}(M_{n+1}) = M_n$) ; on a donc une "orthogonalité" entre $\mathscr{B}_n$ et d_{n+1} (il y aurait vraiment orthogonalité si l'on était dans $L^2(\Omega, \mathscr{A}, \mathbb{P})$). On peut donc espérer pour la convergence des martingales des résultats analogues à ceux obtenus pour les sommes de *v.a.r.* indépendantes et centrées, et nous allons voir que c'est bien le cas.

Théorème 2.3 (inégalité maximale de Doob) *Soit $(M_n)_{n\geqslant 1}$ une martingale adaptée à la filtration $(\mathscr{B}_n)_{n\geqslant 1}$. Alors, pour tout $N \geqslant 1$ et tout $a > 0$, on a :*

$$\boxed{\mathbb{P}\left(\sup_{1\leqslant n\leqslant N} |M_n| > a\right) \leqslant \frac{1}{a}\,\mathbb{E}\,(|M_N|)}.$$

Preuve. Elle est analogue à celle de l'inégalité de Kolmogorov.

On définit le temps d'arrêt ν_a par :

$$\nu_a(\omega) = \begin{cases} +\infty & \text{si } |M_n| = |d_1 + \cdots + d_n| \leqslant a, \text{ pour tout } n = 1, 2, \ldots, N \\ \min\{n \leqslant N\,;\ |M_n| = |d_1 + \cdots + d_n| > a \text{ sinon.} \end{cases}$$

Posons $A = \left\{\sup_{n\leqslant N} |M_n| > a\right\}$, et $A_n = \{\nu_a = n\}$. On a $A_n \in \mathscr{B}_n$, et $A = \bigcup_{1\leqslant n\leqslant N} A_n$. Alors :

$$\begin{aligned} \mathbb{E}\left(\mathbb{1}_{A_n}.|M_n|\right) &= \mathbb{E}\left(\mathbb{1}_{A_n}.\mathrm{sgn}\,(M_n).M_n\right) = \mathbb{E}\left(\mathbb{1}_{A_n}.\mathrm{sgn}\,(M_n).\mathbb{E}^{\mathscr{B}_n}(M_N)\right) \\ &= \mathbb{E}\left(\mathbb{E}^{\mathscr{B}_n}\left[\mathbb{1}_{A_n}.\mathrm{sgn}\,(M_n).M_N\right]\right) = \mathbb{E}\left(\mathbb{1}_{A_n}.\mathrm{sgn}\,(M_n).M_N\right) \\ &\leqslant \mathbb{E}\,(\mathbb{1}_{A_n}.|M_N|)\,. \end{aligned}$$

Par conséquent :

$$\mathbb{P}(A) = \sum_{n=1}^{N} \mathbb{P}(A_n) \leqslant \sum_{n=1}^{N} \mathbb{E}\left(\frac{1}{a}|M_n|.\mathbb{1}_{A_n}\right) \leqslant \frac{1}{a}\sum_{n=1}^{N} \mathbb{E}(\mathbb{1}_{A_n}.|M_N|) = \frac{1}{a}\mathbb{E}(|M_N|),$$

ce qu'on voulait montrer. □

Théorème 2.4 (Théorème de Doob)
1) *Toute martingale terminée converge presque sûrement et pour la norme de* L^1.
2) *Toute martingale bornée dans* L^2 *converge presque sûrement et pour la norme de* L^2.

Pour prouver ce théorème, on aura besoin d'une propriété d'approximation.

Lemme 2.5 *Soit* $\mathscr{C} \subseteq \mathscr{A}$ *une algèbre de Boole, et soit* $A \in \sigma(\mathscr{C})$ *et* $\varepsilon > 0$. *Alors il existe* $B \in \mathscr{C}$ *tel que* $\mathbb{P}(A \Delta B) \leqslant \varepsilon$, *où* Δ *désigne la différence symétrique.*

Preuve. Remarquons que $\mathbb{P}(A\Delta B) = \|\mathbb{1}_A - \mathbb{1}_B\|_1$. Si $\mathscr{D}$ désigne l'ensemble des $A \in \sigma(\mathscr{C})$ ayant la propriété d'approximation de l'énoncé pour *tout* $\varepsilon > 0$, on vérifie facilement que $\mathscr{D}$ est une tribu. Comme elle contient $\mathscr{C}$, elle est égale à $\sigma(\mathscr{C})$. □

Preuve du Théorème de Doob.
1) Soit $(\mathscr{B}_n)_{n\geqslant 1}$ la filtration à laquelle $(M_n)_{n\geqslant 1}$ est adaptée. Par hypothèse, il existe $M \in L^1(\Omega, \mathscr{A}, \mathbb{P})$ telle que $M_n = \mathbb{E}^{\mathscr{B}_n}(M)$ pour tout $n \geqslant 1$.
Soit $\mathscr{C}$ l'algèbre de Boole $\bigcup_{n=1}^{\infty} \mathscr{B}_n$, $\mathscr{B} = \sigma(\mathscr{C})$, et $M_\infty = \mathbb{E}^{\mathscr{B}}(M)$.
Si $\varepsilon > 0$, on peut trouver $Y = \sum_{j=1}^{J} a_j \mathbb{1}_{B_j}$, avec $B_j \in \mathscr{B}$, $1 \leqslant j \leqslant J$, telle que $\|M_\infty - Y\|_1 \leqslant \varepsilon$. Grâce au lemme, on peut trouver, pour chaque j, un $C_j \in \mathscr{C}$ tel que :

$$\|\mathbb{1}_{B_j} - \mathbb{1}_{C_j}\|_1 \leqslant \frac{\varepsilon}{1 + |a_1| + \cdots + |a_J|};$$

il s'ensuit que $Z = \sum_{j=1}^{J} a_j \mathbb{1}_{C_j}$ est $\mathscr{C}$-mesurable, et que :

$$\|M_\infty - Z\|_1 \leqslant \|M_\infty - Y\|_1 + \|Y - Z\|_1 \leqslant 2\varepsilon.$$

Par définition de $\mathscr{C}$, on peut trouver un $n_0 \geqslant 1$ tel que Z soit $\mathscr{B}_{n_0}$-mesurable. On a donc, pour $n \geqslant n_0$:

$$\begin{aligned} M_n - M_\infty &= \mathbb{E}^{\mathscr{B}_n}(M) - M_\infty = \mathbb{E}^{\mathscr{B}_n}(M_\infty) - M_\infty \\ &= \mathbb{E}^{\mathscr{B}_n}(M_\infty) - \mathbb{E}^{\mathscr{B}_n}(Z) + Z - M_\infty; \end{aligned}$$

d'où :

$$\|M_n - M_\infty\|_1 \leqslant \|\mathbb{E}^{\mathscr{B}_n}(M_\infty - Z)\|_1 + \|M_\infty - Z\|_1 \leqslant 2\|M_\infty - Z\|_1 \leqslant 4\varepsilon.$$

Cela montre que $M_n \xrightarrow[n\to\infty]{L^1} M_\infty$.

Appliquons maintenant, pour chaque $n \geqslant 1$, l'inégalité maximale de Doob à la martingale $(M_k - M_n)_{k\geqslant n}$. On obtient :

$$\mathbb{P}\Big(\sup_{n\leqslant k\leqslant N} |M_k - M_n| > a\Big) \leqslant \frac{1}{a}\, \|M_N - M_n\|_1\,.$$

Donc, en faisant tendre N vers l'infini :

$$\mathbb{P}\Big(\sup_{k\geqslant n} |M_k - M_n| > a\Big) \leqslant \frac{1}{a}\, \|M_\infty - M_n\|_1\,.$$

Cette inégalité montre que la fonction maximale $\sup_{k>n} |M_k - M_n|$ converge en probabilité vers 0. Cela signifie que $(M_n)_{n\geqslant 1}$ est presque sûrement de Cauchy, et donc converge presque sûrement.

Comme elle converge vers M_∞ pour la norme de L^1, la limite presque sûre est forcément M_∞.

2) Posons $M_0 = 0$ et $d_n = M_n - M_{n-1}$ pour $n \geqslant 1$. Les accroissements d_n sont orthogonaux dans L^2 car pour $j < n$ on a :

$$\mathbb{E}\,(d_j d_n) = \mathbb{E}\left[\mathbb{E}^{\mathscr{B}_j}(d_j d_n)\right] = \mathbb{E}\left[d_j \mathbb{E}^{\mathscr{B}_j}(d_n)\right] = \mathbb{E}\,(d_j.0) = 0\,.$$

On a donc $\mathbb{E}\,(M_n^2) = \sum_{j=1}^n \mathbb{E}\,(d_j^2)$, et donc, puisque l'on a supposé la martingale $(M_n)_{n\geqslant 1}$ bornée dans L^2 : $\sum_{j=1}^\infty \mathbb{E}\,(d_j^2) < +\infty$. La série $\sum_{j=1}^\infty d_j$ converge donc dans L^2, et puisque $M_n = d_1 + \cdots + d_n$, cela montre que la martingale $(M_n)_{n\geqslant 1}$ converge dans L^2. Notons M sa limite.

Il ne reste plus qu'à voir que $\mathbb{E}^{\mathscr{B}_n}(M) = M_n$, car cela signifiera que $(M_n)_{n\geqslant 1}$ est une martingale terminée, et donc converge presque sûrement, par le 1). On notera que l'on a forcément $M = M_\infty$.

Pour montrer cela, pour chaque $n \geqslant 1$, fixons $B \in \mathscr{B}_n$. Alors, pour $N \geqslant n$, on a :

$$\mathbb{E}\,(M_n \mathbb{1}_B) = \mathbb{E}\left[\mathbb{E}^{\mathscr{B}_n}(M_N)\mathbb{1}_B)\right] = \mathbb{E}\left[\mathbb{E}^{\mathscr{B}_n}(M_N \mathbb{1}_B)\right] = \mathbb{E}\,(M_N \mathbb{1}_B).$$

Comme $|\mathbb{E}\,(M_N \mathbb{1}_B) - \mathbb{E}\,(M \mathbb{1}_B)| \leqslant \|M_N - M\|_1 \leqslant \|M_N - M\|_2$, on obtient, en faisant tendre N vers l'infini : $\mathbb{E}\,(M_n \mathbb{1}_B) = \mathbb{E}\,(M \mathbb{1}_B)$. On a donc bien $M_n = \mathbb{E}^{\mathscr{B}_n}(M)$. □

Nous nous arrêterons ici.

Signalons simplement que l'on peut montrer les résultats suivants :

1) (Doob) Toute martingale bornée dans L^1 converge presque sûrement.

2) Pour toute martingale $(M_n)_{n\geqslant 1}$, il y a équivalence entre :
  a) $(M_n)_{n\geqslant 1}$ converge dans L^1 (pour la norme) ;
  b) $(M_n)_{n\geqslant 1}$ est terminée ;
  c) $\{M_n\,;\ n \geqslant 1\}$ est équi-intégrable.

3) Pour $1 < p < \infty$, si $M_n \in L^p(\Omega, \mathscr{A}, \mathbb{P})$ et si $\|M_n\|_p \leqslant K$ pour tout $n \geqslant 1$, alors les conditions précédentes sont vérifiées, et de plus la convergence a lieu pour la norme de L^p.