Année 2003/2004
Faculté de MathInfo
Maîtrises de Mathématiques
Premier semestre
UV MM3

COURS DE PROBABILITES

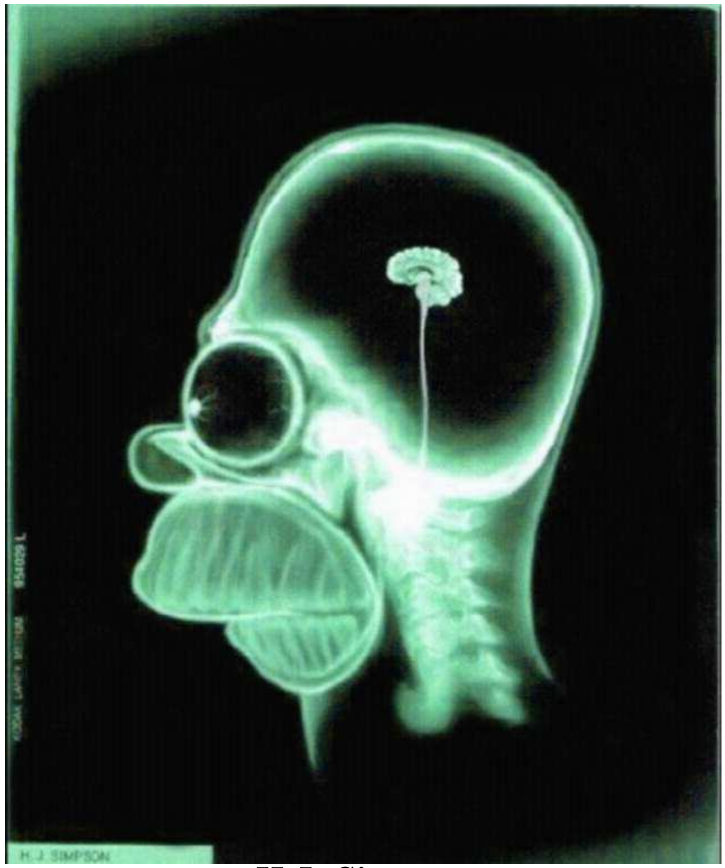
Professeur Y. Lacroix

Andreï Nikolaïevitch Kolmogorov

Extrait de l'encyclopédie Hachette Multimédia édition 2001 :

"Mathématicien russe (Tambov, 1903 - Moscou, 1987). Il est connu pour sa théorie axiomatique des probabilités."

Pour information il a fait, avant de fonder les probabilités modernes, des études d'histoire et travaillé pour les chemins de fer Russes comme conducteur.



H.J. Simpson

Extrait de l'encyclopédie Hachette Multimédia édition 2001 :

“ ”

Jacques 1$^{\text{ier}}$ Bernoulli

Extrait de l'encyclopédie Hachette Multimédia édition 2001 :

"(Bâle, 1654 - id., 1705), a contribué au développement du calcul différentiel et intégral, du calcul des variations, du calcul des probabilités et du calcul exponentiel."

Il est issu d'une famille de Mathématiciens et Physiciens Bâlois.

TABLE DES MATIÈRES

Bibliographie

[B] : P. Billingsley, "*Probability and measure*", Wiley Series in Probability and Mathematical Statistics, Third Edition (1995).

[BL] : P. Barbe & M. Ledoux, "*Probabilité*", Belin, Editions espaces 34 (1998).

[D] : R. Durrett, "*Probability : theory and examples*", Duxbury Press, Second Edition (1996).

[K] : K. Petersen, "*Easy and nearly simultaneous proofs of the Ergodic Theorem and Maximal Ergodic Theorem*", http://www.math.unc.edu/Faculty/petersen/ (1999).

[R] : W. Rudin, "*Real and complex analysis*", McGraw-Hill (1986)

Unicité et existence des probabilités

En intégration de Licence, le lecteur aura traité l'étude des espaces mesurables $(\Omega, \mathcal{B})$, et des mesures positives μ sur ces espaces. Il aura ensuite construit sur ces triplets $(\Omega, \mathcal{B}, \mu)$ la théorie de l'intégration, en partant des indicatrices, puis passant aux fonctions simples, puis aux mesurables positives, et enfin il aura définit l'espace $\mathcal{L}^1(\mu)$ des fonctions $f : \Omega \to \mathbb{R}$ mesurables intégrables, donc celles pour lesquelles la quantité

$$\int_\Omega f(\omega) d\mu(\omega) = \int f d\mu$$

est bien définie. Tous les théorèmes du cours d'intégration de Licence seront valables dans le cours qui va suivre, pour la raison simple qui suit :

les espaces de probabilités sont les espaces mesurés équipés d'une mesure positive $\mathbb{P}$ telle que

$$\mathbb{P}(\Omega) = \mathbf{1} \qquad \text{(Poids total égal à 1)}$$

La mesure finie garantit l'intégrabilité des constantes.

Nous parlerons d'**espérance** plutôt que d'intégrale, et nous noterons

$$\mathbb{E}(\mathbf{f}) = \mathbb{E}_\mu(\mathbf{f}) = \int f d\mu.$$

Avant d'aller vers les spécifités du calcul des probabilités, nous développons dans ce premier chapitre quelques résultats fondamentaux au sujet de l'existence et l'unicité des probabilités.

Pour ce qui est de l'existence nous avons opté pour des critères pratiques. Nous aurions pu développer la théorie de la régularité et le théorème de représentation de Riesz.

Nous noterons $\mathbf{1}_A$ la fonction indicatrice d'un ensemble A.

1.1. Tribu, mesure, algèbre, π-système, générateur.

Dans tout ce qui suit Ω désigne un ensemble (l'espace des **événements élémentaires**), et $\mathcal{B}$ une **tribu** ou **σ-algèbre** de parties de Ω, à savoir

$$\begin{cases} \Omega \in \mathcal{B}; \\ A \in \mathcal{B} \Rightarrow \Omega \setminus A =: A^c \in \mathcal{B}; \\ A_n \in \mathcal{B},\ n \geq 0 \Rightarrow \cup_n A_n \in \mathcal{B}. \end{cases}$$

On appelle $\mathcal{B}$ l'ensemble des **événements observables**, ou **mesurables**. Il est naturel de le munir à priori d'une structure de tribu, dans la mesure où :

– le tout sans restriction (Ω) est observable;

– si on sait observer A, on sait observer non A, i.e. A^c;

– si pour chaque n on sait observer A_n, on sait observer l'événement "l'un des A_n s'est produit", i.e. $\cup_n A_n$.

Observons que si $A_n \in \mathcal{B}$, $n \geq 0$, alors

$$\begin{aligned}&\liminf A_n := \{\omega \in \Omega : \exists n_0(\omega),\ \omega \in \cap_{n \geq n_0(\omega)} A_n\} = \cup_n \cap_{p \geq n} A_p \text{ et}\\ &\limsup A_n := \{\omega \in \Omega : \mathrm{Card}\{n : \omega \in A_n\} = \infty\} = \cap_n \cup_{p \geq n} A_p\end{aligned}$$

sont dans $\mathcal{B}$. On a (exercice)

$$\begin{cases} \cap_n A_n \subset \liminf A_n \subset \limsup A_n \subset \cup_n A_n; \\ \liminf A_n = \limsup A_n \Leftrightarrow \forall \omega \in \Omega,\ (\mathbf{1}_{A_n}(\omega))_n \text{ converge.} \end{cases}$$

Le couple $(\Omega, \mathcal{B})$ est appelé un **espace mesurable**. Une **mesure positive** μ sur ce dernier est une fonction d'ensembles $\mu : \mathcal{B} \to \mathbb{R}^+ \cup \{+\infty\}$ telle que

$$\begin{cases} \mu(\emptyset) = 0; \\ \mu(\cup_n A_n) = \sum_n \mu(A_n) \text{ si les } A_n \text{ sont deux à deux disjoints et dans } \mathcal{B}\ (\sigma\textbf{-additivité}). \end{cases}$$

Un sous-ensemble $\mathcal{E} \subset \mathcal{P}(\Omega)$ est dit **générateur** si l'intersection de toutes les tribus contenant $\mathcal{E}$ - qui en est une, coïncide avec $\mathcal{B}$. On note alors

$$\mathcal{B} = \sigma(\mathcal{E}).$$

Une **algèbre** est un sous ensemble $\mathcal{A} \subset \mathcal{P}(\Omega)$ stable par complémentation et union finie, et contenant $\emptyset$.

Un π**-système** est un sous-ensemble de $\mathcal{P}(\Omega)$ stable par intersections finies.

1.2. Théorème $\pi - \lambda$ et unicité. Théorème des classes monotones.

Un λ**-système** est une classe de parties de Ω contenant Ω, stable par complémentation, et stable par unions dénombrables disjointes. Vous avez déjà vu en Licence les deux résultats suivants :

Théorème $\pi - \lambda$ (*Dynkin*). *Si le π-système Π est contenu dans un λ-système Λ, alors $\sigma(\Pi) \subset \Lambda$.*

Celui-ci permet de démontrer le suivant :

Théorème d'unicité. *Soit Π un π-système tel que $\sigma(\Pi) = \mathcal{B}$. Soient $\mathbb{P}_1$ et $\mathbb{P}_2$ deux mesures de probabilité sur $\mathcal{B}$, et qui coïncident sur Π. Alors elles sont identiques.*

Preuve. La classe des éléments de $\mathcal{B}$ sur lesquels $\mathbb{P}_1$ et $\mathbb{P}_2$ coïncident, est un λ-système, contenant Π. Elle contient donc $\sigma(\Pi) = \mathcal{B}$, et donc $\mathbb{P}_1 = \mathbb{P}_2$ sur $\mathcal{B}$. ■

Nous aurons besoin, au Chapitre 11, d'un théorème des classes monotones : **une classe monotone** est une classe de parties d'un ensemble stable par unions dénombrables croissantes, et intersections dénombrables décroissantes.

Théorème des classes monotones (Halmos). *Si $\mathcal{A}$ est une algèbre de parties d'un ensemble Ω, et si $\mathcal{M}$ est une classe monotone sur Ω, contenant $\mathcal{A}$, alors*

$$\sigma(\mathcal{A}) \subset \mathcal{M}.$$

Preuve. Notons $\mathcal{M}(\mathcal{A})$ la classe monotone engendrée par $\mathcal{A}$: une intersection de classes monotones en est une, il y en a qui contiennent $\mathcal{A}$, et l'intersection de toutes celles qui contiennent $\mathcal{A}$ vaut $\mathcal{M}(\mathcal{A})$, par définition.

On a donc $\mathcal{M}(\mathcal{A}) \subset \mathcal{M}$, et il nous suffit, pour montrer que $\sigma(\mathcal{A}) \subset \mathcal{M}$, de montrer que $\mathcal{M}(\mathcal{A})$ est une σ-algèbre.

Notons $\mathcal{G} = \{A : A^c \in \mathcal{M}(\mathcal{A})\}$. Puisque $\mathcal{M}(\mathcal{A})$ est une classe monotone, $\mathcal{G}$ en est une aussi. Puisque $\mathcal{A}$ est une algèbre, $\mathcal{A} \subset \mathcal{G}$. Et donc $\mathcal{M}(\mathcal{A}) \subset \mathcal{G}$, i.e. $\mathcal{M}(\mathcal{A})$ est stable par complémentation.

Notons $\mathcal{G}_1 = \{A : B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{M}(\mathcal{A})\}$. Alors $\mathcal{G}_1$ est une classe monotone contenant $\mathcal{A}$, et donc $\mathcal{M}(\mathcal{A}) \subset \mathcal{G}_1$.

Notons $\mathcal{G}_2 = \{A : B \in \mathcal{M}(\mathcal{A}) \Rightarrow A \cup B \in \mathcal{M}(\mathcal{A})\}$. C'est une classe monotone, et qui contient $\mathcal{A}$, puisque $\mathcal{M}(\mathcal{A}) \subset \mathcal{G}_1$. Donc $\mathcal{M}(\mathcal{A}) \subset \mathcal{G}_2$, i.e. $\mathcal{M}(\mathcal{A})$ est stable par unions finies.

Et donc $\mathcal{M}(\mathcal{A})$ est une algèbre, et une classe monotone. C'est donc une σ-algèbre. ■

1.3. Existence.

Le théorème d'existence nous permettra de construire effectivement des lois de probabilités. Une **probabilité sur une algèbre** $\mathcal{A}$ se définit à l'instar de ce qu'il en est d'une tribu, à ceci près qu'on va rajouter, pour la σ-additivité, que $\cup_n A_n \in \mathcal{A}$ (ce qui n'est pas automatique).

Théorème d'existence. *Soit $\mathbb{P}$ une probabilité sur une algèbre $\mathcal{A}$. Alors il existe une unique probabilité $\tilde{\mathbb{P}}$ sur $\sigma(\mathcal{A})$, qui coïncide avec $\mathbb{P}$ sur $\mathcal{A}$.*

Preuve. La preuve repose sur l'utilisation de la **mesure extérieure** $\mathbb{P}^*$, et est pour l'essentiel dûe à Carathéodory. On pose

$$\mathbb{P}^*(E) = \inf_{\substack{A_n \in \mathcal{A}, \\ E \subset \cup_n A_n}} \sum_n \mathbb{P}(A_n),\ E \subset \Omega.$$

Considérons

$$\mathcal{M} = \{A \subset \Omega : \forall E \subset \Omega,\ \mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) = \mathbb{P}^*(E)\}.$$

Nous allons montrer que sur $\mathcal{M}$, $\mathbb{P}^*$ est une probabilité, que $\mathcal{M}$ est une tribu, et que $\mathcal{A} \subset \mathcal{M}$, et enfin que $\mathbb{P}^*$ et $\mathbb{P}$ coïncident sur $\mathcal{A}$. L'unicité de l'extension résultera simplement du théorème d'unicité pré-cité.

Les propriétés suivantes de $\mathbb{P}^*$ sont élémentaires à démontrer (propriétés de l'inf) :

$$\begin{array}{rcl} (i) & : & \mathbb{P}^*(\emptyset) = 0; \\ (ii) & : & \mathbb{P}^* \geq 0 \text{ (positivité)}; \\ (iii) & : & A \subset B \Rightarrow \mathbb{P}^*(A) \leq \mathbb{P}^*(B) \text{ (monotonie)}; \\ (iv) & : & \mathbb{P}^*(\cup_n A_n) \leq \sum_n \mathbb{P}^*(A_n) \text{ (sous-}\sigma\text{-additivité)}. \end{array}$$

Remarquons que $A \in \mathcal{M} \Leftrightarrow \mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) \leq \mathbb{P}^*(E)$ pour tout $E \subset \Omega$, puisque l'inégalité contraire résulte de la sous-additivité (iv).

Lemme 1. *$\mathcal{M}$ est une algèbre.*

Preuve du Lemme 1. Par (i), et par définition de $\mathcal{M}$, $\mathcal{M}$ contient $\emptyset$ et est stable par complémentation. Soient $A, B \in \mathcal{M}$ et soit $E \subset \Omega$: alors

$$\begin{array}{rcl} \mathbb{P}^*(E) & = & \mathbb{P}^*(B \cap E) + \mathbb{P}^*(B^c \cap E) \ (B \in \mathcal{M}) \\ & = & \mathbb{P}^*(A \cap B \cap E) + \mathbb{P}^*(A^c \cap B \cap E) + \mathbb{P}^*(A \cap B^c \cap E) \\ & & \qquad + \mathbb{P}^*(A^c \cap B^c \cap E) \ (A \in \mathcal{M}, \text{ deux fois}) \\ & \geq & \mathbb{P}^*(A \cap B \cap E) + \mathbb{P}^*((A \cap B)^c \cap E) \ ((iv)). \end{array}$$

Donc $A \cap B \in \mathcal{M}$. ■

Lemme 2. *Si les A_n constituent une suite finie ou infinie d'éléments deux à deux disjoints de $\mathcal{M}$, alors quel que soit $E \subset \Omega$,*

$$\mathbb{P}^*(E \cap (\cup_n A_n)) = \sum_n \mathbb{P}^*(E \cap A_n).$$

Preuve du Lemme 2. Supposons que la suite soit finie, et à deux termes A_1, A_2, pour commencer. Alors $\mathbb{P}^*(E \cap (A_1 \cup A_2)) = \mathbb{P}^*(E \cap (A_1 \cup A_2) \cap A_1) + \mathbb{P}^*(E \cap (A_1 \cup A_2) \cap A_1^c) = \mathbb{P}^*(E \cap A_1) + \mathbb{P}^*(E \cap A_2)$ car $A_1 \cap A_2 = \emptyset$ (on a simplement écrit la condition $A_1 \in \mathcal{M}$ pour l'ensemble $\tilde{E} = E \cap (A_1 \cup A_2)$).

Si la suite est finie, raisonner par récurrence, en écrivant la condition $A_n \in \mathcal{M}$, pour l'ensemble $\tilde{E} = E \cap (\cup_{k<n} A_k)$.

Dans le cas infini, utiliser la monotonie (iii) : $\mathbb{P}^*(E \cap (\cup_k A_k)) \geq \mathbb{P}^*(E \cap (\cup_{k \leq n} A_k) = \sum_{k \leq n} \mathbb{P}^*(E \cap A_k)$, pour conclure que $\sum_n \mathbb{P}^*(E \cap A_k) \leq \mathbb{P}^*(E \cap (\cup_n A_n))$. L'inégalité contraire résulte de la sous-σ-additivité (iv). ■

Lemme 3. *$\mathcal{M}$ est une tribu, et $\mathbb{P}^*$ restreinte à $\mathcal{M}$ est σ-additive.*

Preuve du Lemme 3. Puisque $\mathcal{M}$ est une algèbre, il suffit, pour montrer qu'elle est une tribu, de montrer qu'elle est stable par unions dénombrables disjointes.

Soit (A_n) une suite disjointe dans $\mathcal{M}$ et posons $U_n = \cup_{k\leq n} A_k$, $n \geq 1$. Posons aussi $U = \cup_n A_n$. On a $\mathbb{P}^*(E \cap U_n) + \mathbb{P}^*(E \cap U_n^c) = \mathbb{P}^*(E)$, pour tout n et $E \subset \Omega$.

Par monotonie, $\mathbb{P}^*(E \cap U_n^c) \geq \mathbb{P}^*(E \cap U^c)$, et par le Lemme 2, $\mathbb{P}^*(E \cap U_n) = \sum_{k\leq n} \mathbb{P}^*(E \cap A_k)$. Donc pour chaque n, $\mathbb{P}^*(E) \geq \sum_{k\leq n} \mathbb{P}^*(E \cap A_k) + \mathbb{P}^*(E \cap U^c)$, et par passage à la limite, $\mathbb{P}^*(E) \geq \sum_n \mathbb{P}^*(E \cap A_n) + \mathbb{P}^*(E \cap U^c)$, ce qui par le Lemme 2, encore, s'écrit $\mathbb{P}^*(E) \geq \mathbb{P}^*(E \cap U) + \mathbb{P}^*(E \cap U^c)$. Donc $U \in \mathcal{M}$. ■

Lemme 4. *$\mathcal{A} \subset \mathcal{M}$.*

Preuve du Lemme 4. Soient $A \in \mathcal{A}$, $\varepsilon > 0$, et $E \subset \Omega$. Soit (A_n) dans $\mathcal{A}$ telle que $\sum_n \mathbb{P}(A_n) \leq \mathbb{P}^*(E) + \varepsilon$, et $E \subset \cup_n A_n$. Soient $B_n = A \cap A_n$ et $C_n = A^c \cap A_n$: $B_n, C_n \in \mathcal{A}$.

On a $E \cap A \subset \cup_n B_n$, et $E \cap A^c \subset \cup_n C_n$. Par définition de $\mathbb{P}^*$ et additivité finie de $\mathbb{P}$, on a

$$\mathbb{P}^*(E \cap A) + \mathbb{P}^*(E \cap A^c) \leq \sum_n \mathbb{P}(B_n) + \sum_n \mathbb{P}(C_n) = \sum_n \mathbb{P}(A_n) \leq \mathbb{P}^*(E) + \varepsilon,$$

ce qui démontre le lemme si l'on fait tendre ε vers 0. ■

Fin de la preuve du théorème : il ne reste qu'à vérifier que $\mathbb{P}^* = \mathbb{P}$ sur $\mathcal{A}$. Par définition de $\mathbb{P}^*$, $\mathbb{P}^*(A) \leq \mathbb{P}(A)$ si $A \in \mathcal{A}$.

Si (A_n) recouvre A dans $\mathcal{A}$, par sous-σ-additivité de $\mathbb{P}$ sur $\mathcal{A}$, il vient $\mathbb{P}(A) \leq \sum_n \mathbb{P}(A_n)$, et en passant à l'inf sur ces recouvrements, on obtient $\mathbb{P}(A) \leq \mathbb{P}^*(A)$. ■

Remarque. La classe $\mathcal{M}$ à laquelle s'étend $\mathbb{P}$ est souvent bien plus large que $\sigma(\mathcal{A})$. Elle contient notamment les ensembles $\mathbb{P}$-négligeables, i.e. les sous-ensembles des ensembles de $\sigma(\mathcal{A})$ qui sont de mesure nulle.

Exemple. Construction de la mesure de Lebesgue m sur $[0, 1]$. On considère l'algèbre $\mathcal{A}$ des unions finies disjointes de sous-intervalles de $[0, 1]$ (vérifier que c'est une algèbre).

Pour $\cup_i I_i \in \mathcal{A}$, on pose $m(\cup_i I_i) = \sum_i \text{longueur}(I_i)$. Il faut vérifier que c'est une bonne définition, puis que si une telle union s'écrit comme union dénombrable disjointe de telles unions, alors la série des sommes de longueur redonne bien la somme de longueurs initiales. On peut faire cela autrement, ainsi que nous l'indique le Lemme qui suit, en vérifiant que si $(A_n) \in \mathcal{A}$ est telle que $A_{n+1} \subset A_n$ et que $\cap_n A_n = \emptyset$, alors $m(A_n) \to \emptyset$.

Lemme (Continuité en $\emptyset$). *Soit $\mathbb{P} : \mathcal{A} \to [0, 1]$ telle que $\mathcal{A}$ soit une algèbre, $\mathbb{P}(\Omega) = 1$, et si $A \cap B = \emptyset$ avec $A, B \in \mathcal{A}$, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. Alors $\mathbb{P}$ est une* *probabilité sur $\mathcal{A}$ si et seulement si* $\mathbb{P}$ **est continue en** $\emptyset$, *à savoir si* (A_n) *dans* $\mathcal{A}$ *et* $A_{n+1} \subset A_n$ *et* $\cap_n A_n = \emptyset$, *alors* $\mathbb{P}(A_n) \to 0$.

Preuve. Supposons que $\mathbb{P}$ soit une probabilité. Alors si la suite (A_n) décroît dans $\mathcal{A}$ vers $\emptyset$, posons $B_n = A_n \setminus A_{n+1}$. On a $B_n \in \mathcal{A}$ et les B_n sont deux à deux disjoints, d'union A_0. Donc $\sum_n \mathbb{P}(B_n) = \mathbb{P}(A_0)$. Et $\mathbb{P}(A_N) = \sum_{n \geq N} \mathbb{P}(B_n) \to 0$.

Réciproquement, soit (B_n) une suite dans $\mathcal{A}$, de termes deux à deux disjoints, et telle que $\cup_n B_n = A_0 \in \mathcal{A}$. Alors si nous posons $A_n = \cup_{k \geq n} B_k$, la suite (A_n) est dans $\mathcal{A}$, et décroît, et $\cap A_n = \emptyset$. Donc $\mathbb{P}(A_n) \to 0$. Mais $\mathbb{P}(A_0) = \mathbb{P}((\cup_{k<n} B_k) \cup A_n) = \sum_{k<n} \mathbb{P}(B_k) + \mathbb{P}(A_n)$, et donc $\mathbb{P}(\cup_n B_n) = \sum_n \mathbb{P}(B_n)$. ■

Continuation de l'exemple. Donnons nous donc (A_n) une suite décroissante d'unions finies d'intervalles de $[0,1]$ telle que $\cap_n A_n = \emptyset$. Nous voulons vérifier que $m(A_n) \to 0$. Soit $\varepsilon > 0$. Pour chaque n, nous pouvons trouver une union finie disjointe d'intervalles fermés $K_n^0 \subset A_n$ telle que

$$m(A_n \setminus K_n^0) \leq \varepsilon \;/\; 2^{n+1}.$$

Posons $K_n = K_1^0 \cap \ldots \cap K_n^0$. Alors la suite (K_n) est une suite décroissante de compacts de $\mathbb{R}$, et d'intersection vide. Il s'ensuit qu'il existe n_0 tel que $n \geq n_0 \Rightarrow K_n = \emptyset$. Alors pour $n \geq n_0$,

$$m(A_n) = m(A_n \setminus K_n) = m(A_n \cap (\cap_{s \leq n} K_s^0)^c) \leq \sum_{s \leq n} m(A_s \cap (K_s^0)^c) \leq \sum_s \varepsilon \;/\; 2^{s+1} = \varepsilon,$$

et donc on a montré que $m(A_n) \to 0$.

Par application du Lemme de continuité en $\emptyset$, il s'ensuit que m est une probabilité sur $\mathcal{A}$. Le théorème d'existence nous donne alors l'existence d'une unique mesure de probabilité sur $([0,1], \sigma(\mathcal{A}))$ telle que $m(I) = \textit{longueur}(I)$ si $I \subset [0,1]$ est un intervalle. C'est la mesure de Lebesgue sur $[0,1]$.

Remarquons que $\sigma(\mathcal{A}) = \{A \cap [0,1] : A \in \mathcal{B}(\mathbb{R})\}$, où $\mathcal{B}(\mathbb{R})$ désigne la tribu borélienne de $\mathbb{R}$.

1.4. Loi image, et conditions de compatibilité de Kolmogorov.

D'abord, supposons donnés deux espaces probabilisables $(\Omega, \mathcal{B})$ et $(\Omega', \mathcal{B}')$, et une application $f : \Omega \to \Omega'$ qui soit $(\mathcal{B}, \mathcal{B}')$-mesurable. Si $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{B})$, alors posons, pour $B' \in \mathcal{B}'$,

$$\mathbb{P}_f(B') = \mathbb{P}(f^{-1}B').$$

On vérifie sans peine que $\mathbb{P}_f$ **est une probabilité sur** $(\Omega', \mathcal{B}')$, dite **loi image de** $\mathbb{P}$ **par** f.

Quelques notations préliminaires s'imposent : notons $\mathcal{T}_n := \mathcal{B}(\mathbb{R}^n)$ la **tribu borélienne de** $\mathbb{R}^n$, à savoir celle engendrée par les ouverts. Notons $\phi_n : \mathbb{R}^{n+1} \to \mathbb{R}^n$ la projection canonique définie par $\phi_n(x_0, \ldots, x_n) = (x_0, \ldots, x_{n-1})$. Notons $\pi_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^n$ la projection définie par $\pi_n((x_k)) = (x_0, \ldots, x_{n-1})$.

Notons $\mathcal{T} = \sigma(\cup_n \pi_n^{-1}(\mathcal{T}_n))$. C'est une tribu sur $\mathbb{R}^{\mathbb{N}}$.

Théorème d'extension de Kolmogorov. *Supposons donnée* $(\mathbb{P}_n)$ *une suite telle que*

$$\begin{cases} (i) & : \quad \mathbb{P}_n \text{ est une probabilité sur } (\mathbb{R}^n, \mathcal{T}_n), \text{ pour chaque } n \geq 0; \\ (ii) & : \quad \text{l'image de } \mathbb{P}_{n+1} \text{ par } \phi_n \text{ est } \mathbb{P}_n \ ((\mathbb{P}_{n+1})_{\phi_n} = \mathbb{P}_n), \ n \geq 0. \end{cases}$$

Alors il existe une unique probabilité $\mathbb{P}$ *sur* $(\mathbb{R}^{\mathbb{N}}, \mathcal{T})$*, telle que pour chaque* n,

$$\mathbb{P}_{\pi_n} = \mathbb{P}_n.$$

Preuve. Montrer que $\mathcal{A} := \cup_n \pi_n^{-1}(\mathcal{A}_n)$ est une algèbre, où $\mathcal{A}_n$ est l'algèbre sur $\mathbb{R}^n$ des unions finies disjointes de pavés de $\mathbb{R}^n$, un pavé étant le produit cartésien de n intervalles réels.

Poser, pour $A \in \mathcal{A}$, $\mathbb{P}(A) = \mathbb{P}_n(\pi_n(A))$, si $A \in \pi_n^{-1}(\mathcal{A}_n)$, et vérifier que c'est une bonne définition, en utilisant (ii), et le fait que $\phi_n \circ \phi_{n+1} \ldots \circ \phi_{n+m} \circ \pi_{n+m+1} = \pi_n$, si $n, m \geq 1$.

En utilisant (i) et (ii), il est facile de vérifier que $\mathbb{P} : \mathcal{A} \to [0,1]$ est additive, et que $\mathbb{P}(\mathbb{R}^{\mathbb{N}}) = 1$.

Mais il faut aussi montrer la continuité de $\mathbb{P}$ en $\emptyset$: supposons donc que (A_k) soit décroissante dans $\mathcal{A}$, et que $\cap_k A_k = \emptyset$. Par définition de $\mathcal{A}$, pour chaque entier k, il en existe un autre, n_k, tel que $A_k = \pi_{n_k}^{-1}(B_{n_k})$, où B_{n_k} est une union finie disjointe de pavés de $\mathbb{R}^{n_k}$.

Lemme de régularité. *Soit* $\tilde{\mathbb{P}}$ *une mesure de probabilité borélienne sur* $\mathbb{R}^n$*. Alors quel que soit le pavé* P *de* $\mathbb{R}^n$*,* $\tilde{\mathbb{P}}(P)$ *est égal au sup de l'ensemble des* $\tilde{\mathbb{P}}(K)$*, où* K *parcours l'ensemble des pavés compacts contenus dans* P*.*

Preuve du Lemme de régularité. Pour l'essentiel il s'agit de vérifier les différents cas de figures possibles concernant P. D'autre part, le cas de $n = 2$ révèle déjà l'ensemble des circonstances éventuellement rencontrées.

Nous nous limiterons donc à ne démontrer le Lemme que dans le cas très (mais pas trop) particulier $P = [0, 1[\times[5, +\infty[$, à titre d'exemple. Notons alors $K_t = [0, 1-\frac{1}{t}[\times[5, t]$, $t \geq 1$. Il est clair que $K_t \uparrow P$, et donc $\tilde{\mathbb{P}}$ étant une mesure de probabilités, $\tilde{\mathbb{P}}(K_t) \uparrow \tilde{\mathbb{P}}(P)$. Enfin l'on remarque que K_t est compact. ■

Fin de la preuve du théorème d'extension de Kolmogorov : soit $\varepsilon > 0$ et pour chaque $k \geq 1$ et B_{n_k}, choisissons $K_{n_k}^0 \subset B_{n_k}$ compact et tel que $\mathbb{P}_{n_k}(B_{n_k} \setminus K_{n_k}^0) < \varepsilon \ / \ 2^{n_k+1}$. Nous pouvons supposer en outre, sans contrainte supplémentaire, que la suite $(n_k)_k$ est strictement croissante.

Posons $K_k = \cap_{s \leq k} \pi_{n_s}^{-1}(K_{n_s}^0)$, $k \geq 1$. Alors la suite $(K_k)_{k \geq 1}$ est une suite décroissante dans $\mathbb{R}^{\mathbb{N}}$.

Supposons, contrairement à la conclusion recherchée, que $\inf_{k\geq 1}\mathbb{P}(A_k) > \varepsilon$. Alors

$$\begin{array}{rcl}\mathbb{P}(A_k \setminus K_k) & = & \mathbb{P}(A_k \cap (\cap_{s\leq k}\pi_{n_s}^{-1}(K_{n_s}^0))^c) \\ & \leq & \sum_{s\leq k}\mathbb{P}(A_k \cap (\pi_{n_s}^{-1}(K_{n_s}^0)^c)) \\ & \leq & \sum_{s\leq k}\mathbb{P}(A_s \cap \pi_{n_s}^{-1}((K_{n_s}^0)^c)) \\ & = & \sum_{s\leq k}\mathbb{P}(\pi_{n_s}^{-1}(B_{n_s} \cap (K_{n_s}^0)^c)) \\ & = & \sum_{s\leq k}\mathbb{P}_{n_s}(B_{n_s} \cap (K_{n_s}^0)^c) \\ & \leq & \sum_{s\leq k}\varepsilon \;/\; 2^{n_s+1} \leq \varepsilon \;/\; 2,\end{array}$$

et donc $\mathbb{P}(K_k) > \varepsilon/2$. Ainsi $(K_k)_{k\geq 1}$ est une suite décroissante non-vide.

Pour chaque $k \geq 1$, choisissons $x^{(k)} \in K_k$. Alors $\pi_{n_k}(x^{(s)}) \in K_{n_k}^0$, si $s \geq k$. Par compacité de $K_{n_k}^0$, et par extraction diagonale, il existe une suite $(k_j)_{j\geq 1}$ croissante telle que pour chaque n, il existe $x_n = \lim_j x_n^{(k_j)} \in \mathbb{R}$.

Mais alors, pour chaque k, $(x_0, \ldots, x_{n_k-1}) = \lim_j (x_0^{(k_j)}, \ldots, x_{n_k-1}^{(k_j)}) \in K_{n_k}^0$. Et donc si $x = (x_n)_{n\geq 0}$, il s'ensuit que pour chaque k, $x \in K_k \subset A_k$, et donc que $x \in \cap_k A_k$, ce qui contredit $\cap_k A_k = \emptyset$. Conclure en utilisant le théorème d'extension. ■

Exemple : notons $\mathbb{P}$ la probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par

$$\mathbb{P}(A) = (1-p)\mathbf{1}_A(0) + p\mathbf{1}_A(1),$$

où $p \in]0,1[$ est un paramètre fixé. Vérifier que c'est une probabilité.

Poser sur $(\mathbb{R}^n, \mathcal{T}_n)$, $\mathbb{P}_n = \mathbb{P}\otimes \ldots \otimes\mathbb{P}$ (n-fois), la mesure produit, qui est une probabilité.

Vérifier que les hypothèses du théorème sont satisfaites, et en déduire l'existence d'une unique probabilité $\mathbb{P}_\infty$ sur $(\mathbb{R}^{\mathbb{N}}, \mathcal{T})$. Elle est parfois notée $\mathbb{P}^{\otimes\mathbb{N}}$.

Remarquer que

$$\mathbb{P}_\infty(\{(x_n) \in \mathbb{R}^{\mathbb{N}} : \exists n \geq 0, x_n \notin \{0,1\}\}) = 0,$$

et donc on peut s'astreindre à ne considérer $\mathbb{P}_\infty$ que comme une probabilité sur $(\{0,1\}^{\mathbb{N}}, \mathcal{T} \cap \{0,1\}^{\mathbb{N}})$, où $\mathcal{T} \cap \{0,1\}^{\mathbb{N}}$ est un raccourci pour $\{A \cap \{0,1\}^{\mathbb{N}} : A \in \mathcal{T}\}$.

L'espace ainsi défini $(\{0,1\}^{\mathbb{N}}, \mathcal{T} \cap \{0,1\}^{\mathbb{N}}, \mathbb{P}_\infty)$ sert de modèle à l'expérience aléatoire qui consiste en une suite de jets indépendants d'une pièce d'équilibrage p : si $x \in \cap\{0,1\}^{\mathbb{N}}$ est le résultat de cette expérience, nous dirons que le i$^{\text{ème}}$ lancé fut pile si $x_i = 1$, face sinon.

Un autre exemple. Soit $(\tilde{\mathbb{P}}_n)_{n\geq 0}$ une suite de probabilités boréliennes (i.e. sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$). Notons $\mathbb{P}_n = \tilde{\mathbb{P}}_0 \otimes \ldots \otimes \tilde{\mathbb{P}}_n$ la probabilité produit associée sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Alors la suite $(\mathbb{P}_n)$ satisfait aux conditions d'application du théorème d'extension de Kolmogorov.

1.5. Autres constructions de probabilités.

1.5.1. Le cas des espaces finis ou dénombrables.

La rubrique ici correspond à la configuration suivante : Ω **est fini ou dénombrable, et** $\mathcal{B} = \mathcal{P}(\Omega)$.

Si $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{B})$, alors puisque $(\{\omega\})_{\omega\in\Omega}$ est finie ou dénombrable, $\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 1$, et si $A \subset \Omega$, $\mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\{\omega\})$.

Réciproquement, si $(p_\omega)_{\omega\in\Omega}$ est une suite indexée par Ω telle que $p_\omega \geq 0$ et $\sum_\omega p_\omega = 1$, alors il est aisé de vérifier que $\mathbb{P}(A) := \sum_{\omega\in A} p_\omega$ définit bien une probabilité sur $(\Omega, \mathcal{B})$.

Exemple : la **loi de Poisson de paramètre** $\lambda > 0$**,** $\mathcal{P}(\lambda)$, est la probabilité $\mathbb{P}$ définie sur $\mathbb{N}$ telle que

$$p_n = e^{-\lambda}\frac{\lambda^n}{n!}.$$

1.5.2. Les absolument continues sur $\mathbb{R}^n$.

Supposons $(\Omega, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Soit $f : \mathbb{R}^n \to \mathbb{R}$ $(\mathcal{B}(\mathbb{R}^n), \mathcal{B}(\mathbb{R}))$-mesurable, positive, et telle que

$$\int f dm_n = 1,$$

où m_n désigne la mesure de Lebesgue sur $\mathbb{R}^n$. Posons

$$\mathbb{P}(\mathbf{A}) = \int \mathbf{f1_A dm_n},\ \mathbf{A} \in \mathcal{B}(\mathbb{R}^\mathbf{n}).$$

Alors $\mathbb{P}$ est bien définie puisque $0 \leq f\mathbf{1}_A \leq f$ et $f \in \mathcal{L}^1(m_n)$. De plus $\mathbb{P}(\mathbb{R}^n) = \int f dm_n = 1$. Enfin $A\cap B = \emptyset \Rightarrow \mathbf{1}_A + \mathbf{1}_B = \mathbf{1}_{A\cup B}$, et donc $\mathbb{P}(A\cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. La σ-additivité de $\mathbb{P}$ se vérifie grâce au théorème de Beppo-Lévi (convergence monotone), en observant que si les A_n sont deux à deux disjoints,

$$\sum_n \mathbf{1}_{A_n} = \mathbf{1}_{\cup_n A_n}.$$

Les probabilités boréliennes ainsi définies sur $\mathbb{R}^n$ sont dites **absolument continues**. On appelle la fonction f, unique modulo les ensembles négligeables, sa **densité**.

Exemple : la loi Normale centrée réduite $\mathcal{N}(\mathbf{0}, \mathbf{1})$ est la probabilité absolument continue sur $\mathbb{R}$ de densité

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

On vérifie que f est mesurable car continue, positive, intégrable par théorème de comparaison en $\pm\infty$.

Il reste à s'assurer de ce que $\int f dm = 1$. L'astuce pour ce faire est la suivante (on passe en coordonnées polaires dans le plan) :

$$\begin{aligned}(\textstyle\int f dm)^2 &= \textstyle\int_{\mathbb{R}^2} f(x)f(y)dxdy \\ &= \frac{1}{2\pi}\textstyle\int_{\theta\in[0,2\pi[}\int_{0<\rho<\infty} \rho e^{-\frac{\rho^2}{2}} d\rho d\theta \\ &= 1.\end{aligned}$$

Chapitre 2
Décomposition des mesures et dérivées de Radon-Nykodim

2.1. Mesures complexes et variation totale.

Pour $E \in \mathcal{B}$, nous notons $Part(E)$ l'ensemble des partitions finies ou dénombrables de E en éléments de $\mathcal{B}$ deux à deux disjoints. Une **mesure complexe** sur $(\Omega, \mathcal{B})$ est une fonction d'ensembles $\mu : \mathcal{B} \to \mathbb{C}$ telle que quel que soit $E \in \mathcal{B}$, et $(E_i) \in Part(E)$,

$$\mu(E) = \sum_i \mu(E_i).$$

Remarquons que par réindexation, si $(E_i) \in Part(E)$, la définition impose que $\sum_i |\mu(E_i)| < \infty$, puisque la convergence commutative équivaut à la convergence absolue dans $\mathbb{C}$. Définissons pour $E \in \mathcal{B}$,

$$|\mu|(\mathbf{E}) = \sup \sum_{\mathbf{i}} |\mu(\mathbf{E_i})|,$$

le sup étant pris sur l'ensemble des éléments $(E_i) \in Part(E)$. On appelle $|\mu|$ ainsi définie **la variation totale de** μ.

Théorème 1. *Si μ est une mesure complexe sur $(\Omega, \mathcal{B})$, alors sa variation totale $|\mu|$ est une mesure positive sur $(\Omega, \mathcal{B})$.*

Preuve. Il s'agit de montrer la σ-additivité de $|\mu|$ sur $\mathcal{B}$. Soient $E \in \mathcal{B}$, $(E_i) \in Part(E)$, et pour chaque i, $t_i < |\mu|(E_i)$. Soit $(A_{i,j})_j \in Part(E_i)$ telle que $\sum_j |\mu(A_{i,j})| > t_i$, pour chaque i. Alors $(A_{i,j}) \in Part(E)$, et donc $\sum_i t_i < \sum_{i,j} |\mu(A_{i,j})| \leq |\mu|(E)$, par propriété du sup. Les t_i étaient arbitraires, et donc $\sum_i |\mu|(E_i) \leq |\mu|(E)$.

Réciproquement, si $(A_j) \in Part(E)$, posons $A_{i,j} = E_i \cap A_j$. Alors

$$\sum_j |\mu(A_j)| = \sum_j \left| \sum_i \mu(A_j \cap E_i) \right| \leq \sum_{i,j} |\mu(A_j \cap E_i)| \leq \sum_i |\mu|(E_i),$$

et par passage au sup on en déduit l'inégalité contraire. ■

Théorème 2. *$|\mu|(\Omega) < \infty$ si μ est une mesure complexe. Donc la variation totale d'une mesure complexe est une mesure positive finie.*

Preuve. Il s'agit de montrer que $|\mu|(\Omega) < \infty$, d'après le Théorème qui précède. Pour cela utilisons un petit lemme technique :

Lemme 1. *Si $z_1, \ldots, z_n \in \mathbb{C}$, il existe $S \subset \{1, \ldots, n\}$ tel que*

$$\left|\sum_{j\in S} z_j\right| \geq \frac{1}{6}\sum_i |z_i|.$$

Preuve du Lemme 1. Notons $w = \sum_i |z_i|$. Découpons le plan complexe $\mathbb{C} = \mathbb{R}^2$ en quatre quadrans délimités par les deux droites d'équations $y = \pm x$. L'un de ces quadrans, disons Q, est tel que

$$\sum_{z_j\in Q} |z_j| \geq \frac{1}{4}w.$$

Pour simplifier, supposons que $Q = \{|y| \leq x\}$. Alors si $z \in Q$, $Re(z) \geq |z| \,/\, \sqrt{2}$, et si $S = \{j : z_j \in Q\}$, il vient

$$|\sum_{j\in S} z_j| \geq \sum_{j\in S} Re(z_j) \geq \frac{1}{\sqrt{2}}\sum_{j\in S} |z_j| \geq \frac{1}{4\sqrt{2}}w.$$

■

Lemme 2. *Si $E \in \mathcal{B}$ est tel que $|\mu|(E) = \infty$, alors il existe $(A, B) \in Part(E)$ telle que $|\mu(A)| > 1$ et $|\mu|(B) = \infty$.*

Preuve du Lemme 2. Soit $\varepsilon > 0$. On sait que pour $t < \infty$, il existe $(E_i) \in Part(E)$ telle que $t < \sum_i |\mu(E_i)|$. Choisissons $t = 6(1 + |\mu(E)| + \varepsilon)$.

Alors il existe n tel que $\sum_{i\leq n} |\mu(E_i)| > t$. Appliquons le Lemme 1 à $z_i = \mu(E_i)$, $i \leq n$: il vient un S et si $A = \cup_{i\in S} E_i$,

$$|\mu(A)| > t/6 \geq 1.$$

Ensuite posons $B = E - A$. On a

$$|\mu(B)| \geq |\mu(E) - \mu(A)| \geq |\mu(A)| - |\mu(E)| > t/6 - |\mu(E)| = 1.$$

Puisque $|\mu|(E) = |\mu|(A) + |\mu|(B)$, l'un des deux est infini. Echanger au besoin les ròles de A et B pour conclure. ■

Pour prouver le Théorème 2, nous allons raisonner par l'absurde et supposer $|\mu|(\Omega) = \infty$. On pose $B_0 = \Omega$. Par le Lemme 2, de façon inductive, on construit deux suites $(A_n)_{n\geq 1}$ et $(B_n)_{n\geq 1}$ dans $\mathcal{B}$ telles que

– $(B_n) \downarrow$ et pour chaque n, $|\mu|(B_n) = \infty$;

– les A_n sont deux à deux disjoints, et $|\mu(A_n)| > 1$.

Posons alors $A = \cup A_n$: on a $\mu(A) = \sum_n \mu(A_n)$, et donc par convergence de la série, il faut que $\mu(A_n) \to 0$. Or $|\mu(A_n)| > 1$. ■

Notons $\mathcal{M}(\Omega, \mathcal{B})$ l'ensemble des mesures complexes sur $(\Omega, \mathcal{B})$. C'est un $\mathbb{C}$-e.v. muni des lois

$$(\lambda\mu + \beta\nu)(A) = \lambda\mu(A) + \beta\nu(A),\ A \in \mathcal{B}.$$

Il est normé par

$$\|\mu\| = |\mu|(\Omega).$$

Pour information, une **mesure signée** est une mesure complexe à valeurs réelles. Les **parties positives et négatives d'une mesure signée** sont définies par

$$\mu^+ = \frac{1}{2}(|\mu| + \mu) \text{ et } \mu^- = \frac{1}{2}(|\mu| - \mu).$$

Ce sont des mesures positives finies, et l'on a la **décomposition de Jordan** de μ :

$$\mu = \mu^+ - \mu^-.$$

Exemple. Soit μ une mesure positive sur $(\Omega, \mathcal{B})$, et soit $f \in L^1(\mu)$. Alors

$$\lambda_f(E) = \int_E f d\mu,\ E \in \mathcal{B},$$

définit une mesure complexe sur $(\Omega, \mathcal{B})$ (utiliser ici non plus Beppo-Lévi, mais le théorème de convergence dominée).

2.2. Absolue continuité et singularité.

Soient λ et μ deux mesures complexes sur $(\Omega, \mathcal{B})$. Nous dirons que λ **est absolument continue relativement à** μ**, et nous noterons** $\lambda \ll \mu$, si

$$\mu(E) = 0 \Rightarrow \lambda(E) = 0,\ E \in \mathcal{B}.$$

Nous dirons que μ **est portée par** $A \in \mathcal{B}$ si $E \in \mathcal{B} \Rightarrow \mu(E) = \mu(E \cap A)$.

Nous dirons que λ **et** μ **sont singulières, ou étrangères, et nous noterons** $\lambda \perp \mu$, si il existe $A, B \in \mathcal{B}$, disjoints, tels que A porte λ et B porte μ.

Voici quelques propiétés :

PROPOSITION. *Supposons que μ, $\lambda, \lambda_1, \lambda_2$ soient des mesures complexes sur $(\Omega, \mathcal{B})$, avec μ positive. Alors*

(a) : *si λ est portée par A, $|\lambda|$ aussi;*
(b) : *si $\lambda_1 \perp \lambda_2$, alors $|\lambda_1| \perp |\lambda_2|$;*
(c) : *si $\lambda_1 \perp \lambda$ et $\lambda_2 \perp \lambda$, alors $\lambda_1 + \lambda_2 \perp \lambda$;*
(d) : *si $\lambda_1 \ll \lambda$ et $\lambda_2 \ll \lambda$, alors $\lambda_1 + \lambda_2 \ll \lambda$;*
(e) : *si $\lambda \ll \mu$, alors $|\lambda| \ll \mu$;*
(f) : *si $\lambda_1 \ll \lambda$ et $\lambda_2 \perp \lambda$, alors $\lambda_1 \perp \lambda_2$;*
(g) : *si $\lambda_1 \perp \lambda$ et $\lambda_1 \ll \lambda$, alors $\lambda \equiv 0$.*

PREUVE. Exercice. ■

2.3. Décomposition de Lebesgue et dérivée de Radon-Nikodym.

LEMME PRÉLIMINAIRE 1. *Soit $(H, |\ |)$ un espace de Hilbert sur le corps K réel ou complexe, avec le produit scalaire noté $< h, h' >$. Si $u \in \mathcal{L}(H, K)$, il existe un unique $h \in H$ tel que pour tout $t \in H$,*

$$u(t) = < t, h > .$$

PREUVE. Si $u \equiv 0$, prendre $h = 0$. Sinon $H = Ker(u) \oplus Ker(u)^{\perp}$, une somme directe orthogonale de sous-espaces fermés. Soit $v \in Ker(u)^{\perp}$ tel que $u(v) = 1$. Soit $s \in Ker(u)^{\perp}$: alors $s - u(s)v \in Ker(u) \cap Ker(u)^{\perp} = \{0\}$, et donc $s = u(s)v$.

Soit $k \in H$: alors on a la décompostion en somme directe $k = y + w$, $y \in Ker(u)$, $w \in Ker(u)^{\perp}$. Donc $k = y + u(w)v = y + u(k)v$. Et

$$< k, v > = < y, v > + u(k) < v, v > = u(k)|v|^2.$$

Choisir $h = v \ / \ |v|^2$, et vérifier l'unicité. ■

LEMME PRÉLIMINAIRE 2. *Soit μ une mesure positive finie sur $(\Omega, \mathcal{B})$, et $\Delta \subset \mathbb{C}$ un sous-ensemble fermé. Soit $f \in \mathcal{L}^1(\mu)$ et supposons que pour tout $E \in \mathcal{B}$, si $\mu(E) > 0$,*

$$A_E(f) := \frac{1}{\mu(E)} \int f \mathbf{1}_E d\mu \in \Delta.$$

Alors $\mu\{\omega \in \Omega : f \notin \Delta\} = 0$.

PREUVE. Soit B une boule fermée de centre α et de rayon $r > 0$ incluse dans Δ^c. Il suffit de montrer que pour une telle boule, si $E = f^{-1}B$, $\mu(E) = 0$. Sinon, nous aurions

$$|A_E(f) - \alpha| = \left| \frac{1}{\mu(E)} \int \mathbf{1}_E (f - \alpha) d\mu \right| \leq \frac{1}{\mu(E)} \int_E |f - \alpha| d\mu \leq r,$$

et donc $A_E(f) \in B \subset \Delta^c$, ce qui est absurde. ■

Théorème de Radon-Nikodym. *Soient λ et μ deux mesures complexes sur $(\Omega, \mathcal{B})$, avec μ positive.*

(i) : il existe un unique couple (λ_s, λ_a) de mesures complexes sur $(\Omega, \mathcal{B})$ tel que $\lambda_s \perp \mu$, $\lambda_a \ll \mu$, et $\lambda = \lambda_s + \lambda_a$.

(ii) il existe un unique $h \in L^1(\mu)$ tel que pour $A \in \mathcal{B}$, $\lambda_a(A) = \int \mathbf{1}_A h d\mu$. On appelle h le dérivée de Radon-Nikodym de λ_a relativement à μ, et on note

$$d\lambda_a = hd\mu \text{ ou encore } h = d\lambda_a \,/\, d\mu.$$

Preuve. Commençons par l'unicité dans (i). Si un autre couple existe, alors $\lambda_a - \lambda'_a = \lambda'_s - \lambda_s$, qui, d'après la proposition précédente, est une mesure complexe à la fois absolument continue et étrangère relativement à μ. Et donc elle est nulle.

Pour l'existence, commençons par remarquer que $\lambda = Re(\lambda)^+ - Re(\lambda)^- + i(Im(\lambda)^+ - Im(\lambda)^-)$, et que donc si nous prouvons l'existence dans le cas de λ positive, elle en découlera dans le cas général.

Si $\lambda \geq 0$, alors $\nu = \lambda + \mu$ est une mesure positive finie sur $(\Omega, \mathcal{B})$. Il en découle (indicatrices, simples, puis Beppo-Lévi pour limites croissantes de simples positives), que si $f : \Omega \to \mathbb{R}$ est mesurable positive, puis dans $\mathcal{L}^1(\nu)$, alors

$$\int f d\nu = \int f d\lambda + \int f d\mu.$$

Si $f \in L^2(\nu)$, alors par Cauchy-Schwarz,

$$\left| \int f d\lambda \right| \leq \int |f| d\lambda \leq \int |f| d\nu \leq \sqrt{\int |f|^2 d\nu} \sqrt{\nu(\Omega)},$$

et donc $f \mapsto \int f d\lambda$ est une forme linéaire continue sur $L^2(\nu)$.

Puisque $L^2(\nu)$ est un Hilbert, il s'ensuit qu'il existe un unique $g \in L^2(\nu)$ tel que pour toute $f \in L^2(\nu)$,

$$\int f d\lambda = \int fg d\nu = \int fg d\lambda + \int fg d\mu.$$

Soit $E \in \mathcal{B}$ tel que $\nu(E) > 0$, faisons $f = \mathbf{1}_E$: alors $\lambda(E) = \int_E g d\nu$, et puisque $0 \leq \lambda(E) \leq \nu(E)$, il s'ensuit que

$$A_E(g) = \frac{1}{\nu(E)} \int_E g d\nu \in [0, 1].$$

Donc on peut supposer $0 \leq g \leq 1$, et alors on a pour $f \in L^2(\nu)$,

$$\int f(1-g) d\lambda = \int fg d\mu. \qquad (\star)$$

Posons $A = \{g < 1\}$ et $B = \{g = 1\}$. Puis posons

$$\lambda_a(E) = \lambda(E \cap A),\ \lambda_s(E) = \lambda(E \cap B).$$

Alors on a $\lambda = \lambda_a + \lambda_s$, et $\lambda_a \perp \lambda_s$.

Si $f = \mathbf{1}_B$, alors dans $(\star)$ on obtient $\mu(B) = 0$. Donc puisque λ_s est portée par B, on a $\lambda_s \perp \mu$.

Soit $n \geq 1$ et prenons $f = (1 + g + \ldots + g^n)\mathbf{1}_E$ avec $E \in \mathcal{B}$, dans $(\star)$. Alors

$$\int_E (1 - g^{n+1})d\lambda = \int_E g(1 + \ldots + g^n)d\mu.$$

Le membre de gauche ci-dessus converge vers $\lambda_a(E)$, et celui de droite vers $\int_E g(\sum_{n\geq 0} g^n)d\mu$. Prendre enfin

$$h = g(1 + g + g^2 + \ldots).$$

Il nous faut conclure par l'unicité de h. Supposons que $k \in L^1(\mu)$ soit tel que pour tout $A \in \mathcal{B}$, $\int_A hd\mu = \int_A kd\mu$. Alors $h - k \in L^1(\mu)$, et quel que soit A, avec $\mu(A) > 0$,

$$\frac{1}{\mu(A)} \int_A (h - k)d\mu \in \Delta = \{0\},$$

ce qui montre $h - k = 0$ μ-p.s., par le Lemme préliminaire 2. ■

Extension des hypothèses du théorème précédent : (exercice)
il reste vrai pour μ une mesure σ-finie sur $(\Omega, \mathcal{B})$.

Exemple : nous avons vu que si $0 \leq f$ est mesurable et Lebesgue intégrable sur $\mathbb{R}$, et que $\int f dm = 1$, alors $A \mapsto \mathbb{P}_f(A) = \int f\mathbf{1}_A dm$ définit une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On sait (Licence) que $m(A) = 0 \Rightarrow f\mathbf{1}_A = 0\ m - p.p. \Rightarrow \int f\mathbf{1}_A dm = \mathbb{P}_f(A) = 0$, et donc $\mathbb{P}_f \ll m$.

Donc $f = d\mathbb{P}_f \,/\, dm \in L^1(m)$ d'après le théorème précédent.

2.4. Quelques compléments.

Le résultat suivant éclaire la terminologie "absolument continu":

Théorème 1. *Soient μ et λ deux mesures complexes sur $(\Omega, \mathcal{B})$, avec μ positive. Alors les deux conditions suivantes sont équivalentes :*

(a) : $\lambda \ll \mu$;

(b) : $\forall \varepsilon > 0$, $\exists \delta > 0$, $E \in \mathcal{B}$ et $\mu(E) < \delta \Rightarrow |\lambda(E)| < \varepsilon$.

Preuve. Supposons (b), et soit E tel que $\mu(E) = 0$. Alors quelque soit $\varepsilon > 0$, $|\lambda(E)| < \varepsilon$, et donc $\lambda(E) = 0$.

Supposons non (b) : alors il existe $\varepsilon > 0$ et pour chaque n un ensemble $E_n \in \mathcal{B}$ tel que $\mu(E_n) < 2^{-n}$ et $|\lambda(E_n)| \geq \varepsilon$. Alors $|\lambda|(E_n) \geq \varepsilon$. Posons $A = \limsup_n E_n$. Alors $A = \cap_n \downarrow (\cup_{k \geq n} E_k)$, et comme $\mu(\cup_{k \geq n} E_k) \leq \sum_{k \geq n} 2^{-k} = 2^{1-n}$, il vient que $\mu(A) = 0$. D'autre part $|\lambda|(A) = \lim_n |\lambda|(\cup_{k \geq n} E_k) \geq \varepsilon > 0$, et donc λ n'est pas absolument continue relativement à μ. ■

Théorème 2. *Soit μ une mesure complexe sur $(\Omega, \mathcal{B})$. Alors $\mu \ll |\mu|$ et*

$$\big| d\mu \;/\; d|\mu| \big| = 1 \; |\mu| - p.s..$$

Preuve. L'affirmation $\mu \ll |\mu|$ est triviale. Et $|\mu|$ est une mesure positive finie sur $(\Omega, \mathcal{B})$. Soit donc $h = d\mu \;/\; d|\mu| \in L^1(|\mu|)$. Notons $A_r = \{|h| < r\}$, $r > 0$. Soit (E_i) une partition de A_r. Alors

$$\sum_i |\mu(E_i)| = \sum_i \left| \int_{E_i} h d|\mu| \right| \leq r \sum_i |\mu|(E_i) = r|\mu|(A_r),$$

et donc, en passant au sup sur les partitions, on a $|\mu|(A_r) \leq r|\mu|(A_r)$. Donc si $r < 1$, on a $|\mu|(A_r) = 0$. Et donc $|h| \geq 1$.

D'autre part, si $|\mu|(E) > 0$, alors

$$\left| \frac{1}{|\mu|(E)} \int_E h d|\mu| \right| = \frac{|\mu(E)|}{|\mu|(E)} \leq 1,$$

et par notre second lemme préliminaire ($\{|z| = 1\}$ est fermé dans $\mathbb{C}$), cela impose $|h| \leq 1$. ■

Théorème 3. *Soit μ une mesure positive sur $(\Omega, \mathcal{B})$, et $g \in L^1(\mu)$. Alors si la mesure complexe λ est définie par*

$$\lambda(E) = \int_E g d\mu,$$

on a $|\lambda|(E) = \int_E |g| d\mu$, $E \in \mathcal{B}$.

Preuve. On sait qu'il existe $h \in L^1(|\lambda|)$ telle que $|h| = 1$ et $h = d\lambda \;/\; d|\lambda|$. Alors $hd|\lambda| = gd\mu$, et donc $d|\lambda| = \bar{h} g d\mu$.

Mais $|\lambda|$ et μ sont positives donc $\bar{h} g \geq 0$, et donc $\bar{h} g = |\bar{h} g| = |\bar{h}||g| = |g|$. ■

Théorème de décomposition de Hahn. *Soit μ une mesure signée sur $(\Omega, \mathcal{B})$. Alors il existe $(A, B) \in Part(\Omega)$ telle que les variations positives et négatives de μ satisfassent*

$$\mu^+(E) = \mu(A \cap E) \text{ et } \mu^-(E) = -\mu(E \cap B),\ E \in \mathcal{B}.$$

Preuve. On a $d\mu = hd|\mu|$ avec $|h| = 1$. Comme μ est réelle, h l'est aussi. Etant de module 1, h ne peut valoir que ± 1. Posons $A = \{h = 1\}$ et $B = \{h = -1\}$. Alors

$$\frac{1}{2}(1 + h) = \mathbf{1}_A \text{ et } \frac{1}{2}(1 - h) = \mathbf{1}_B,$$

et donc puisque $\mu^+ = \frac{1}{2}(|\mu| + \mu)$, il s'ensuit que

$$\mu^+(E) = \frac{1}{2} \int_E (1 + h) d|\mu| = \int_{E \cap A} hd|\mu| = \mu(E \cap A).$$

Comme $\mu(E) = \mu(E \cap A) + \mu(E \cap B)$, et $\mu = \mu^+ - \mu^-$, il s'ensuit que $\mu^-(E) = -\mu(E \cap B)$. ■

Corollaire : minimalité dans la décomposition de Jordan. *Soit μ signée sur $(\Omega, \mathcal{B})$, et supposons que λ_1 et λ_2 sont deux mesures positives finies sur $(\Omega, \mathcal{B})$ telles que $\mu = \lambda_1 - \lambda_2$. Alors $\lambda_1 \geq \mu^+$ et $\lambda_2 \geq \mu^-$.*

Preuve. Puisque $\mu \leq \lambda_1$, on a $\mu^+(E) = \mu(E \cap A) \leq \lambda_1(E \cap A) \leq \lambda_1(E)$. De même $\mu^-(E) = -\mu(E \cap B) \leq \lambda_2(E \cap B) \leq \lambda_2(E)$. ■

Exercices.

1) : supposons que λ, β, et μ soient trois mesures complexes sur $(\Omega, \mathcal{B})$, et que β et μ soient positives. Montrer que si $\lambda \ll \beta \ll \mu$, alors

$$\frac{d\lambda}{d\mu} = \left(\frac{d\lambda}{d\beta}\right)\left(\frac{d\beta}{d\mu}\right).$$

2) : supposons que $\mu \perp \nu$ soient deux mesures complexes positives étrangères sur $(\Omega, \mathcal{B})$. Supposons que λ et β soient deux mesures complexes sur $(\Omega, \mathcal{B})$ telles que $\lambda \ll \mu$ et $\beta \ll \nu$. Montrer qu'alors

$$\lambda + \beta \ll \mu + \nu, \text{ et que } \frac{d(\lambda + \beta)}{d(\mu + \nu)} = \left(\frac{d\lambda}{d\mu}\right) + \left(\frac{d\beta}{d\nu}\right) \text{ !!}$$

3) : notons $p_1(x,y)=x$ et $p_2(x,y)=y$ les deux projections canoniques de $\mathbb{R}^2$ dans $\mathbb{R}$. Soit μ une mesure complexe positive sur $(\mathbb{R}^2,\mathcal{B}(\mathbb{R}^2))$, et soit $\lambda \ll \mu$ une mesure complexe. Montrer que $\lambda_{p_i} \ll \mu_{p_i}$.

Supposons en plus que μ soit une mesure produit, i.e. que $\mu = \frac{1}{\mu(\mathbb{R}^2)}\mu_{p_1}\otimes\mu_{p_2}$. Montrer qu'alors μ_{p_1}-presque sûrement en x,

$$\frac{d\lambda_{p_1}}{d\mu_{p_1}}(x) = \frac{1}{\mu(\mathbb{R}^2)}\int \frac{d\lambda}{d\mu}(x,y)d\mu_{p_2}(y).$$

4) : soit μ la mesure complexe sur $[0,2\pi]$ absolument continue relativement à la mesure de Lebesgue m et dont la dérivée est donnée par

$$d\mu \Big/ dm(x) = \sin x.$$

Déterminer la décomposition de Jordan de μ.

5) : soit μ la mesure sur $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ définie par

$$\mu(A) = \left(\sum_{\substack{r\in\mathbb{Q}\cap A\cap[0,1],\\ r>0, r=p/q, p\wedge q=1}} \frac{1}{q2^q}\right) + \int_{A\cap[0,1]} 3x^2dx,\ A\in\mathcal{B}.$$

Montrer que c'est une mesure complexe portée par $[0,1]$. Que vaut $\mu(\mathbb{R})$? Calculer $\mathbb{E}_\mu(x)$ puis $\mathbb{E}_\mu(x\mathbf{1}_\mathbb{Q}(x))$.

Probabilités conditionnelles et indépendance

3.1. Probabilités conditionnelles.

Nous verrons à l'occasion de l'étude des martingales comment le théorème de Radon-Nykodim permettra de donner une version "*locale*" des probabilités conditionnelles. Ce bref paragraphe ne fait que rappeler des éléments simples de la théorie classique.

Soit $(\Omega, \mathcal{B}, \mathbb{P})$ un espace probabilisé, $A, B \in \mathcal{B}$ et $\mathbb{P}(B) > 0$. La **probabilité de A sachant B, notée $\mathbb{P}(A|B)$, est définie par**

$$\mathbb{P}(A|B) = \mathbb{P}(A \cap B) \;/\; \mathbb{P}(B).$$

Identité multiplicative. *Si $A_1, \dots, A_n \in \mathcal{B}$, alors*

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) \ldots \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

Preuve. Triviale. ■

Formule des probabilités totales. *Si (H_i) est une partition finie ou dénombrable de Ω par des éléments de $\mathcal{B}$ de mesures positives, alors*

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|H_i)\mathbb{P}(H_i).$$

Preuve. Triviale. ■

Formule de Bayes. *Sous les mêmes hypothèses,*

$$\mathbb{P}(H_i|A) = \frac{\mathbb{P}(H_i)\mathbb{P}(A|H_i)}{\sum_j \mathbb{P}(H_j)\mathbb{P}(A|H_j)}.$$

Preuve. Triviale. ■

3.2. Indépendance.

Nous dirons que A **est indépendant de** B**, noté** $A \perp\!\!\!\perp B$, si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Plus généralement, nous dirons qu'une suite dévènements $(A_i)_{i \in I}$ **est indépendante** si quel que soit $J \subset I$ fini,

$$\mathbb{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j).$$

Nous dirons que la suite $(A_i)_{i \in I}$ **est deux à deux indépendante** si quels que soient $i \neq j \in I$, $A_i \perp\!\!\!\perp A_j$.

Soit $(\mathcal{B}_i)_{i \in I}$ une famille finie ou dénombrable de parties de $\mathcal{B}$. Nous dirons que **la suite** $(\mathcal{B}_i)$ **est indépendante** si quelle que soit la suite $(A_i) \in \prod \mathcal{B}_i$, elle est indépendante.

Nous dirons que **la suite** $(\mathcal{B}_i)$ **est deux à deux indépendante** si quels que soient $i \neq j \in I$, $A_i \in \mathcal{B}_i$, et $A_j \in \mathcal{B}_j$, $A_i \perp\!\!\!\perp A_j$.

Théorème. *Soit $(\mathcal{A}_i)_{i \in I}$ une famille finie ou dénombrable de π-systèmes, indépendante. Alors la suite $(\sigma(\mathcal{A}_i))$ est indépendante.*

Preuve. Supposons $I = \mathbb{N}$. Montrons que pour chaque $n \geq 1$, $(\sigma(\mathcal{A}_i))_{1 \leq i \leq n}$ est indépendante. Pour cela soient $A_2, \ldots, A_n$ des éléments de $\mathcal{A}_2, \ldots, \mathcal{A}_n$, respectivement.

Soit $\mathcal{M}$ l'ensemble des éléments de $\sigma(\mathcal{A}_1)$, qui sont indépendants de $A^+ = A_2 \cap \ldots \cap A_n$. C'est une classe de parties de Ω stable par complémentation, union dénombrable disjointe, et qui contient Ω. C'est donc une classe monotone. En outre $\mathcal{A}_1 \subset \mathcal{M}$, et $\mathcal{A}_1$ est un π-système. Donc par le théorème $\pi - \lambda$, $\sigma(\mathcal{A}_1) \subset \mathcal{M}$.

En procédant ainsi de proche en proche, la conclusion arrive. ■

3.3. Lemmes de Borel-Cantelli.

Soit $(A_i)_{i \geq 1}$ une suite dans $\mathcal{B}$.

Premier Lemme de Borel-Cantelli. *Si $\sum_n \mathbb{P}(A_n) < \infty$, alors $\mathbb{P}(\limsup A_n) = 0$.*

Preuve. Notons $B_n = \cup_{p \geq n} A_p$: alors $\limsup A_n = \cap_n \downarrow B_n$, car $B_{n+1} \subset B_n$. Donc, par monotonie des probabilités, $\mathbb{P}(\limsup A_n) = \lim_n \downarrow \mathbb{P}(B_n)$. Or si $\sum_n \mathbb{P}(A_n) < \infty$,

$$0 \leq \mathbb{P}(B_n) \leq \sum_{p \geq n} \mathbb{P}(A_p) \to 0,$$

et donc. ■

Second Lemme de Borel-Cantelli. *Si de plus la suite* (A_n) *est indépendante, alors on a la réciproque : si* $\sum_n \mathbb{P}(A_n) = \infty$, $\mathbb{P}(\limsup A_n) = 1$.

Preuve. Montrons que $\mathbb{P}(\liminf A_n^c) = 0$. Si $C_n = \cap_{p \geq n} A_p^c$, alors $\liminf A_n^c = \cup_n \uparrow C_n$, et donc $\mathbb{P}(\liminf A_n^c) = \lim_n \uparrow \mathbb{P}(C_n)$. Or si $x \geq 0$, $1 - x \leq e^{-x}$, et donc

$$\mathbb{P}(C_n) = \lim_k \prod_{i=n}^{n+k} \mathbb{P}(A_i^c) = \lim_k \prod_{i=n}^{n+k} (1 - \mathbb{P}(A_i)) \leq \lim_k e^{-\sum_{i=n}^{n+k} \mathbb{P}(A_i)} = 0!$$

Donc pour chaque n, $\mathbb{P}(C_n) = 0$, et donc. ■

Exercice : sur $(\{0,1\}^{\mathbb{N}}, \mathcal{T} \cap \{0,1\}^{\mathbb{N}}, \mathbb{P}_\infty)$ (cf. **1.1.4.**), posons $A_k = \{\omega_k = 1\}$. Vérifier que la suite (A_k) est indépendante. Evaluer $\sum_k \mathbb{P}_\infty(A_k)$, et en déduire que $\mathbb{P}_\infty$-presque sûrement, ω possède une infinité de 1 et de 0. Utiliser deux méthodes : l'une par le premier lemme de Borel-Cantelli, l'autre par le second.

3.4. Loi $\mathbf{0-1}$ de Kolmogorov.

Loi $\mathbf{0-1}$ de Kolmogorov. *Soit* $(B_n)_{n\geq 0}$ *une suite indépendante de* $\mathcal{B}$. *Notons*

$$\mathcal{B}_\infty = \cap_{n\geq 1} \sigma(B_n, B_{n+1}, \dots),$$

la σ*-algèbre des évènements asymptotiques associés. Alors si* $A \in \mathcal{B}_\infty$,

$$\mathbb{P}(A) = 0 \ \textit{ou} \ \mathbb{P}(A) = 1.$$

Autrement dit, $\mathcal{B}_\infty$ *est triviale.*

Preuve. Nous aurons besoin du

Lemme. *Supposons que la suite double*

$$\begin{array}{ccc} A_{11} & A_{12} & \dots \\ A_{21} & A_{22} & \dots \\ \vdots & \vdots & \end{array}$$

soit indépendante, où le nombre de lignes ou de colonnes est fini ou dénombrable. Alors si $\mathcal{B}_i$ *est la* σ*-algèbre engendrée par la* $i^{\text{ième}}$ *ligne, la suite* $(\mathcal{B}_i)$ *est indépendante.*

Preuve du Lemme. Notons $\mathcal{A}_i$ l'ensemble des intersections finies d'éléments de la $i^{\text{ième}}$ ligne, alors c'est un π-système et $\sigma(\mathcal{A}_i) = \mathcal{B}_i$.

Il suffit donc de montrer que la suite $(\mathcal{A}_i)$ est indépendante. Soit donc I un ensemble fini d'indices de lignes, et pour chaque $i \in I$, J_i un ensemble fini d'indices de colonnes et $C_i = \cap_{j \in I_i} A_{ij}$ l'élément de $\mathcal{A}_i$ correspondant.

Alors
$$\mathbb{P}(\cap_i C_i) = \mathbb{P}(\cap_i(\cap_j A_{ij})) = \prod_i \prod_{j \in J_i} \mathbb{P}(A_{ij}) = \prod_i \mathbb{P}(C_i).$$

■

Fin de la preuve de la loi $0-1$: soit $A \in \mathcal{B}_\infty$, et montrons que $\mathbb{P}(A) = \mathbb{P}(A)^2$, ce qui découlera si $A \perp\!\!\!\perp A$. Pour chaque $n \geq 1$, la suite $(A, B_1, \dots, B_n)$ est indépendante puisque $A \in \sigma(B_{n+1}, B_{n+2}, \dots)$. Donc la suite $(A, B_1, B_2, \dots)$ est indépendante, et donc, par le Lemme préliminaire ci-dessus, $A \perp\!\!\!\perp \sigma(B_1, B_2, \dots)$. Or $A \in \sigma(B_1, B_2, \dots)$, et donc $A \perp\!\!\!\perp A$. ■

Exemples importants : si $(A_n)_{n\geq 0}$ est indépendante, alors $\limsup_n A_n$ et $\liminf A_n$ sont des évènements asymptotiques, i.e. dans $\cap_n \sigma(A_n, A_{n+1}, \dots)$. Cela se voit en écrivant
$$\limsup_n A_n = \cap_n \downarrow \cup_{p\geq n} A_p \quad \text{et} \quad \liminf_n A_n = \cup_n \uparrow \cap_{p \geq n} A_p.$$

Variables et vecteurs aléatoires

4.1. Variables et vecteurs aléatoires.

On se donne un espace probabilisé $(\Omega, \mathcal{B}, \mathbb{P})$, et $X_i : \Omega \to \mathbb{R}$, $1 \leq i \leq d$, d-applications $(\mathcal{B}, \mathcal{B}(\mathbb{R}))$-mesurables. Ce sont des **variables aléatoires**, par définition.

Un **vecteur aléatoire** est une application $(X_1, \dots, X_d)\Omega \to \mathbb{R}$ dont les composantes sont des variables aléatoires.

Souvent on adopte la notation abrégée **v.a.** pour *variable aléatoire* ou *vecteur aléatoire.*

Du point de vue probabiliste, ce qui nous intéressera le plus seront les **lois des variables ou vecteurs aléatoires** : $\mathbb{P}_X$ ou $\mathbb{P}_{(X_1,\dots,X_d)}$.

L'espérance d'un vecteur aléatoire est définie par

$$\mathbb{E}((X_1, \dots, X_d)) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d)).$$

Pour les vecteurs aléatoires, nous bénéficions des mêmes théorèmes généraux d'interversion de limite et d'intégrale que pour les variables aléatoires, à ceci près qu'il faut en adapter l'énoncé de sorte à ce qu'ils intègrent simultanément les conditions requises composante à composante, à savoir pour chaque X_i.

Les variables aléatoires ne nous seront accessibles dans la pratique qu'au travers de leurs lois, qui sont des probabilités boréliennes sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ ($d > 1$ pour les vecteurs). De ce point de vue, le théorème suivant, qui résulte directement de la définition de $\mathbb{P}_X$, est essentiel :

Théorème de transfert. *Soit $(\Omega', \mathcal{B}')$ un autre espace probabilisable et soit $f : \Omega \to \Omega'$ une application $(\mathcal{B}, \mathcal{B}')$-mesurable. Alors pour toute application borélienne $g : \Omega' \to \mathbb{R}^d$, on a*

$$g \in \mathcal{L}^1(\mathbb{P}_f) \Longleftrightarrow g \circ f \in \mathcal{L}^1(\mathbb{P}).$$

Si tel est le cas, alors

$$\mathbb{E}_{\mathbb{P}}(g \circ f) = \mathbb{E}_{\mathbb{P}_f}(g).$$

Commentaire : supposons que nous ne connaissions de la variable aléatoire $X : \Omega \to \mathbb{R}$ que sa loi $\mathbb{P}_X$ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors, bien que nous ne connaissions pas l'application X elle-même, nous pouvons décider de l'intégrabilité de $g(X)$ où $g : \mathbb{R} \to \mathbb{R}$ est borélienne, et en plus calculer $\mathbb{E}_{\mathbb{P}}(g(X))$. Il suffit en effet pour cela de vérifier que $g \in \mathcal{L}^1(\mathbb{P}_X)$, et si c'est le cas, de calculer

$$\mathbb{E}_{\mathbb{P}}(g(X)) = \mathbb{E}_{\mathbb{P}_X}(g) = \int_{\mathbb{R}} g(x) d\mathbb{P}_X(x).$$

Preuve. Par définition, si $g : \Omega' \to \mathbb{R}^d$, son intégrabilité équivaut à celle de toutes ses composantes, et son intégrale est le vecteur des intégrales de ses composantes. Donc restreignons nous à ne traiter que le cas de $g : \Omega' \to \mathbb{R}$.

Par définition de $\mathbb{P}_f$, si $A \in \mathcal{B}'$, et $g = \mathbf{1}_A$, $g \circ f = \mathbf{1}_{f^{-1}A}$ sont intégrables, et

$$\mathbb{E}_{\mathbb{P}}(g \circ f) = \mathbb{P}(f^{-1}A) = \mathbb{P}_f(A) = \mathbb{E}_{\mathbb{P}_f}(g).$$

Donc l'énoncé est vrai pour g une variable aléatoire simple, i.e.

$$g = \sum_{finie} \alpha_i \mathbf{1}_{g^{-1}\{\alpha_i\}},$$

où $(g^{-1}\{\alpha_i\})$ est une partition $\mathcal{B}'$-mesurable de Ω'.

Si g est positive alors elle est limite d'une suite croissante (g_n) de variables aléatoires simples et positives, et donc par Beppo-Lévi, et la validité de notre énoncé pour les fonctions simples,

$$\mathbb{E}_{\mathbb{P}_f}(g) = \lim_n \uparrow \mathbb{E}_{\mathbb{P}_f}(g_n) = \lim_n \uparrow \mathbb{E}_{\mathbb{P}}(g_n \circ f) = \mathbb{E}_{\mathbb{P}}(g \circ f),$$

car $g_n \circ f \to g \circ f$ simplement, en étant minoré par zéro, et en croissant.

Il en résulte notamment que si g est positive,

$$g \in \mathcal{L}^1(\mathbb{P}_f) \Leftrightarrow \mathbb{E}_{\mathbb{P}_f}(g) < \infty \Leftrightarrow \mathbb{E}_{\mathbb{P}}(g \circ f) < \infty \Leftrightarrow g \circ f \in \mathcal{L}^1(\mathbb{P}).$$

Donc l'énoncé est vrai pour les variables aléatoires positives. Dans le cas général, on passe par les parties positives $g^+ = \max(0, g)$ et $g^- = \max(0, -g)$. On remarque que $(g \circ f)^+ = g^+ \circ f$ et $(g \circ f)^- = g^- \circ f$, et on écrit

$$\begin{aligned}
& g \in \mathcal{L}^1(\mathbb{P}_f) \\
& \Leftrightarrow g^+ \text{ et } g^- \in \mathcal{L}^1(\mathbb{P}_f) \Leftrightarrow \max(\mathbb{E}_{\mathbb{P}_f}(g^+), \mathbb{E}_{\mathbb{P}_f}(g^-)) < \infty \\
& \Leftrightarrow \max(\mathbb{E}_{\mathbb{P}}(g^+ \circ f), \mathbb{E}_{\mathbb{P}}(g^- \circ f)) < \infty \Leftrightarrow (g \circ f)^+ \text{ et } (g \circ f)^- \in \mathcal{L}^1(\mathbb{P}) \\
& \Leftrightarrow g \circ f \in \mathcal{L}^1(\mathbb{P}).
\end{aligned}$$

Puis on conclut par $\mathbb{E}_{\mathbb{P}_f}(g) = \mathbb{E}_{\mathbb{P}_f}(g^+) - \mathbb{E}_{\mathbb{P}_f}(g^-) = \mathbb{E}_{\mathbb{P}}((g \circ f)^+) - \mathbb{E}_{\mathbb{P}}((g \circ f)^-) = \mathbb{E}_{\mathbb{P}}(g \circ f)$. ■

Exemple : supposons que X prenne des valeurs entières et suive une loi de Poisson de paramètre $\lambda > 0$. Calculer $\mathbb{E}_{\mathbb{P}}(X)$! A priori, on ne connaît pas X et on peut être surpris de devoir cependant chercher à calculer $\int_\Omega X d\mathbb{P}$, d'autant qu'on ne connaît pas non plus ni Ω, ni $\mathcal{B}$, ni $\mathbb{P}$!

Ce n'est pas un problème, puisqu'on nous indique que $\mathbb{P}_X$ est la loi de Poisson de paramètre $\lambda > 0$. Certes c'est un loi de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, mais bien plutôt sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, puisque $X \in \mathbb{N}$ $\mathbb{P}$-p.s.. Et nous sont donnés les valeurs $\mathbb{P}_X(\{n\}) = e^{-\lambda}\lambda^n \,/\, n!$.

Donc par construction de l'opérateur d'intégration $\mathbb{E}_{\mathbb{P}_X}$, il vient, par application du théorème de transfert,

$$\mathbb{E}_{\mathbb{P}}(X) = \mathbb{E}_{\mathbb{P}_X}(x) = \int_{\mathbb{R}} x d\mathbb{P}_X(x) = \sum_{n \geq O} n\mathbb{P}_X(\{n\}) = \ldots = \lambda,$$

où les points de suspension sont à compléter par le lecteur.

4.2. Quelques définitions spécifiques aux variables aléatoires.
L'espérance : l'espérance de la variable aléatoire X est définie si $X \in \mathcal{L}^1(\mathbb{P})$, notée $\mathbb{E}(X)$ ou $\mathbb{E}_{\mathbb{P}}(X)$ si le contexte suscite de spécifier $\mathbb{P}$, et vaut

$$\mathbb{E}(X) = \mathbb{E}_{\mathbb{P}_X}(x) = \int_{\mathbb{R}} x d\mathbb{P}_X(x).$$

C'est un **paramètre de localisation**, parfois appelé **valeur moyenne** ou encore simplement la **moyenne**. Bien entendu, en tant qu'intégrale, l'espérance est **linéaire**.
Exemple : si la loi de X est normale centrée réduite (de densité $d\mathbb{P}_X \ / \ dm(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$), alors

$$\mathbb{E}(X) = \int_{\mathbb{R}} \frac{x e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = 0,$$

par imparité.
Exemple : si la loi de X est Cauchy de paramètre 1 (i.e. $d\mathbb{P}_X \ / \ dm(x) = \frac{1}{\pi(1+x^2)}$), alors X n'a pas d'espérance. Pourquoi ?

Variance : si $X \in \mathcal{L}^2(\mathbb{P})$, alors elle a une **variance**, notée $\mathbf{Var(X)}$, définie par

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

(remarquons que $X \in \mathcal{L}^2(\mathbb{P}) \Rightarrow X \in \mathcal{L}^1(\mathbb{P})$). C'est un **paramètre de dispersion**.

LEMME. $Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$ *si elle existe.*

PREUVE. Les constantes sont dans $\mathcal{L}^2(\mathbb{P})$ et donc $X - \mathbb{E}(X)$ existe et se trouve dans $\mathcal{L}^2(\mathbb{P})$. Alors il suffit de développer le carré dans l'espérance, et d'utiliser la linéarité de l'espérance. ∎

Propriétés de la variance. La variance est invariante par translation i.e.

$$\mathbf{Var(X + c) = Var(X)},$$

et une variable sans dispersion est constante $\mathbb{P}$-p.s., à savoir que

$$\mathbf{Var(X) = 0 \Rightarrow X = \mathbb{E}(X)\ \mathbb{P} - p.s..}$$

Ecart-type *: c'est la racine de la variance,* $\sigma(\mathbf{X}) = \sqrt{\mathbf{Var(X)}}$.

Matrice des variances-covariances *: dans le cas d'un vecteur aléatoire* $\mathcal{X} = (X_1, \ldots, X_d)$, *avec* $X_i \in \mathcal{L}^2(\mathbb{P})$, *on remplace la variance par la matrice*

$$\mathbf{Var}(\mathcal{X}) = (\mathbf{Cov(X_i, X_j)})_{\mathbf{1 \leq i,j \leq d}},$$

où $Cov(X_i, X_j)$ désigne la **covariance des variables** X_i **et** X_j, définie, pour deux variables aléatoires $Y, Z \in \mathcal{L}^2(\mathbb{P})$, par

$$\mathbf{Cov(Y, Z)} = \mathbb{E}\left((\mathbf{Y} - \mathbb{E}(\mathbf{Y}))(\mathbf{Z} - \mathbb{E}(\mathbf{Z}))\right).$$

Fonction de répartition : c'est la fonction $F_X : \mathbb{R} \to [0, 1]$ définie par

$$\mathbf{F_X(t)} = \mathbb{P}(\mathbf{X} \leq \mathbf{t}) = \mathbb{P}_\mathbf{X}(]-\infty, \mathbf{t}]), \ \ \mathbf{t} \in \mathbb{R}.$$

C'est un exercice instructif de démontrer que F_X a les propriétés suivantes :

$$\begin{cases} (i) & : \ F_X \text{ est croissante;} \\ (ii) & : \ F_X \text{ est continue à droite;} \\ (iii) & : \ \lim_{t\to-\infty} F_X(t) = 0; \\ (iv) & : \ \lim_{t\to+\infty} F_X(t) = 1. \end{cases}$$

Fonction de répartition d'un vecteur aléatoire : soit $\mathcal{X} = (X_1, \ldots, X_d)$ un vecteur aléatoire. Sa fonction de répartition est par définition l'application $F_\mathcal{X} : \mathbb{R}^d \to [0, 1]$ définie par

$$\mathbf{F}_\mathcal{X}(\mathbf{t_1}, \ldots, \mathbf{t_d}) = \mathbb{P}((\mathbf{X_1} \leq \mathbf{t_1}) \wedge \ldots \wedge (\mathbf{X_d} \leq \mathbf{t_d})).$$

LEMME. *Si $F : \mathbb{R} \to [0, 1]$ satisfait aux propriétés $(i$ —$iv)$ énoncées précédemment, alors il existe $(\Omega, \mathcal{B}, \mathbb{P})$ et une variable aléatoire X sur $(\Omega, \mathcal{B})$ tels que $F = F_X$.*

De plus si X et Y sont deux variables aléatoires telles que $F_X = F_Y$, alors elles ont même loi, i.e. $\mathbb{P}_X = \mathbb{P}_Y$.

PREUVE. Remarquons que puisque F est croissante $((i))$, elle a une limite à droite $F(a^+)$ et une limite à gauche $F(a^-)$ en tout point a de $\mathbb{R}$: notons $\delta_F(a) = F(a^+) - F(a^-)$ le saut de discontinuité de F en a (remarquons que $F(a^+) = F(a)$ par (ii)). Remarquons aussi que $\sum_{a\in\mathbb{R}} \delta_F(a) \leq 1$, par $(ii$ —$iv)$. Nous noterons $F(-\infty) = 0$ et $F(+\infty) = 1$.

Appelons $\mathcal{A}$ l'algèbre des unions finies disjointes d'intervalles de $\mathbb{R}$. Si I est un intervalle de $\mathbb{R}$, d'extrémités $a \leq b$, posons

$$\bar{\mathbb{P}}(I) = F(b) - F(a) - \delta_F(b)\mathbf{1}_{\mathbb{R}\setminus I}(b)\mathbf{1}_\mathbb{R}(b) + \delta_F(a)\mathbf{1}_I(a)\mathbf{1}_\mathbb{R}(a).$$

Cela peut paraître compliqué mais par exemple on a $\bar{\mathbb{P}}([c, c]) = \bar{\mathbb{P}}(\{c\}) = \delta_F(c)$ et $\bar{\mathbb{P}}([a, b[) = \bar{\mathbb{P}}(]a, b]) + \bar{\mathbb{P}}([a, a]) - \bar{\mathbb{P}}([b, b]) = F(b) - F(a) + \delta_F(a) - \delta_F(b)$ si $a, b, c \in \mathbb{R}$.

On vérifie sans peine que si $A \in \mathcal{A}$ et J est fini et $A = \cup_{j\in J} I_j$ avec les intervalles I_j deux à deux disjoints, alors en posant $\bar{\mathbb{P}}(A) = \sum_{j\in J} \bar{\mathbb{P}}(I_j)$, on obtient une bonne définition, à savoir que si $\cup_{j\in J} I_j = \cup_{j'\in J'} I'_{j'}$, alors $\sum_{j\in J} \bar{\mathbb{P}}(I_j) = \sum_{j'\in J'} \bar{\mathbb{P}}(I'_{j'})$.

On a de plus $\bar{\mathbb{P}}(\emptyset) = \bar{\mathbb{P}}(]a, a[) = F(a) - F(a) = 0$, et $\bar{\mathbb{P}}(\mathbb{R}) = F(+\infty) - F(-\infty) = 1$, et si A, A' sont disjoints dans $\mathcal{A}$, $\bar{\mathbb{P}}(A \cup A') = \bar{\mathbb{P}}(A) + \bar{\mathbb{P}}(A')$. De plus si $A, B \in \mathcal{A}$ et $A \subset B$, on vérifie sans peine que $\bar{\mathbb{P}}(B \setminus A) = \bar{\mathbb{P}}(B) - \bar{\mathbb{P}}(A)$. Donc $\bar{\mathbb{P}} : \mathcal{A} \to [0, 1]$ est une fonction additive, de masse totale 1, qui vérifie $\bar{\mathbb{P}}(\mathbb{R} \setminus A) = 1 - \bar{\mathbb{P}}(A)$.

Montrons que $\bar{\mathbb{P}}$ est une probabilité sur $\mathcal{A}$, à savoir qu'elle est *continue en* $\emptyset$. A cet effet, montrons d'abord que pour $A \in \mathcal{A}$,

$$\bar{\mathbb{P}}(A) = \sup_{\substack{K \subset A, \\ K \in \mathcal{A}, \\ K \text{ compact}}} \bar{\mathbb{P}}(K).$$

Puisqu'un tel A est une union finie d'intervalles, et qu'une union finie de compacts est compacte, montrons cette propriété simplement dans le cas particulier où A lui-même est un intervalle.

Par exemple supposons $A = [a, b[$ avec $a < b \in \mathbb{R}$. Alors $\bar{\mathbb{P}}(A) = F(b) - F(a) - \delta_F(b) + \delta_F(a) = F(b) - F(a) - (F(b) - F(b^-)) + F(a) - F(a^-) = F(b^-) - F(a^-)$. Soit $(\varepsilon_n) \downarrow 0$ avec $0 < \varepsilon_n < \frac{b-a}{2}$. Notons $K_n = [a, b - \varepsilon_n]$. Alors $K_n \subset A$ est compact et $\bar{\mathbb{P}}(K_n) = F(b - \varepsilon_n) - F(a^-)$, et puisque $\varepsilon_n \downarrow 0$, $F(b - \varepsilon_n) \to F(b^-)$, et donc $\lim_n \bar{\mathbb{P}}(K_n) = \bar{\mathbb{P}}(A)$.

Donc à l'instar de ce qui avait été fait pour la mesure de Lebesgue sur $[0, 1]$, ou à l'occasion de la preuve du théorème d'extension de Kolmogorov, nous avons ici montré que $\bar{\mathbb{P}}$ a une certaine **propriété de régularité sur les pavés de** $\mathbb{R}$.

C'était alors la clef de la preuve de la continuité en $\emptyset$ (qui permet d'appliquer le théorème d'extension), et cela le reste ici : supposons $(A_n) \in \mathcal{A}$ et que $A_n \downarrow \emptyset$. Soit $\varepsilon > 0$ et pour chaque n, soit K_n^0 compact dans $\mathcal{A}$ contenu dans A_n tel que $\bar{\mathbb{P}}(A_n \setminus K_n^0) \leq \varepsilon \ / \ 2^{n+1}$. Puis posons $K_n = \cap_{s \leq n} K_s^0$. Alors

$$\bar{\mathbb{P}}(A_n \setminus K_n) \leq \sum_{s \leq n} \bar{\mathbb{P}}(A_s \setminus K_s^0) \leq \varepsilon.$$

Et la suite (K_n) est une suite de compacts, décroissante, et d'intersection vide. Donc il existe n_0 tel que si $n \geq n_0$, $K_n = \emptyset$. Alors

$$n \geq n_0 \Rightarrow 0 \leq \bar{\mathbb{P}}(A_n) = \bar{\mathbb{P}}(A_n \setminus K_n) \leq \varepsilon,$$

ce qui entraîne que $\bar{\mathbb{P}}(A_n) \to 0$.

Nous n'avons plus qu'à appliquer le théorème d'existence pour déduire l'existence d'une unique probabilité $\tilde{\mathbb{P}}$ sur $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$ telle que $\tilde{\mathbb{P}} = \bar{\mathbb{P}}$ sur $\mathcal{A}$.

Alors posons $(\Omega, \mathcal{B}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \tilde{\mathbb{P}})$, et $X : \mathbb{R} \to \mathbb{R}$ définie par $X(x) = x$, autrement dit $X = Id_{\mathbb{R}}$. On a alors

$$F_X(t) = \mathbb{P}_X(] - \infty, t]) = \tilde{\mathbb{P}}(X^{-1}] - \infty, t]) = \tilde{\mathbb{P}}(] - \infty, t]) = \bar{\mathbb{P}}(] - \infty, t]) = F(t), \quad t \in \mathbb{R}.$$

Montrons à présent la seconde assertion du Lemme : soient X et Y deux v.a. telles que $F_X = F_Y$. Alors puisque $F_X(t) = \mathbb{P}_X(] - \infty, t])$ et que $F_Y(t) = \mathbb{P}_Y(] - \infty, t])$, il vient $\mathbb{P}_X = \mathbb{P}_Y$ sur la classe $\mathcal{M} = \{] - \infty, t] : t \in \mathbb{R}\}$.

Mais $\mathcal{M}$ est un π-système générateur de $\mathcal{B}(\mathbb{R})$, et donc par le théorème d'unicité, $\mathbb{P}_X = \mathbb{P}_Y$ sur $\mathcal{B}(\mathbb{R})$. ■

Remarque importante : *dans le cas d'un vecteur aléatoire, il faut rajouter une condition supplémentaire pour pouvoir caractériser les applications qui sont exactement les fonctions de répartitions des vecteurs aléatoires en dimension* $d > 1$.

Pour comprendre l'existence de cette condition supplémentaire, pensons à $\mathbb{R}^2$, $\mathcal{X} = (X_1, X_2)$, et soient $x_1 < y_1$ et $x_2 < y_2$. Alors $0 \leq \mathbb{P}((x_1 < X_1 \leq y_1) \wedge (x_2 < X_2 \leq y_2))$, et donc

$$\mathbf{F}_{\mathcal{X}}(\mathbf{y_1}, \mathbf{y_2}) + \mathbf{F}_{\mathcal{X}}(\mathbf{x_1}, \mathbf{x_2}) - \mathbf{F}_{\mathcal{X}}(\mathbf{x_1}, \mathbf{y_2}) - \mathbf{F}_{\mathcal{X}}(\mathbf{y_1}, \mathbf{x_2}) \geq \mathbf{0}.$$

La condition supplémentaire est une condition de "*positivité sur les contours*".

Fonction caractéristique d'une v.a. : soient $(t_1, \ldots, t_d)$ et $(s_1, \ldots, s_d) \in \mathbb{R}^d$. Notons $< s, t >= \sum_i s_i t_i$ leur produit scalaire. Alors si $\mathcal{X} = (X_1, \ldots, X_d)$ est un v.a., $|e^{i<t,\mathcal{X}>}| = 1$ pour tout $t \in \mathbb{R}^d$, et donc $e^{i<t,\mathcal{X}>} \in \mathcal{L}^1(\mathbb{P})$. On appelle **fonction caractéristique du v.a.** $\mathcal{X}$ l'application $\phi_{\mathcal{X}} : \mathbb{R}^d \to \mathbb{C}$ définie par

$$\phi_{\mathcal{X}}(\mathbf{t}) = \mathbb{E}(\mathbf{e}^{\mathbf{i}<\mathbf{t},\mathcal{X}>}), \quad \mathbf{t} \in \mathbb{R}^{\mathbf{d}}.$$

Nous verrons plus loin que la fonction caractéristique est effectivement caractéristique, au sens où elle caractérise $\mathbb{P}_{\mathcal{X}}$.

Fonction génératrice : supposons que la v.a. X prenne $\mathbb{P}$-p.s. des valeurs entières. Alors **sa fonction génératrice, notée** G_X, est définie par

$$\mathbf{G_X}(\mathbf{s}) = \sum_{\mathbf{n} \geq \mathbf{0}} \mathbf{s}^{\mathbf{n}} \mathbb{P}_{\mathbf{X}}(\{\mathbf{n}\}) = \mathbb{E}(\mathbf{s}^{\mathbf{X}}), \quad |\mathbf{s}| \leq \mathbf{1}.$$

Remarquons que la série converge normalement pour $|s| \leq 1$ puisque $G_X(1) = \mathbb{E}(\mathbf{1}) = 1$.

Lemme. *Soient* X *et* Y *deux v.a. prenant des valeurs entières et supposons que* $G_X = G_Y$. *Alors* X *et* Y *ont la même loi, i.e.* $\mathbb{P}_X = \mathbb{P}_Y$.

Preuve. G_X et G_Y sont analytiques et coïncident sur l'ouvert $\{|s| < 1\}$. Donc pour chaque $n \geq 0$, $n!\mathbb{P}_X(\{n\}) = G_X^{(n)}(0) = G_Y^{(n)}(0) = n!\mathbb{P}_Y(\{n\})$, et donc $\mathbb{P}_X = \mathbb{P}_Y$. ■

Nous renvoyons le lecteur en Annexe 1 où il trouvera une liste de lois usuelles et pourra déjà s'entraîner à en calculer les paramètres que nous venons d'introduire, à savoir

- *espérance,*
- *variance,*
- *fonction de répartition,*
- *fonction caractéristique,*
- *fonction génératrice le cas échéant.*

4.3. Les inégalités de base.

Nous notons, pour $p > 0$, et X une v.a., $|X|_p = \mathbb{E}(|X|^p)^{\frac{1}{p}}$ (rappelons que ce n'est une norme que sur les classes d'équivalence de v.a. presque sûrement égales, sinon ce n'est qu'une semi-norme).
Hölder : soient $p, q \geq 1$ conjugués, i.e. $\frac{1}{p} + \frac{1}{q} = 1$. Soit $X \in \mathcal{L}^p(\mathbb{P})$ et $Y \in \mathcal{L}^q(\mathbb{P})$. Alors

$$\mathbf{XY} \in \mathcal{L}^{\mathbf{1}}(\mathbb{P}) \textbf{ et } |\mathbf{XY}|_{\mathbf{1}} \leq |\mathbf{X}|_{\mathbf{p}} |\mathbf{Y}|_{\mathbf{q}}.$$

Minkowski : c'est l'inégalité triangulaire de $|\ |_p$. Si $X, Y \in \mathcal{L}^p(\mathbb{P})$, pour un certain $p \geq 1$, alors $X + Y \in \mathcal{L}^p(\mathbb{P})$ et de plus

$$|\mathbf{X} + \mathbf{Y}|_{\mathbf{p}} \leq |\mathbf{X}|_{\mathbf{p}} + |\mathbf{Y}|_{\mathbf{q}}.$$

Ces deux premières inégalités ne sont pas spécifiques aux espaces probabilisés (masse totale 1). Elles sont valables pour les espaces mesurés positifs.

Les suivantes par contre sont spécifiques. Avant d'énoncer la première d'entre elles, l'inégalité de Jensen, nous avons besoin de quelques notions préliminaires.

Soit $\phi :]a, b[\to \mathbb{R}$, avec $-\infty \leq a < b \leq +\infty$. Nous dirons qu'elle est **convexe** si

$$a < x < y < b \text{ et } 0 \leq \lambda \leq 1 \Rightarrow \phi((1-\lambda)x + \lambda y) \leq (1-\lambda)\phi(x) + \lambda\phi(y).$$

Autrement dit, si $X = (x, \phi(x))$, $Y = (y, \phi(Y))$ et si $Z = (z, \phi(z))$, avec $x < z < y$, alors le point du graphe de ϕ, Z, est sous le segment $[X, Y]$. La convexité de ϕ équivaut donc à ce que la pente de la droite (XZ) soit moindre que celle de la droite (ZY), ou encore,

$$\mathbf{a} < \mathbf{s} < \mathbf{t} < \mathbf{u} < \mathbf{b} \Rightarrow \frac{\phi(\mathbf{t}) - \phi(\mathbf{s})}{\mathbf{t} - \mathbf{s}} \leq \frac{\phi(\mathbf{u}) - \phi(\mathbf{t})}{\mathbf{u} - \mathbf{t}}. \tag{Q}$$

Lemme. *ϕ convexe entraîne ϕ continue.*

Preuve. Montrons que ϕ est continue à droite et à gauche en tout point $a < x < b$. Pour la continuité à droite en x, considérons $a < s < x < y < t < b$. Notons $S = (s, \phi(s))$ et $T = (t, \phi(t))$. Faire un dessin. Puisque X est au-dessous (SY), Y est au-dessus de (SX). Et Y est au-dessous de (XT). Si $y \to x^+$, le point limite $L = (x, \ell)$ doit continuer à être au-dessus de (SX), et au-dessous de (XT), ce qui impose $\phi(x) \leq \ell \leq \phi(x)$, donc $\ell = \phi(x)$, et ϕ est continue à droite.

Pour la continuité à gauche le raisonnement est similaire. ■

Jensen. *Soit X une v.a. sur $(\Omega, \mathcal{B}, \mathbb{P})$ telle que X prenne ses valeurs dans $]a, b[$. Soit $\phi :]a, b[\to \mathbb{R}$ convexe. Supposons que $X \in \mathcal{L}^1(\mathbb{P})$ et que $\phi \circ X \in \mathcal{L}^1(\mathbb{P})$. Alors*

$$\phi(\mathbb{E}(\mathbf{X})) \leq \mathbb{E}(\phi(\mathbf{X})).$$

Preuve. Soit $t = \mathbb{E}(X)$. Comme $a < X < b$, il vient $a = \mathbb{E}(\mathbf{a}) \leq \mathbb{E}(X) \leq \mathbb{E}(\mathbf{b}) = b$. Montrons qu'en fait $a < t < b$. L'inégalité stricte est automatique si $a = -\infty$ ou $b = +\infty$ puisque par hypothèse $X \in \mathcal{L}^1(\mathbb{P})$.

Supposons $a \in \mathbb{R}$ et $t = a = \mathbb{E}(X)$. Alors $X - a \geq 0$ intégrable d'intégrale nulle et donc $\mathbb{P}$-p.s., $X = a$, ce qui contredit $X \in]a, b[$. Idem en b pour le cas $b \in \mathbb{R}$.

Soient $a < s < t < u$, et notons $\beta = \sup_{s:a<s<t} \frac{\phi(t)-\phi(s)}{t-s}$. Alors d'après (Q),

$$t < u < b \Rightarrow \beta \leq \frac{\phi(u) - \phi(t)}{u - t},$$

et donc

$$a < s < b \Rightarrow \beta(s - t) + \phi(t) \leq \phi(s),$$

ce qui reste vrai si $s = X...$. En intégrant (tout est dans $\mathcal{L}^1(\mathbb{P})$), il vient

$$\beta\mathbb{E}(X - \mathbb{E}(X)) + \phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)),$$

et $\mathbb{E}(X - \mathbb{E}(X)) = 0$. ■

Voici une première application de l'inégalité de Jensen :

Lyapunov. *si* $0 < s < t$, *et* $X \in \mathcal{L}^t(\mathbb{P})$, *alors* $X \in \mathcal{L}^s(\mathbb{P})$ *et*

$$|\mathbf{X}|_\mathbf{s} \leq |\mathbf{X}|_\mathbf{t}.$$

Preuve. Poser $r = t/s > 1$. Donc $\phi(x) = |x|^r$ est convexe sur $\mathbb{R}$. Poser $Y = |X|^s$: alors $Y \in \mathcal{L}^r(\mathbb{P})$, donc dans $\mathcal{L}^1(\mathbb{P})$, et $\phi(Y) = |X|^t \in \mathcal{L}^1(\mathbb{P})$. On applique l'inégalité de Jensen pour déduire $\phi(\mathbb{E}(Y)) = (|X|_s^s)^r \leq \mathbb{E}(\phi(Y)) = |X|_t^t$, puis on passe à la puissance $1/t$. ■

On peut obtenir d'autres inégalités par le processus général suivant :

Une classe d'inégalités. *Soit* $\psi : \mathbb{R} \to \mathbb{R}^+$ *borélienne et pour* $A \in \mathcal{B}(\mathbb{R})$, *posons*

$$\mathbf{i_A} = \inf\{\psi(\mathbf{x}) : \mathbf{x} \in \mathbf{A}\}.$$

Alors si X *est une v.a. sur* $(\Omega, \mathcal{B}, \mathbb{P})$,

$$\mathbb{P}(\mathbf{X} \in \mathbf{A}) \leq \frac{\mathbf{1}}{\mathbf{i_A}}\mathbb{E}(\psi(\mathbf{X})).$$

Preuve. Ecrire

$$\begin{aligned} i_A\mathbb{P}(X \in A) &= \mathbb{E}(i_A\mathbf{1}_A(X)) = \mathbb{E}_{\mathbb{P}_X}(i_A\mathbf{1}_A(x)) \\ &\leq \mathbb{E}_{\mathbb{P}_X}(\psi(x)\mathbf{1}_A(x)) \leq \mathbb{E}_{\mathbb{P}_X}(\psi(x)) = \mathbb{E}(\psi(X)), \end{aligned}$$

où aucune condition d'intégrabilité n'est à vérifier puisque ψ est borélienne positive. ■

Passons aux applications :
Bienaymé-Chebyshev : on prend $\psi(x) = x^2$, et $X \in \mathcal{L}^2(\mathbb{P})$. On pose $Y = X - \mathbb{E}(X) \in \mathcal{L}^2(\mathbb{P})$, puis pour $\alpha > 0$, $A = [\alpha, +\infty[$. Alors $Var(X) = Var(Y) = \mathbb{E}(\psi(Y))$ et on obtient

$$\mathbb{P}(|\mathbf{X} - \mathbb{E}(\mathbf{X})| \geq \alpha) \leq \frac{\mathbf{Var(X)}}{\alpha^{\mathbf{2}}}.$$

Cette inégalité se généralise en la suivante :
Markov : on prend $r \geq 1$ et $X \in \mathcal{L}^r(\mathbb{P})$. Puis $\alpha > 0$ et $A = [\alpha, +\infty[$. Enfin on prend $\psi(x) = |x|^r$. Alors

$$\mathbb{P}(|\mathbf{X}| \geq \alpha) \leq \frac{\mathbb{E}(|\mathbf{X}|^{\mathbf{r}})}{\alpha^{\mathbf{r}}}.$$

Un autre corollaire est :
Cantelli : on suppose $X \in \mathcal{L}^2(\mathbb{P})$, avec $\mathbb{E}(X) = 0$. On pose $\sigma^{\mathbf{2}} = \mathbf{Var(X)}$, et on prend $a > 0$, et $\psi(x) = (x + \frac{\sigma^2}{a})^2$. Enfin on prend $A = [a, +\infty[$. Alors on a

$$\mathbb{P}(\mathbf{X} \geq \mathbf{a}) \leq \frac{\sigma^{\mathbf{2}}}{\sigma^{\mathbf{2}} + \mathbf{a^2}}.$$

EXERCICES

I : calculer, en les justifiant, les limites suivantes :

$$\lim_{n\to\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n e^{x/2} dx, \qquad \lim_{n\to\infty} \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-2x} dx.$$

II : soit μ une mesure positive sur $(X, \mathcal{B})$, $f : X \to \mathbb{C}$ mesurable, et $p > 0$. On pose

$$\phi(p) = \int_X |f|^p d\mu = \|f\|_p^p.$$

On pose ensuite $E = \{p : \phi(p) < \infty\}$, et on suppose que $\|f\|_\infty > 0$.

a) Pour $r < p < s$, avec $r, s \in E$, montrer que $p \in E$.

b) Montrer que $\log \phi$ est convexe à l'intérieur de E et que ϕ est continue sur E.

c) D'après a), E est connexe. Est-il nécessairement ouvert, fermé ? Peut-il être réduit à un point ? Peut-il être n'importe quel sous ensemble connexe de $]0, +\infty[$?

d) Pour $r < p < s$, montrer que $\|f\|_p \leq \max\{\|f\|_r, \|f\|_s\}$. En conclure $L^p(\mu) \subset L^s(\mu) \cap L^r(\mu)$.

e) Supposons qu'il existe $r > 0$ tel que $\|f\|_r < \infty$. Montrer qu'alors $\|f\|_p \to \|f\|_\infty$.

III : on reprend les hypothèses du **II** mais en outre on suppose que $\mu(X) = 1$.

a) Démontrer que $\|f\|_r \leq \|f\|_s$ si $0 < r < s \leq \infty$.

b) A quelles conditions peut-on avoir $0 < r < s \leq \infty$ et $\|f\|_r = \|f\|_s < \infty$?

c) En supposant que $\|f\|_r < \infty$ pour un certain $r > 0$, montrer que

$$\lim_{p \to 0^+} \|f\|_p = \exp\left(\int_X \log|f| d\mu\right),$$

en posant $\exp(-\infty) = 0$.

IV : on suppose toujours $\mu(X) = 1$. Soient f et g deux fonctions mesurables et positives telles que $fg \geq 1$. Montrer qu'alors

$$\int_X f d\mu \int_X g d\mu \geq 1.$$

V : toujours si $\mu(X) = 1$, soit $h > 0$ mesurable et $A = \int_X h d\mu$. Montrer que

$$\sqrt{1 + A^2} \leq \mathbb{E}_\mu(\sqrt{1 + h^2}) \leq 1 + A.$$

Si μ est la mesure de Lebesgue sur $[0, 1]$ et si h est continue, avec $h = f'$, les inégalités ci-dessus ont une interprétation géométrique simple. Laquelle ?

VI : soit $1 < p < \infty$, $f \in L^p(]0, +\infty[)$, relativement a la mesure de Lebesgue. Posons pour $x > 0$,

$$F(x) = \frac{1}{x} \int_0^x f(t) dt.$$

a) Démontrer l'inégalité de Hardy

$$\|F\|_p \leq \frac{p}{p-1} \|f\|_p,$$

qui assure que F envoie L^p dans lui-même.

b) Démontrer que l'on a l'égalité si et seulement si $f = 0$ p.s..

c) Démontrer que la constante $\frac{p}{p-1}$ est la plus petite possible.

d) Si $f > 0$ et $f \in L^1$, montrer que $F \notin L^1$.

<u>*Indications :*</u> pour le a), considérer d'abord $f \geq 0$ continue à support compact, et intégrer par parties pour obtenir $\|F\|_p^p = -p \int_0^\infty F^{p-1}(x) x F'(x) dx$. Remarquer que $xF' = f - F$, et appliquer l'inégalité de Hölder à $\int F^{p-1} f dx$. Pour c), faire $f(x) = x^{-1/p}$ sur $[1, A]$, 0 ailleurs, pour A assez grand.

VII : "*Théorème d'Egoroff*" ; supposons $\mu(X) < \infty$. Soit (f_n) une suite de fonctions complexes mesurables qui converge simplement sur X. Alors quelque soit $\varepsilon > 0$, il existe un sous ensemble mesurable E de X tel que $\mu(X \setminus E) < \varepsilon$ et que (f_n) converge uniformément sur E.

<u>*Indication :*</u> poser

$$S(n,k) = \cap_{i,j>n} \left\{ x : |f_i(x) - f_j(x)| < \frac{1}{k} \right\},$$

et montrer qu'il existe une suite croissante d'entiers (n_k) tel que $E = \cap_k S(n_k, k)$ ait la propriété désirée.

VIII : soit (X, Y) un couple de variables aléatoires réelles. On note m_2 la mesure de Lebesgue sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, et m la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On suppose que $\mathbb{P}_{(X,Y)} \ll m_2$, et on note

$$\frac{d\mathbb{P}_{(X,Y)}}{dm_2}(x,y) = f(x,y)$$

sa densité.

VIII-1 : montrer que $\mathbb{P}_X \ll m$ et que $\mathbb{P}_Y \ll m$.

VIII-2 : exprimer $\frac{d\mathbb{P}_X}{dm}(x)$ sous la forme d'une intégrale faisant intervenir f. Faire de même en ce qui concerne $\frac{d\mathbb{P}_Y}{dm}(y)$.

VIII-3 : soit $A \in \mathcal{B}(\mathbb{R}^2)$. On note $A_y = \{x \in \mathbb{R} : (x,y) \in A\}$. On rappelle que m presque sûrement en y, $A_y \in \mathcal{B}(\mathbb{R})$, et que $y \mapsto m(A_y)$ est mesurable, et que $m_2(A) = \int m(A_y)dy$.

Soient μ et ν deux mesures de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, absolument continues relativement à la mesure de Lebesgue m. Montrer qu'alors $\mu \otimes \nu \ll m_2$, et déterminer $\frac{d(\mu\otimes\nu)}{dm_2}$ en fonction de $\frac{d\mu}{dm}$ et de $\frac{d\nu}{dm}$.

VIII-4 : en déduire une CNS pour que $X \perp\!\!\!\perp Y$, concernant la forme algébrique de f.

VIII-5 : on suppose que la loi d'un couple de v.a.r. (X, Y), $\mathbb{P}_{(X,Y)}$, a une densité $f(x,y) = \frac{1}{\pi(1+x^2)} \frac{1}{\pi(1+y^2)}$ relativement à m_2. Les variables X et Y sont-elles indépendantes ?

IX : soit (X, Y) un couple de variables aléatoires réelles tel que pour chaque $A \in \mathcal{B}(\mathbb{R}^2)$,

$$\mathbb{P}((X,Y) \in A) = \sum_{n \geq 1} \frac{1}{2^{n+1}} \mathbf{1}_A(n,n) + C \int_A (1 - x^2y^2) \mathbf{1}_{[0,1]^2}(x,y) dm_2(x,y),$$

où C est une constante positive.

IX-1 : calculer C.

IX-2 : déterminer $\mathbb{P}_X$, puis $\mathbb{P}_Y$ sans calcul supplémentaire.

IX-3 : déterminer la décomposition de Lebesgue de $\mathbb{P}_X$ relativement à m.

IX-4 : X et Y sont-elles indépendantes ?

IX-5 : montrer que X est intégrable et calculer $\mathbb{E}(X)$.

X : soit $(A_n)_{n\geq 1}$ une suite d'évènements mesurables d'un espace probabilisé $(\Omega, \mathcal{B}, \mathbb{P})$, telle que $\mathbb{P}(A_1) > 0$. On note

$$N_n = \mathbf{1}_{A_1} + \ldots + \mathbf{1}_{A_n}, \text{ et } \theta_n = \frac{\sum_{j,k \leq n} \mathbb{P}(A_j \cap A_k)}{\left(\sum_{k \leq n} \mathbb{P}(A_k)\right)^2}, \ \ n \geq 1.$$

X-1 : montrer que $\limsup_n A_n = \{\lim_n \uparrow N_n = +\infty\}$.

X-2 : on note $p_k = \mathbb{P}(A_k)$ et $m_n = p_1 + \ldots + p_n$. Calculer $Var(N_n)$ en fonction de θ_n et de m_n.

X-3 : soit $x \in \mathbb{R}$ tel que $m_n > x$. Montrer que $\mathbb{P}(N_n \leq x) \leq \mathbb{P}(|N_n - m_n| \geq m_n - x)$, et en déduire que $\mathbb{P}(N_n \leq x) \leq \frac{(\theta_n - 1)m_n^2}{(m_n - x)^2}$.

X-4 : on suppose désormais que $\sum_n p_n = +\infty$, et que $\liminf_n \theta_n \leq 1$. Montrer qu'alors quel que soit $x \in \mathbb{R}$, $\liminf_n \mathbb{P}(N_n \leq x) = 0$.

X-5 : montrer que $\mathbb{P}(\sup_k N_k \leq x) \leq \mathbb{P}(N_n \leq x)$. En déduire que $\mathbb{P}(\limsup_n A_n) = 1$.

X-6 : déduire de ce qui précède que le second Lemme de Borel-Cantelli est encore valable si l'on suppose les A_i deux à deux indépendants.

Chapitre 5
Convergence presque sûre

5.1. Indépendance des v.a. - critères.

Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{B}, \mathbb{P})$. Nous appelons **tribu engendrée par** X**, notée** $\sigma(\mathbf{X})$, la tribu $X^{-1}(\mathcal{B}(\mathbb{R}))$.

Soit $(X_i : \Omega \to \mathbb{R})_{i\in I}$ une suite finie ou dénombrable de variables aléatoires. Soit $J \subset I$ fini.

Nous dirons que **la suite finie** $(X_j)_{j\in J}$ **est indépendante** si

$$(\mathbf{A_j})_{\mathbf{j}\in\mathbf{J}} \in \prod_{\mathbf{j}\in\mathbf{J}} \sigma(\mathbf{X_j}) \Longrightarrow \mathbb{P}(\cap_{\mathbf{j}\in\mathbf{J}}\mathbf{A_j}) = \prod_{\mathbf{j}\in\mathbf{J}} \mathbb{P}(\mathbf{A_j}).$$

Nous dirons que **la suite** $(X_i)_{i\in I}$ **est indépendante** si quelque soit $J \subset I$ fini, la suite $(X_j)_{j\in J}$ est indépendante.

Nous dirons que **la suite** $(X_i)_{i\in I}$ **est deux à deux indépendante** si quels que soient $i \neq j \in I$, X_i est indépendante de X_j (**noté encore** $\mathbf{X_i} \perp\!\!\!\perp \mathbf{X_j}$).

Théorème 1. *La suite* $(X_i)_{i\in I}$ *est indépendante si et seulement si quel que soit* $J \subset I$ *fini, si* $\mathcal{X}_J = (X_j)_{j\in J}$*, alors*

$$\mathbb{P}_{\mathcal{X}_\mathbf{j}} = \bigotimes_{\mathbf{j}\in\mathbf{J}} \mathbb{P}_{\mathbf{X_j}}.$$

Preuve. Soit $(A_j)_{j\in J} \in \mathcal{B}(\mathbb{R})^J$. La loi $\mathbb{P}_{\mathcal{X}_J}$ est entièrement déterminée par ses valeurs du type $\mathbb{P}_{\mathcal{X}_J}(\prod_{j\in J} A_j)$. Il en est de même de l'autre loi en question sur $\mathbb{R}^J$, à savoir $\bigotimes_{j\in J} \mathbb{P}_{X_j}$. Mais par indépendance,

$$\begin{array}{rcl} \mathbb{P}_{\mathcal{X}_J}(\prod_{j\in J} A_j) & = & \mathbb{P}(\cap_{j\in J} X_j^{-1}(A_j)) \\ & = & \prod_{j\in J} \mathbb{P}(X_j^{-1}(A_j)) \\ & = & \prod_{j\in J} \mathbb{P}_{X_j}(A_j) \\ & = & \left(\bigotimes_{j\in J} \mathbb{P}_{X_j}\right)(\prod_{j\in J} A_j). \end{array}$$

Donc l'indépendance entraîne que la loi est la loi produit (π-système générateur des pavés).

Réciproquement, si $\mathbb{P}_{\mathcal{X}_J} = \bigotimes_{j\in J} \mathbb{P}_{X_j}$, l'indépendance est évidente. ∎

Théorème 2. *La suite $(X_i)_{i\in I}$ est indépendante si et seulement si quel que soit $J \subset I$ fini, si $\mathcal{X}_J = (X_j)_{j\in J}$, alors*

$$\mathbf{F}_{\mathcal{X}_\mathbf{J}}((\mathbf{t_j})_{\mathbf{j}\in\mathbf{J}}) = \prod_{\mathbf{j}\in\mathbf{J}} \mathbf{F}_{\mathbf{X_j}}(\mathbf{t_j}) \qquad (i.e.\ \mathbf{F}_{\mathcal{X}_\mathbf{J}} = \bigotimes_{\mathbf{j}\in\mathbf{J}} \mathbf{F}_{\mathbf{X_j}}).$$

Preuve. Soit $(t_j)_{j\in J} \in \mathbb{R}^J$. Si on a l'indépendance, alors

$$F_{\mathcal{X}_J}((t_j)_{j\in J}) = \mathbb{P}(\cap_{j\in J} X_j^{-1}(]-\infty, t_j])) = \prod_{j\in J} \mathbb{P}(X_j^{-1}(]-\infty, t_j])) = \prod_{j\in J} F_{X_j}(t_j).$$

Réciproquement, l'ensemble des pavés $\prod_{j\in J}]-\infty, t_j]$ forme un π-système générateur de $\mathcal{B}(\mathbb{R}^J)$. Si donc l'on suppose que $F_{\mathcal{X}_J} = \bigotimes_{j\in J} F_{X_j}$, c'est donc que $\mathbb{P}_{\mathcal{X}_J}$ et $\bigotimes_{j\in J} \mathbb{P}_{X_j}$ coïncident sur ce π-système générateur, et donc sont égales.

Par le Théorème 1 ci-dessus, on obtient l'indépendance. ■

Corollaire. *Si $X \perp\!\!\!\perp Y$, alors quels que soient $g, h : \mathbb{R} \to \mathbb{R}$ boréliennes telles que $g(X)$ et $h(Y)$ soient dans $\mathcal{L}^2(\mathbb{P})$, on a*

$$\mathbb{E}(\mathbf{g}(\mathbf{X})\mathbf{h}(\mathbf{Y})) = \mathbb{E}(\mathbf{g}(\mathbf{X}))\mathbb{E}(\mathbf{h}(\mathbf{Y})).$$

En particulier si $X, Y \in \mathcal{L}^2(\mathbb{P})$,

$$\mathbf{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}.$$

Et alors $\mathbf{Var}(\mathbf{X}+\mathbf{Y}) = \mathbf{Var}(\mathbf{X}) + \mathbf{Var}(\mathbf{Y})$.

Preuve. Il suffit de le vérifier d'abord pour les fonctions simples g, h, puis pour leurs limites croissantes boréliennes, et enfin de considérer les parties positives et négatives.

On a donc

$$Cov(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(X - \mathbb{E}(X))\mathbb{E}(Y - \mathbb{E}(Y)) = 0^2 = 0.$$

Enfin $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X, Y)$. ■

Pour conclure ce paragraphe, voici quelques exercices classiques :

I : soient X et Y deux var indépendantes et équidistribuées sur $\{0, \ldots, n\}$ et $\{0, \ldots, m\}$ respectivement.

Déterminer la loi de $Z = X + Y$: on appelle cette loi la loi *trapèzoïdale*.

II : soient X et Y deux var à valeurs entières dont la loi conjointe (à savoir **la loi du couple ou vecteur** (X,Y)) est donnée par

$$\mathbb{P}((X,Y)=(x,y)) = \frac{C}{(x+y-1)(x+y)(x+y+1)}, \qquad x,y \geq 1.$$

Déterminer les lois de $U = X+Y$ et $V = X-Y$.

III : soient X et Y deux var indépendantes de lois de Poisson de paramètres respectifs λ et β. Montrer que $\mathcal{L}(X+Y) = \mathcal{P}(\lambda+\beta)$.

IV : si X est une var de loi géométrique, à valeurs dans $\mathbb{N}$, montrer que quel que soit $k,n \geq 1$,

$$\mathbb{P}(X = n+k | X > n) = \mathbb{P}(X = k).$$

Pourquoi appelle-t-on cette propriété la propriété d'"*absence de mémoire*" ?

Existe-t-il une autre v.a. à valeurs dans $\mathbb{N}$ ayant cette propriété ?

V : on note $\{p_1 < p_2 < \dots\} = \mathcal{P}$ l'ensemble des nombres premiers, et $(N_i)_{i\geq 1}$ une suite de var indépendantes telles que

$$\mathbb{P}(N_i = k) = (1-\gamma_i)\gamma_i^k, \qquad k \geq 0,$$

où $\gamma_i = p_i^{-\beta}$, pour chaque i, et un certain $\beta > 1$. Montrer qu'alors

$$M = \prod_{i\geq 1} p_i^{N_i}$$

est un entier aléatoire, de loi $\mathbb{P}(M = m) = Cm^{-\beta}$ pour $m \geq 1$ (c'est la distribution dite de "*Dirichlet*"), où C est une constante telle que

$$C = \prod_{i\geq 1}\left(1 - \frac{1}{p_i^{\beta}}\right) = \left(\sum_{m\geq 1}\frac{1}{m^{\beta}}\right)^{-1}.$$

5.2. Loi forte d'Etemadi.

Commençons par un petit Lemme, qui a son propre intérêt au demeurant :

Lemme. *Si $p \geq 1$ et $X \in \mathcal{L}^p(\mathbb{P})$, $X : \Omega \to \mathbb{R}^+$, alors*

$$\mathbb{E}(\mathbf{X^p}) = \int_{\mathbf{0}}^{\infty} \mathbf{p x^{p-1}(1 - F_X(t))dt}.$$

Preuve. On a $1-F_X(t) = \mathbb{P}(X > t) = \int_0^\infty \mathbf{1}_{]t,+\infty[}d\mathbb{P}_X(u)$, et donc par Fubini-Tonelli, et le théorème de transfert,

$$\begin{aligned} \int_0^\infty px^{p-1}(1-F_X(t))dt &= \int_{(\mathbb{R}^+)^2} pt^{p-1}\mathbf{1}_{]t,+\infty[}(u)d\mathbb{P}_X(u)dt \\ &= \int_0^\infty \left(\int_0^u pt^{p-1}dt\right) d\mathbb{P}_X(u) = \int_0^\infty u^p d\mathbb{P}_X(u) \\ &= \mathbb{E}(X^p). \qquad \blacksquare \end{aligned}$$

La loi forte "*classique*" suppose l'indépendance de la suite et non pas simplement l'indépendance deux à deux. Le résultat qui suit est dû au mathématicien indien Etemadi, et date de vingt ans. Il généralise le cas classique.

Loi forte d'Etemadi. *Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires identiquement distribuées, intégrables, et deux à deux indépendantes. Alors*

$$\lim_{\mathbf{n}} \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{k}\leq \mathbf{n}} \mathbf{X_k} = \mathbb{E}(\mathbf{X_1})\ \mathbb{P}-\mathbf{p.s.}.$$

Preuve. D'abord remarquons que sous nos hypothèses, les mêmes hypothèses ont lieu pour les deux suites $(X_k^+)_{k\geq 1}$ et $(X_k^-)_{k\geq 1}$. Et donc il est clair que si le résultat est vrai sous l'hypothèse supplémentaire de positivité des X_k, il en découlera dans le cas général par décomposition en parties positives et négatives. Ainsi nous pouvons sans perte de généralité nous restreindre au cas $\mathbb{P}(X_1 \geq 0) = 1$. Nous noterons $\mathbf{S_n} = \sum_{\mathbf{k}\leq \mathbf{n}} \mathbf{X_k}$.

Posons $\mathbf{Y_k} = \mathbf{X_k} \mathbf{1}_{\mathbf{X_k}\leq \mathbf{k}}$, et $\mathbf{S^*_n} = \sum_{\mathbf{k}\leq \mathbf{n}} \mathbf{Y_k}$. Remarquons que la suite (Y_k) est une suite deux à deux indépendante, et que $Y_k \in \mathcal{L}^\infty(\mathbb{P})$.

Soit $\alpha > \mathbf{1}$, fixé, et posons $\mathbf{u_n} = [\alpha^{\mathbf{n}}]$. Nous allons d'abord démontrer que, si $\varepsilon > 0$ est donné,

$$\Sigma_\varepsilon = \sum_{\mathbf{n}} \mathbb{P}\left[\left|\frac{\mathbf{S^*_{u_n}} - \mathbb{E}(\mathbf{S^*_{u_n}})}{\mathbf{u_n}}\right| > \varepsilon\right] < \infty.$$

Par indépendance et distribution identique des X_k, on a

$$Var(S_n^*) = \sum_{k\leq n} Var(Y_k) \leq \sum_{k\leq n} \mathbb{E}(Y_k^2) = \sum_{k\leq n} \mathbb{E}(X_1^2 \mathbf{1}_{X_1\leq k}) \leq n\mathbb{E}(X_1^2 \mathbf{1}_{X_1\leq n}).$$

Par application de l'inégalité de Bienaymé-Chebyshev, il s'ensuit que

$$\Sigma_\varepsilon \leq \frac{1}{\varepsilon^2} \sum_n \frac{1}{u_n} \mathbb{E}(X_1^2 \mathbf{1}_{X_1\leq u_n}).$$

Posons $\mathbf{K} = \mathbf{2}\alpha \,/\, (\alpha - \mathbf{1})$, et soit $\mathbf{x} > 0$. Soit $N = N(x) = \min\{n : u_n \geq x\}$: alors $\alpha^N \geq x$, et comme pour $y \geq 1$, on a toujours $y \leq 2[y]$, il s'ensuit que

$$\sum_{u_n\geq x} \frac{1}{u_n} \leq 2 \sum_{n\geq N} \alpha^{-n} = K\alpha^{-N} \leq \frac{K}{x}.$$

Et donc en substituant X_1 à x, il vient, compte tenu de la majoration précédente de Σ_ε, et par convergence monotone,

$$\Sigma_\varepsilon \leq \frac{\mathbf{1}}{\varepsilon^{\mathbf{2}}} \mathbb{E}(\mathbf{X_1^2} \mathbf{1}_{\mathbf{X_1}>\mathbf{0}} (\sum_{\mathbf{u_n}\geq \mathbf{X_1}} \frac{\mathbf{1}}{\mathbf{u_n}})) \leq \frac{\mathbf{1}}{\varepsilon^{\mathbf{2}}} \mathbb{E}(\mathbf{X_1^2} \mathbf{1}_{\mathbf{X_1}>\mathbf{0}} \frac{\mathbf{K}}{\mathbf{X_1}}) = \frac{\mathbf{K}}{\varepsilon^{\mathbf{2}}} \mathbb{E}(\mathbf{X_1}) < \infty.$$

La convergence de Σ_ε entraîne, par le premier lemme de Borel-Cantelli, en prenant une suite $\varepsilon_p \downarrow 0$, que

$$\mathbb{P}(\limsup_{\mathbf{n}} \left|\frac{\mathbf{S^*_{u_n}} - \mathbb{E}(\mathbf{S^*_{u_n}})}{\mathbf{u_n}}\right| > \mathbf{0}) = \mathbf{0},$$

autrement dit que $\mathbb{P}$-p.s., la suite $(S^*_{u_n} - \mathbb{E}(S^*_{u_n})) \,/\, u_n$ tend vers 0.

D'autre part, $\mathbb{E}(Y_k) \to \mathbb{E}(X_1)$, par convergence monotone. Et donc puisque $u_n \to \infty$, il s'ensuit que $\mathbb{E}(S^*_{u_n}) \,/\, u_n \to \mathbb{E}(X_1)$, et donc nous concluons que

$$\frac{\mathbf{S^*_{u_n}}}{\mathbf{u_n}} \to \mathbb{E}(\mathbf{X_1}) \qquad \mathbb{P}-\mathbf{p.s.}.$$

Ensuite on a, en utilisant le lemme préliminaire,

$$\sum_{\mathbf{k}} \mathbb{P}(\mathbf{X_k} \neq \mathbf{Y_k}) = \sum_{\mathbf{k}} \mathbb{P}(\mathbf{X_k} > \mathbf{k}) \leq \int_{\mathbf{0}}^{\infty} \mathbb{P}(\mathbf{X_1} > \mathbf{t})\mathbf{dt} = \mathbb{E}(\mathbf{X_1}) < \infty,$$

et donc à nouveau avec le premier lemme de Borel-Cantelli, il s'ensuit que

$$\mathbb{P}(\limsup_{\mathbf{k}} \{\mathbf{X_k} \neq \mathbf{Y_k}\}) = \mathbf{0}.$$

Et donc $\mathbb{P}$-p.s., il existe $n(\omega)$ tel que si $n \geq n(\omega)$, $(S_n - S^*_n)(\omega) = S_{n(\omega)}(\omega) - S^*_{n(\omega)}(\omega)$, et donc

$$\lim_{\mathbf{n}} \frac{\mathbf{S_n} - \mathbf{S^*_n}}{\mathbf{n}} = \mathbf{0} \qquad \mathbb{P}-\mathbf{p.s.}.$$

Il s'ensuit que

$$\lim_{\mathbf{n}} \frac{\mathbf{S_{u_n}}}{\mathbf{u_n}} = \mathbb{E}(\mathbf{X_1}) \qquad \mathbb{P}-\mathbf{p.s.}. \tag{C1}$$

Soit $n_0 = \min\{n : \frac{1}{u_n} < \alpha - 1\}$: alors la suite $(u_n)_{n \geq n_0}$ est strictement croissante. Soit $k \geq u_{n_0}$ et soit n l'unique entier tel que $u_n \leq k < u_{n+1}$. Par positivité des X_k, nous avons

$$\frac{u_n}{u_{n+1}} \frac{S_{u_n}}{u_n} \leq \frac{S_k}{k} \leq \frac{u_{n+1}}{u_n} \frac{S_{u_{n+1}}}{u_{n+1}},$$

dont il découle, puisque $u_{n+1} \,/\, u_n \to \alpha$, et compte tenu de $(C1)$, que

$$\textbf{quel que soit } \alpha > \mathbf{1},\ \mathbb{P}\left(\frac{\mathbf{1}}{\alpha}\mathbb{E}(\mathbf{X_1}) \leq \liminf_{\mathbf{k}} \frac{\mathbf{S_k}}{\mathbf{k}} \leq \limsup_{\mathbf{k}} \frac{\mathbf{S_k}}{\mathbf{k}} \leq \alpha\mathbb{E}(\mathbf{X_1})\right) = \mathbf{1}.$$

Il suffit alors d'intersecter les ensembles correspondants de mesures pleines le long d'une suite $\alpha_n \downarrow 1$ pour conclure. ■

Corollaire (ou réciproque partielle). *Supposons $X_1^- \in \mathcal{L}^1(\mathbb{P})$ et $\mathbb{E}(X_1^+) = +\infty$. Alors sous les mêmes hypothèses que celles de la loi forte, on a*

$$\frac{\mathbf{S_n}}{\mathbf{n}} \to +\infty \qquad \mathbb{P}-\mathbf{p.s.}.$$

Preuve. Par la loi forte, si nous posons $S_{n,-} = \sum_{k\leq n} X_k^-$, alors $S_{n,-} \,/\, n \to \mathbb{E}(X_1^-)$ $\mathbb{P}$-p.s.. Donc il suffit de montrer que le corollaire est vrai si $X_1 \geq 0$ et $\mathbb{E}(X_1) = +\infty$. Soit alors $p \in \mathbb{N}$ et posons $X_{k,p} = \min(X_1, p)$. Alors par la loi forte, si $S_{n,p} = \sum_{k\leq n} X_{k,p}$, on a $\lim_n S_{n,p}/n = \mathbb{E}(X_{1,p})$, et puisque $S_{n,p} \leq S_n$, on a

$$\mathbb{E}(\mathbf{X_{1,p}}) \leq \liminf_{\mathbf{n}} \frac{\mathbf{S_n}}{\mathbf{n}}.$$

Comme $X_{1,p} \uparrow X_1$ si $p \to \infty$, il s'ensuit que $+\infty = \lim_{\mathbf{p}} \uparrow \mathbb{E}(\mathbf{X_{1,p}}) \leq \liminf_{\mathbf{n}} \frac{\mathbf{S_n}}{\mathbf{n}}$. ■

5.3. Inégalités maximales.

Inégalité maximale de Kolmogorov. *Soit $(X_i)_{i\geq 1}$ une suite de v.a.r. indépendante. Supposons les X_i dans $\mathcal{L}^2(\mathbb{P})$, et centrées ($\mathbb{E}(X_i) = 0$). Alors, si $\alpha > 0$, et $S_n = \sum_{k\leq n} X_k$,*

$$\mathbb{P}(\max_{\mathbf{1\leq k\leq n}} |\mathbf{S_k}| \geq \alpha) \leq \frac{\mathbf{1}}{\alpha^{\mathbf{2}}}\mathbf{Var}(\mathbf{S_n}).$$

Preuve. Posons

$$A_k(\alpha) = A_k = \{\max_{j<k} |S_j| < \alpha\} \cap \{|S_k| \geq \alpha\}.$$

Posons

$$\Omega_n = \{\max_{1\leq k\leq n} |S_k| \geq \alpha\}.$$

Alors Ω_n est l'union disjointe des A_k, $1 \leq k \leq n$. D'autre part, les X_k étant centrées, les S_n le sont aussi, et donc $Var(S_n) = \mathbb{E}(S_n^2)$. Il vient

$$Var(S_n) \geq \sum_{k\leq n} \int_{A_k} S_n^2 d\mathbb{P}.$$

On a $S_n = (S_n - S_k) + S_k$, et donc

$$Var(S_n) \geq \sum_{k\leq n} \int_{A_k} S_n^2 d\mathbb{P} \geq \sum_{k\leq n} \int_{A_k} (S_k^2 + 2S_k(S_n - S_k))d\mathbb{P},$$

puisque $S_k^2 \geq 0$.

Ensuite $S_n - S_k = f(X_{k+1}, \ldots, X_n)$ avec $f(x_{k+1}, \ldots, x_n) = x_{k+1} + \ldots + x_n$, et $S_k = g(X_1, \ldots, X_k)$ de façon analogue. De plus, $\mathbf{1}_{A_k} = h(X_1, \ldots, X_k)$ pour une application borélienne $h : \mathbb{R}^k \to [0, 1]$. Par indépendance, on a

$$\mathbb{P}_{(X_1,\ldots,X_n)} = \mathbb{P}_{(X_1,\ldots,X_k)} \otimes \mathbb{P}_{(X_{k+1},\ldots,X_n)},$$

et donc f et gh étant boréliennes, il s'ensuit que $(S_n - S_k) \perp\!\!\!\perp S_k$. Dans $\mathcal{L}^2(\mathbb{P})$, cela impose

$$\int_{A_k} (S_n - S_k) S_k d\mathbb{P} = \mathbb{E}((S_n - S_k)(S_k \mathbf{1}_{A_k})) = \mathbb{E}(S_n - S_k)\mathbb{E}(S_k \mathbf{1}_{A_k}) = 0,$$

puisque $S_n - S_k$ est centrée. Et donc

$$\mathbf{Var}(\mathbf{S_n}) \geq \sum_{\mathbf{k \leq n}} \int_{\mathbf{A_k}} \mathbf{S_k^2 d}\mathbb{P} \geq \sum_{\mathbf{k \leq n}} \int_{\mathbf{A_k}} \alpha^{\mathbf{2}} \mathbf{d}\mathbb{P} = \alpha^{\mathbf{2}} \mathbb{P}(\Omega_{\mathbf{n}}).$$

■

Inégalité maximale d'Etemadi. *Sous les mêmes hypothèses, on a*

$$\mathbb{P}(\max_{\mathbf{1 \leq k \leq n}} |\mathbf{S_k}| \geq \mathbf{3}\alpha) \leq \mathbf{3} \max_{\mathbf{1 \leq k \leq n}} \mathbb{P}(|\mathbf{S_k}| \geq \alpha).$$

Preuve. Reprenons les notations de la preuve précédente, et notons $A_k = A_k(3\alpha)$ et $\Omega_n = \cup_{k \leq n} A_k$ l'union disjointe correspondante. Alors

$$\mathbb{P}(\Omega_n) = \mathbb{P}(\Omega_n \cap \{|S_n| \geq \alpha\}) + \mathbb{P}(\Omega_n \cap \{|S_n| < \alpha\}) \leq \mathbb{P}(|S_n| \geq \alpha) + \sum_{k \leq n-1} \mathbb{P}(A_k \cap \{|S_n| < \alpha\}),$$

puisque $A_n \cap \{|S_n| < \alpha\} = \emptyset$.

Puisque $S_n = (S_n - S_k) + S_k$, pour que $|S_k| \geq 3\alpha$, sachant que $|S_n| < \alpha$, il est nécessaire que $|S_n - S_k| \geq 2\alpha$, et donc si $k \leq n-1$, $A_k \cap \{|S_n| < \alpha\} \subset A_k \cap \{|S_n - S_k| \geq 2\alpha\}$, et donc

$$\mathbb{P}(\Omega_{\mathbf{n}}) \leq \mathbb{P}(|\mathbf{S_n}| \geq \alpha) + \sum_{\mathbf{k \leq n-1}} \mathbb{P}(\mathbf{A_k} \cap \{|\mathbf{S_n} - \mathbf{S_k}| \geq \mathbf{2}\alpha\}).$$

A nouveau par indépendance, on a $\mathbf{1}_{A_k} \perp\!\!\!\perp \mathbf{1}_{|S_n - S_k| \geq 2\alpha}$, et donc

$$\mathbb{P}(A_k \cap \{|S_n - S_k| \geq 2\alpha\}) = \mathbb{P}(A_k)\mathbb{P}(|S_n - S_k| \geq 2\alpha).$$

Donc, puisque $\sum_{k \leq n-1} \mathbb{P}(A_k) \leq 1$, il vient

$$\mathbb{P}(\Omega_{\mathbf{n}}) \leq \mathbb{P}(|\mathbf{S_n}| \geq \alpha) + \max_{\mathbf{1 \leq k \leq n}} \mathbb{P}(|\mathbf{S_n} - \mathbf{S_k}| \geq \mathbf{2}\alpha).$$

Ensuite pour que $|S_n - S_k| \geq 2\alpha$, il faut que ou bien $|S_n| \geq \alpha$, ou bien $|S_k| \geq \alpha$, et donc

$$\mathbb{P}(\Omega_n) \leq \mathbb{P}(|S_n| \geq \alpha) + \max_{k \leq n-1} \mathbb{P}(|S_n| \geq \alpha) + \mathbb{P}(|S_k| \geq \alpha)),$$

et la conclusion s'en suit. ■

Application : convergence p.s. de séries de v.a. indépendantes. *Soit (X_n) une suite indépendante de v.a. satisfaisant aux mêmes hypothèses que les précédentes. Notons $S_n = \sum_{1\le k\le n} X_k$. Soit S une v.a.. Alors*

$$\mathbf{(S_n)\ CVPS\ vers\ une\ v.a.\ S} \Leftrightarrow \forall\varepsilon > \mathbf{0},\ \ \mathbb{P}(|\mathbf{S_n} - \mathbf{S}| \ge \varepsilon) \underset{\mathbf{n}\to\infty}{\longrightarrow} \mathbf{0}$$

(on dit alors que S_n converge vers S en probabilité).

Preuve.
$\underline{\Rightarrow}$: si S_n CVPS vers S, alors pour $\varepsilon > 0$, $\mathbb{P}(\limsup\{|S_n - S| \ge \varepsilon\}) = 0$. Mais $\limsup\{|S_n - S| \ge \varepsilon\} = \cap_n \downarrow \cup_{m\ge n}\{|S_m - S| \ge \varepsilon\}$, et donc par monotonie de $\mathbb{P}$, il vient

$$\begin{aligned} 0 \le \liminf_n \mathbb{P}(\{|S_n - S| \ge \varepsilon\}) &\le \limsup_n \mathbb{P}(\{|S_n - S| \ge \varepsilon\}) \\ &\le \limsup_n \mathbb{P}(\cup_{m\ge n}\{|S_m - S| \ge \varepsilon\}) = 0. \end{aligned}$$

$\underline{\Leftarrow}$: on va montrer que $\mathbb{P}$-p.s., (S_n) est Cauchy. Pour cela observons d'abord que

$$\mathbb{P}(|S_{n+j} - S_n| \ge \varepsilon) \le \mathbb{P}(|S_{n+j} - S| \ge \frac{\varepsilon}{2}) + \mathbb{P}(|S_n - S| \ge \frac{\varepsilon}{2}).$$

L'hypothèse entraîne donc que

$$\lim_n \sup_{j\ge 1} \mathbb{P}(|S_{n+j} - S_n| \ge \varepsilon) = 0.$$

Par l'inégalité maximale d'Etemadi, on a

$$\mathbb{P}(\max_{1\le j\le k} |S_{n+j} - S_n| \ge \varepsilon) \le 3 \max_{1\le j\le k} \mathbb{P}(|S_{n+j} - S_n| \ge \frac{\varepsilon}{3}),$$

dont il découle que

$$\mathbb{P}(\sup_{k\ge 1} |S_{n+k} - S_n| > \varepsilon) \le 3 \sup_{k\ge 1} \mathbb{P}(|S_{n+k} - S_n| \ge \frac{\varepsilon}{3}).$$

Donc $\mathbb{P}(\sup_{\mathbf{k}\ge \mathbf{1}} |\mathbf{S_{n+k}} - \mathbf{S_n}| > \varepsilon) \to_{\mathbf{n}} \mathbf{0}$. D'autre part on a l'inclusion

$$\{\sup_{j,k\ge n} |S_j - S_k| > \varepsilon\} \subset \{\sup_{k\ge 1} |S_{n+k} - S_n| > \frac{\varepsilon}{2}\},$$

et donc il s'ensuit que la condition de Cauchy pour $\varepsilon > 0$ est satisfaite avec probabilité 1, quel que soit $\varepsilon > 0$. ■

Les deux inégalités maximales précédentes majorent la probabilité pour que le maximum dépasse quelquechose. La suivante la minore :

Inégalité de Kolmogorov à gauche. *Soit $c > 0$ et supposons que (Z_n) soit une suite indépendante de v.a., centrées, et uniformément bornées par c ($|Z_n| \leq c$). Soit $\varepsilon > 0$ et notons $S_k = Z_1 + \ldots + Z_k$. Alors*

$$\mathbf{1} - \frac{(\varepsilon + \mathbf{c})^{\mathbf{2}}}{\sum_{\mathbf{k} \leq \mathbf{n}} \mathbf{Var}(\mathbf{Z_k})} \leq \mathbb{P}(\max_{\mathbf{k} \leq \mathbf{n}} |\mathbf{S_k}| \geq \varepsilon).$$

Preuve. Toujours avec les mêmes notations, en notant $B_k = \{\max_{j \leq k} |S_j| < \varepsilon\}$, avec $A_k = A_k(\varepsilon)$, on a $B_k \cap A_k = \emptyset$ et $B_k = B_{k+1} \cup A_{k+1}$. On en déduit que

$$\mathbf{S_{k-1} 1_{B_{k-1}}} + \mathbf{Z_k 1_{B_{k-1}}} = \mathbf{S_k 1_{B_{k-1}}} = \mathbf{S_k 1_{B_k}} + \mathbf{S_k 1_{A_k}}. \tag{$\spadesuit$}$$

En élevant tout au carré, en prenant l'espérance, en tenant compte du fait que les variables sont centrées, avec par indépendance $Z_k \perp\!\!\!\perp \mathbf{1}_{B_{k-1}}$ et $Z_k \perp\!\!\!\perp S_{k_1} \mathbf{1}_{B_{k-1}}$, et du fait que $\mathbf{1}_{A_k} \mathbf{1}_{B_k} = 0$, il découle de ($\spadesuit$) que

$$\mathbb{E}(\mathbf{S^2_{k-1} 1_{B_{k-1}}}) + \mathbb{P}(\mathbf{B_{k-1}}) \mathbf{Var}(\mathbf{Z_k}) = \mathbb{E}(\mathbf{S^2_k 1_{B_k}}) + \mathbb{E}(\mathbf{S^2_k 1_{A_k}}). \tag{$\diamondsuit$}$$

De la définition de A_k et du fait que $|Z_k| \leq c$, il découle de ($\spadesuit$) que

$$|\mathbf{S_k 1_{A_k}}| \leq |\mathbf{S_{k-1} 1_{A_k}}| + |\mathbf{Z_k}| \mathbf{1_{A_k}} \leq (\varepsilon + \mathbf{c}) \mathbf{1_{A_k}}. \tag{$\heartsuit$}$$

De ($\diamondsuit$) et ($\heartsuit$), il découle que, puisque $B_n \subset B_{k-1}$,

$$\mathbb{E}(S^2_{k-1} \mathbf{1}_{B_{k-1}}) + \mathbb{P}(B_n) Var(Z_k) \leq \mathbb{E}(S^2_k \mathbf{1}_{B_k}) + (\varepsilon + c)^2 \mathbb{P}(A_k),$$

dont il suit en sommant sur $1 \leq k \leq n$, et puisque $\sum_{k \leq n} \mathbb{P}(A_k) = \mathbb{P}(\Omega_n)$ et $\mathbb{P}(B_n) = 1 - \mathbb{P}(\Omega_n)$, que

$$(\mathbf{1} - \mathbb{P}(\Omega_\mathbf{n}))(\sum_{\mathbf{1} \leq \mathbf{k} \leq \mathbf{n}} \mathbf{Var}(\mathbf{Z_k})) \leq (\varepsilon + \mathbf{c})^{\mathbf{2}} \mathbb{P}(\Omega_\mathbf{n}) + \varepsilon^{\mathbf{2}} \mathbb{P}(\Omega_\mathbf{n}). \tag{$\clubsuit$}$$

Notons $\Sigma = \sum_{1 \leq k \leq n} Var(Z_k)$ et $p = \mathbb{P}(\Omega_n)$. On veut $\Sigma - (\varepsilon + c)^2 \leq p\Sigma$. Pour l'instant on a, d'après ($\clubsuit$), que

$$\begin{array}{cl} & (1-p)\Sigma \leq (\varepsilon + c)^2 p + \varepsilon^2 (1-p) \\ \Leftrightarrow & \Sigma - \varepsilon^2 \leq p((\varepsilon + c)^2 + \Sigma - \varepsilon^2) \\ \underset{(0 \leq p \leq 1)}{\Rightarrow} & \Sigma - \varepsilon^2 \leq p\Sigma + (\varepsilon + c)^2 - \varepsilon^2 \\ \Leftrightarrow & \Sigma - (\varepsilon + c)^2 \leq p\Sigma. \end{array}$$

■

5.4. Convergence de séries aléatoires.

Si la série des variances converge, la série CVPS. *Soit* $(X_i)_{i\geq 1}$ *une suite de v.a. indépendante. Supposons les centrées et dans* $\mathcal{L}^2(\mathbb{P})$*. Alors*

$$\sum_i Var(X_i) < \infty \Longrightarrow \sum_i X_i \textbf{ converge } \mathbb{P}\textbf{-p.s..} \tag{C2}$$

Preuve. On pose $S_n = X_1 + \ldots + X_n$, et il s'agit de montrer que (S_n) est $\mathbb{P}$-p.s. de Cauchy.

Soit $\varepsilon > 0$: alors par l'inégalité maximale de Kolmogorov, si $m, n \geq 1$, on a

$$\mathbb{P}(\max_{1\leq k\leq n} |S_{n+k} - S_n| > \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{1\leq k\leq n} Var(X_{m+k}),$$

et par monotonie de $\mathbb{P}$, il s'ensuit que

$$\mathbb{P}(\sup_{\mathbf{k\geq 1}} |\mathbf{S_{m+k}} - \mathbf{S_m}| > \varepsilon) \leq \frac{\mathbf{1}}{\varepsilon^{\mathbf{2}}} \sum_{\mathbf{k\geq 1}} \mathbf{Var}(\mathbf{X_{m+k}}).$$

Comme $\sum_t Var(X_t) < \infty$, la série queue tend vers 0, et donc

$$\lim_{\mathbf{m}} \underbrace{\mathbb{P}(\sup_{k\geq 1} |S_{m+k} - S_m| > \varepsilon)}_{\mathbf{A}(\mathbf{m},\varepsilon)} = \mathbf{0},\ \forall \varepsilon > \mathbf{0}.$$

Notons $\mathbf{E}(\mathbf{n},\varepsilon) = \{\sup_{\mathbf{j,k\geq n}} |\mathbf{S_j} - \mathbf{S_k}| > \mathbf{2}\varepsilon\}$, et soit $\mathbf{E}(\varepsilon) = \cap_{\mathbf{n\geq 1}} \downarrow \mathbf{E}(\mathbf{n},\varepsilon)$. Alors

$$(\mathbf{S_n})\ \mathbb{P} - \mathbf{p.s.\ de\ Cauchy} \ \Leftrightarrow \forall \varepsilon > \mathbf{0},\ \mathbb{P}(\mathbf{E}(\varepsilon)) = \mathbf{0}.$$

Mais puisque si $j, k \geq m$ et $|S_j - S_k| = |(S_j - S_m) + (S_m - S_k)| > 2\varepsilon$ impose que $|S_j - S_m| > \varepsilon$ ou bien que $|S_k - S_m| > \varepsilon$, il s'ensuit que

$$E(m,\varepsilon) \subset A(m,\varepsilon).$$

Donc $0 \leq \mathbb{P}(E(\varepsilon) \leq \lim_m \downarrow \mathbb{P}(A(m,\varepsilon) = 0$. ■

Corollaire - Convergence des séries de v.a. uniformément bornées. *Soit* (Y_n) *une suite indépendante de v.a. telle qu'il existe* $c > 0$ *satisfaisant* $|Y_n| \leq c$ $\mathbb{P}$*-p.s., pour chaque* n*. Alors*

$$\sum_{\mathbf{n}} \mathbf{Y_n}\ \ \mathbf{CVPS} \Rightarrow \sum_{\mathbf{n}} \mathbb{E}(\mathbf{Y_n})\ \mathbf{CV}\ \ \mathbf{et}\ \sum_{\mathbf{n}} \mathbf{Var}(\mathbf{Y_n})\ \mathbf{CV}. \tag{C3}$$

PREUVE. Soit $(\Omega, \mathcal{B}, \mathbb{P})$ l'espace probabilisé sous-jacent et notons

$$(\tilde{\Omega}, \tilde{\mathcal{B}}, \tilde{\mathbb{P}}) = (\Omega \times \Omega, \mathcal{B} \otimes \mathcal{B}, \mathbb{P} \otimes \mathbb{P}), \quad \pi_1(\omega, \omega') = \omega, \quad \pi_2(\omega, \omega') = \omega'$$

l'espace produit et les projections associées. Notons $\tilde{Y}_n = Y_n \circ \pi_1$ et $\tilde{Y}'_n = Y_n \circ \pi_2$. Puisque $\tilde{\mathbb{P}}_{\pi_1} = \tilde{\mathbb{P}}_{\pi_2} = \mathbb{P}$, on a

$$\mathbb{P}_{\tilde{Y}_n} = \mathbb{P}_{\tilde{Y}'_n} = \mathbb{P}_{Y_n},$$

et donc $\mathbb{E}(Y_n) = \mathbb{E}(\tilde{Y}_n) = \mathbb{E}(\tilde{Y}'_n)$ et $Var(Y_n) = Var(\tilde{Y}_n) = Var(\tilde{Y}'_n)$, par le théorème de transfert. La suite de v.a. $(\tilde{Y}'_n)$ est une **copie indépendante** de la suite (Y_n) dans la mesure où par construction il découle directement des définitions que

$$\begin{cases} (a) & : \quad \tilde{\mathbb{P}}_{(\tilde{Y}'_1, \ldots, \tilde{Y}'_n)} = \tilde{\mathbb{P}}_{(\tilde{Y}_1, \ldots, \tilde{Y}_n)} = \mathbb{P}_{(Y_1, \ldots, Y_n)}, \ n \geq 1; \\ (b) & : \quad (\tilde{Y}_1, \ldots, \tilde{Y}_n) \perp\!\!\!\perp (\tilde{Y}'_1, \ldots, \tilde{Y}'_n), \ n \geq 1. \end{cases}$$

Observons aussi que, si $S_n = \tilde{Y}_1 + \ldots + \tilde{Y}_n$ et $S'_n = \tilde{Y}'_1 + \ldots + \tilde{Y}'_n$,

$$\begin{aligned} \tilde{\mathbb{P}}(\textstyle\sum_n \tilde{Y}'_n \ CV) &= \tilde{\mathbb{P}}(\{(\omega, \omega') : (S'_n(\omega, \omega')) \ Cauchy\}) \\ &= \mathbb{P}(\{\omega' \in \Omega : (S_n(\omega')) \ Cauchy\}) = 1, \end{aligned}$$

par hypothèse. Donc $\sum_n \tilde{Y}'_n$ CVPS, et de même $\sum_n \tilde{Y}_n$ aussi.

Formons la suite $Z_n = \tilde{Y}_n - \tilde{Y}'_n$. Alors la suite (Z_n) est une suite de v.a. indépendante, uniformément bornées par $2c$, centrées. En outre $Var(Z_n) = 2Var(Y_n)$.

Supposons que $\sum_n Var(Z_n) = +\infty$. Alors en appliquant l'inégalité maximale de Kolmogorov à gauche, on obtient, pour tout $m \geq 1$ et $\varepsilon > 0$,

$$\begin{aligned} &\mathbf{1} = \mathbf{1} - \lim_{\mathbf{n}} \frac{(\varepsilon + \mathbf{2c})^{\mathbf{2}}}{\sum_{\mathbf{k} \leq \mathbf{n}} \mathbf{Var}(\mathbf{Z}_{\mathbf{m}+\mathbf{k}})} \\ &\leq \liminf_{\mathbf{n}} \mathbb{P}(\max_{\mathbf{1} \leq \mathbf{k} \leq \mathbf{n}} |\mathbf{S}_{\mathbf{m}+\mathbf{k}} - \mathbf{S}_{\mathbf{m}}| \geq \varepsilon) \\ &= \mathbb{P}(\cup_{\mathbf{n}} \{\max_{\mathbf{1} \leq \mathbf{k} \leq \mathbf{n}} |\mathbf{S}_{\mathbf{m}+\mathbf{k}} - \mathbf{S}_{\mathbf{m}}| \geq \varepsilon\}), \end{aligned}$$

par monotonie de $\mathbb{P}$ et croissance des ensembles $\{\max_{1 \leq k \leq n} |S_{m+k} - S_m| \geq \varepsilon\}$ indexés par n.

Il s'en suit que quel que soit $\varepsilon > 0$, pour presque tout $\omega \in \Omega$, pour tout $m \geq 1$, il existe un $k \geq 1$ tel que

$$|S_{m+k}(\omega) - S_m(\omega)| \geq \varepsilon,$$

et donc sauf peut-être sur un ensemble négligeable, la suite (S_n) ne satisfait jamais la condition de Cauchy. Or $\sum_n Z_n$ CVPS comme différence de deux séries qui CVPS. Une contradiction.

Donc $\sum_n Var(Z_n) < \infty$, et donc $\sum_{\mathbf{n}} \mathbf{Var}(\mathbf{Y}_{\mathbf{n}}) < \infty$.

Mais $Var(Y_n) = Var(Y_n - \mathbb{E}(Y_n))$, et $(Y_n - \mathbb{E}(Y_n))$ est une suite indépendante de v.a., centrées, dont la série des variances converge. Par application de $(C2)$, il s'ensuit que $\sum_n (Y_n - \mathbb{E}(Y_n)$ CVPS. Et comme $\sum_n Y_n$ CVPS par hypothèse, il s'ensuit que $\sum_n \mathbb{E}(Y_n)$ CVPS, par différence. Mais ceci signifie, puisqu'il s'agit de constantes, que $\sum_n \mathbb{E}(Y_n)$

converge tout simplement. ■

Suivent quelques définitions et un lemme préliminaire. Soient (Y_n) et (Y'_n) deux suites de v.a.. Nous dirons qu'elles sont **queue équivalentes, noté $\mathbf{Y_n \overset{q}{\equiv} Y'_n}$**, si

$$\mathbb{P}(\limsup_{\mathbf{n}}\{\mathbf{Y_n \neq Y'_n}\}) = \mathbf{0}.$$

Nous dirons qu'elles sont **convergence équivalentes, noté $\mathbf{Y_n \overset{c}{\equiv} Y'_n}$**, si

$$\mathbb{P}\left((\sum_{\mathbf{n}} \mathbf{Y_n}\ \mathbf{CV\ et} \sum_{\mathbf{n}} \mathbf{Y'_n}\ \mathbf{CV}) \vee (\sum_{\mathbf{n}} \mathbf{Y_n}\ \mathbf{DV\ et} \sum_{\mathbf{n}} \mathbf{Y'_n}\ \mathbf{DV})\right) = \mathbf{1}.$$

Lemme d'équivalence. *Si $\sum_n \mathbb{P}(Y_n \neq Y'_n) < \infty$, alors $Y_n \overset{q}{\equiv} Y'_n$ et $Y_n \overset{c}{\equiv} Y'_n$.*

Preuve. Par le premier lemme de Borel-Cantelli, l'affirmation $Y_n \overset{q}{\equiv} Y'_n$ est immédiate.

De fait, il en résulte que $\mathbb{P}$-p.s. en ω, les deux suites $(Y_n(\omega))$ et $(Y'_n(\omega))$ ne diffèrent que par un nombre fini de termes. Et donc les séries associées doivent converger ou diverger simultanément. ■

Théorème des trois séries de Kolmogorov. *Soit $(X_i)_{i\geq 1}$ une suite indépendante de v.a.. Posons pour $c > 0$, $\mathbf{X_n^{(c)}} = \mathbf{X_n 1_{|X_n|\leq c}}$. Notons*

$$\mathbf{S_1(c)} = \sum_{\mathbf{n}} \mathbb{P}(|\mathbf{X_n}| > \mathbf{c}), \qquad \mathbf{S_2(c)} = \sum_{\mathbf{n}} \mathbb{E}(\mathbf{X_n^{(c)}}), \qquad \mathbf{S_3(c)} = \sum_{\mathbf{n}} \mathbf{var}(\mathbf{X_n^{(c)}}).$$

Alors les conditions suivantes sont équivalentes :

$$\left\{\begin{array}{lcl} (i) & : & \sum_{\mathbf{n}} \mathbf{X_n}\ \mathbf{CVPS}; \\ (ii) & : & \exists \mathbf{c} > \mathbf{0}, \max\{\mathbf{S_1(c)}, \mathbf{S_2(c)}, \mathbf{S_3(c)}\} < \infty; \\ (iii) & : & \forall \mathbf{c} > \mathbf{0}, \max\{\mathbf{S_1(c)}, \mathbf{S_2(c)}, \mathbf{S_3(c)}\} < \infty. \end{array}\right.$$

Preuve. Nous montrerons $(iii) \Rightarrow (ii) \Rightarrow (i) \Rightarrow (iii)$. La première de ces trois implications est triviale.
$\underline{\mathbf{(ii)} \Rightarrow \mathbf{(i)}}$: soit $c > 0$ tel que les trois séries convergent. Puisque $\{|X_n| > c\} = \{X_n \neq X_n^{(c)}\}$, il résulte du lemme d'équivalence et de la convergence de $S_1(c)$ que $\sum_n X_n \overset{c}{\equiv} \sum_n X_n^{(c)}$.

Montrons donc que $\sum_n X_n^{(c)}$ CVPS. La convergence de $S_3(c)$ nous indique, d'après $(C2)$, que $\sum_n (X_n^{(c)} - \mathbb{E}(X_n^{(c)}))$ CVPS, et $S_2(c)$ indique que $\sum_n \mathbb{E}(X_n^{(c)})$ CV. Donc $\sum_n X_n^{(c)}$ CVPS.

(i) $\Rightarrow$ **(iii)** : donnons nous $c > 0$.

Si $\sum_n X_n$ CVPS, alors $X_n \to 0$ $\mathbb{P}$-p.s., et donc $\mathbb{P}$-p.s., il n'existe qu'un nombre fini de n tels que $|X_n| > c$. Cela signifie que

$$\mathbb{P}(\limsup\{|\mathbf{X_n}| > \mathbf{c}\}) = \mathbf{0}.$$

Par indépendance, et le second lemme de Borel-Cantelli (appliqué par contraposition), il s'ensuit que $S_1(c) < \infty$ ($S_1(c) = \sum_n \mathbb{P}(X_n \neq X_n^{(c)}) < \infty$). Et donc par le lemme d'équivalence, $X_n \overset{c}{\equiv} X_n^{(c)}$.

Mais puisque $\sum_n X_n$ CVPS, il s'en suit que $\sum_n X_n^{(c)}$ CVPS aussi.

Alors $(X_n^{(c)})$ est indépendante et uniformément bornée par $c > 0$. Et sa série CVPS : par $(C3)$ il s'ensuit que $S_2(c)$ et $S_3(c)$ convergent. ■

5.5. Exercices.

I : "*Intégration de Monte-Carlo*".

Soit $(\Omega, \mathcal{B}, \mathbb{P}) = ([0,1], \mathcal{B}, m)$ (m=mesure de Lebesgue), et $f \in L^2$. Soit $(U_i)_{i\geq 1}$ une suite i.i.d., de loi commune la loi uniforme sur $[0,1]$. On pose

$$I_n = \frac{1}{n}(f(U_1) + \ldots + f(U_n)).$$

a) : montrer que $I_n \overset{\mathbb{P}-p.s.}{\underset{n\to\infty}{\longrightarrow}} \mathbb{E}_m(f)$;

b) : utiliser l'inégalité de Chebyshev pour estimer $\mathbb{P}(|I_n - I| > \frac{a}{\sqrt{n}})$.

II : "*Théorème de Bernstein*".

Soit $f \in \mathcal{C}([0,1])$, et soit $f_n(x) = \sum_{m=0}^n C_n^m x^m (1-x)^{n-m} f(m/n)$ le $n^{\text{ième}}$ polynôme de Bernstein associé à f.

a) : soit S_n la somme de n variables indépendantes de loi de Bernoulli de paramètre p, $0 < p < 1$. Montrer que $\mathbb{E}f(S_n/n) = f_n(p)$.

b) : soit $M = |f|_\infty$. Pour $\varepsilon > 0$, choisir $\delta > 0$ tel que $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$ (justifier).
Montrer qu'alors $|\mathbb{E}(f(S_n/n)) - f(p)| \leq \varepsilon + 2M\mathbb{P}(|S_n/n - p| > \delta)$.

c) : montrer avec Chebyshev que $\mathbb{P}(|S_n/n - p| > \delta) \leq \frac{1}{4n\delta^2}$.

d) : montrer que $|f_n(x) - f(x)|_\infty \to 0$ si $n \to \infty$.

III : "*Loi forte dans L^4*".

Soit $(X_i)_{i\geq 1}$ une suite i.i.d. dans L^4 telle que $\mu = \mathbb{E}(X_1)$. On veut montrer que $S_n/n \to \mu$ $\mathbb{P}$-p.s..

a) : montrer que l'on peut supposer sans perte de généralité que $\mu = 0$.
b) : développer $\mathbb{E}(S_n^4)$ pour en déduire que $\mathbb{E}(S_n^4) \leq Cn^2$.
c) : en déduire avec l'inégalité de Markov que $\mathbb{P}(|S_n| > n\varepsilon) \leq C/(n^2\varepsilon^4)$.
d) : montrer à l'aide du premier Lemme de Borel-Cantelli que $\mathbb{P}(\limsup\{|S_n| > \varepsilon\}) = 0$. Conclure.

IV : "*Un continuum de mesures mutuellement singulières sur* $[0,1]$".
Pour $x \in [0,1]$, on note $(\epsilon_n(x))_{n\geq 1} \in \{0,1\}^{\mathbb{N}^*}$ une suite telle que

$$\mathbf{x} = \sum_{\mathbf{n}\geq \mathbf{1}} \frac{\epsilon_{\mathbf{n}}(\mathbf{x})}{\mathbf{2^n}}.$$

a) : montrer que si x n'est pas rationnel de la forme $\frac{p}{2^m}$, alors la suite $(\epsilon_n(x))$ est unique. Montrer que sinon il en existe au plus deux.
b) : on note $\pi : \{0,1\}^{\mathbb{N}^*} \to [0,1]$ l'application définie par $\pi((\epsilon_n)_{n\geq 1}) = \sum_{n\geq 1} \frac{\epsilon_n}{2^n}$. Sur $\Omega := \{0,1\}^{\mathbb{N}^*}$, on choisit la tribu $\mathcal{B} = \bigotimes_{\mathbb{N}^*} \mathcal{P}(\{0,1\})$. Montrer qu'alors π est $(\mathcal{B} - \mathcal{B}(\mathbb{R}))$-mesurable.
c) : soit $p \in]0,1[$, soit $\mathbb{P}_p$ la probabilité sur $(\{0,1\}, \mathcal{P}(\{0,1\}))$ définie par $\mathbb{P}_p(\{1\}) = p$. Soit enfin $\tilde{\mathbb{P}}_p = \bigotimes_{\mathbb{N}^*} \mathbb{P}_p$ la probabilité produit correspondante sur $(\Omega, \mathcal{B})$. Montrer que $(\tilde{\mathbb{P}}_{\frac{1}{2}})_\pi = m$, la mesure de Lebesgue sur $[0,1]$.
d) : en utilisant la loi forte des grands nombres, une fois, cependant, son utilisation justifiée, montrer que $p \neq q \Rightarrow \tilde{\mathbb{P}}_p \perp \tilde{\mathbb{P}}_q$. Montrer qu'il en est de même de $(\tilde{\mathbb{P}}_p)_\pi$ et de $(\tilde{\mathbb{P}}_q)_\pi$. Ces mesures sont-elles discrètes ? Absolument continues ? *Indication* : considérer $X_k((\epsilon_n)_{n\geq 1}) = \epsilon_k$.

6.1. Intégrabilité uniforme (ou équi-intégrabilité).
Soit $(\Omega, \mathcal{B}, \mathbb{P})$ un espace de probabilité, et $f \in \mathcal{L}^1(\mathbb{P})$. Alors par convergence dominée,

$$\lim_{n\to\infty} \int_{\{|f|\geq n\}} |f| d\mathbb{P} = 0.$$

Une suite (f_k) dans $\mathcal{L}^1(\mathbb{P})$ est dite **uniformément intégrable** si

$$\lim_{\mathbf{n}\to\infty} \sup_{\mathbf{k}} \int_{\{|\mathbf{f_k}|\geq \mathbf{n}\}} |\mathbf{f_k}| \mathbf{d}\mathbb{P} = \mathbf{0}.$$

Remarquons que si (f_k) est uniformément intégrable, alors

$$\sup_{\mathbf{k}} \int |\mathbf{f_k}| \mathbf{d}\mathbb{P} < \infty.$$

Si $f, g \in \mathcal{L}^1(\mathbb{P})$ et si $h = \max\{|f|, |g|)$, alors

$$\int_{|f+g|\geq 2n} |f+g| d\mathbb{P} \leq 2 \int_{h\geq n} h d\mathbb{P} \leq 2 \left(\int_{|f|\geq n} |f| d\mathbb{P} + \int_{|g|\geq n} |g| d\mathbb{P} \right).$$

Et donc si (f_k) et (g_k) sont uniformément intégrables, il en est de même de $(f_k + g_k)$.

REMARQUE. *Si la suite (f_k) est telle qu'il existe g intégrable avec $|f_k| \leq |g|$ (donc si elle est dominée dans $\mathcal{L}^1(\mathbb{P})$), alors elle est équi-intégrable.*

Le résultat suivant généralise le théorème de convergence dominée de Lebesgue.

THÉORÈME 1. *Supposons (f_k) dans $\mathcal{L}^1(\mathbb{P})$. Supposons que $f_k \to f$ $\mathbb{P}$-p.s.. Alors*
a) *: si la suite (f_k) est uniformément intégrable, f est intégrable et $\mathbb{E}(f_k) \to \mathbb{E}(f)$;*
b) *: si $f_k \geq 0$, et f intégrable, et si $\mathbb{E}(f_k) \to \mathbb{E}(f)$, la suite (f_k) est uniformément intégrable.*

PREUVE. Si **a)**, alors par le Lemme de Fatou, $\mathbb{E}(|f|) \leq \sup_k \int |f_k| d\mathbb{P} < +\infty$, et donc $f \in \mathcal{L}^1(\mathbb{P})$. Notons

$$\mathbf{f_k^\alpha} = \mathbf{f_k 1_{|f_k|<\alpha}}, \ \mathbf{f^\alpha} = \mathbf{f 1_{|f|<\alpha}}.$$

Alors $f_k^\alpha \to_{k\to\infty} f^\alpha$ $\mathbb{P}$-p.s. si $\mathbb{P}(|f| = \alpha) = 0$ (remarquons que l'ensemble des α tels que $\mathbb{P}(f = \alpha) > 0$ est au plus dénombrable, puisque $\sum_{\mathbb{P}(f=\alpha)>0} \mathbb{P}(f = \alpha) \leq 1$), et par convergence dominée, $\mathbb{E}(f_k^\alpha) \to_k \mathbb{E}(f^\alpha)$.

D'autre part $\mathbb{E}(f_k - f_k^\alpha) = \int_{|f_k|\geq\alpha} f_k d\mathbb{P}$ et $\mathbb{E}(f - f^\alpha) = \int_{|f|\geq\alpha} f d\mathbb{P}$, donc

$$\limsup_k \left| \int f_k d\mathbb{P} - \int f d\mathbb{P} \right| \leq \sup_k \int_{|f_k|\geq\alpha} f_k d\mathbb{P} + \int_{|f|\geq\alpha} |f| d\mathbb{P} \to_\alpha 0.$$

Si **b)**, alors comme $\mathbb{E}(f_k^\alpha) \to \mathbb{E}(f^\alpha)$ si $\mathbb{P}(f = \alpha) = 0$, il s'ensuit que $\mathbb{E}(|f_k|\mathbf{1}_{|f_k|\geq\alpha}) \to \mathbb{E}(|f|\mathbf{1}_{|f|\geq\alpha})$, par positivité. Soit $\varepsilon > 0$. Il existe α_0 tel que $\mathbb{P}(f = \alpha_0) = 0$, et si $\alpha \geq \alpha_0$,

$$\mathbb{E}(|f|\mathbf{1}_{|f|\geq\alpha}) < \varepsilon.$$

Alors il existe k_0 tel que si $k \geq k_0$, et $\alpha \geq \alpha_0$, et $\mathbb{P}(f = \alpha) = 0$,

$$\mathbb{E}(f_k\mathbf{1}_{|f_k|\geq\alpha}) \leq \mathbb{E}(f_k\mathbf{1}_{|f_k|\geq\alpha_0}) < \varepsilon.$$

Ensuite $\lim_\alpha \max\{\mathbb{E}(|f_k|\mathbf{1}_{|f|\geq\alpha}),\ 1 \leq k < k_0\} = 0$. La conclusion s'en suit. ■

Voici une autre expression de l'équi-intégrabilité :

THÉORÈME 2. *Une suite (f_k) est équi-intégrable si et seulement si*

$$\begin{cases} (i) & : \quad \sup_\mathbf{k} \mathbb{E}(|\mathbf{f_k}|) < \infty; \\ (ii) & : \quad \lim_{\mathbb{P}(\mathbf{A})\to\mathbf{0}} \sup_\mathbf{k} \mathbb{E}(|\mathbf{f_k}|\mathbf{1_A}) = \mathbf{0}. \end{cases}$$

PREUVE.
$\underline{\Leftarrow}$: soit $\varepsilon > 0$ et $\delta > 0$ tel que $\mathbb{P}(A) < \delta \Rightarrow \sup_k \mathbb{E}(|f_k|\mathbf{1}_A) < \varepsilon$. Utilisons l'inégalité de Markov :

$$\sup_k \mathbb{P}(|f_k| \geq n) \leq \frac{1}{n} \sup_k \mathbb{E}(|f_k|) \underset{n\to\infty}{\longrightarrow} 0.$$

Donc pour $n \geq n_0$, pour tout k, $\mathbb{P}(|f_k| \geq n) < \delta$. Et alors pour un tel n, et pour tout k, $\mathbb{E}(|f_k|\mathbf{1}_{|f_k|\geq n}) < \varepsilon$.
$\underline{\Rightarrow}$: soit $A \in \mathcal{B}$: alors

$$\begin{array}{rcl} \mathbb{E}(|f_k|\mathbf{1}_A) & = & \mathbb{E}(|f_k|\mathbf{1}_{A\cap\{|f_k|\geq n\}}) + \mathbb{E}(|f_k|\mathbf{1}_{A\cap\{|f_k|<n\}}) \\ & \leq & \mathbb{E}(|f_k|\mathbf{1}_{|f_k|\geq n}) + n\mathbb{P}(A). \end{array}$$

Il s'ensuit que

$$\sup_k \mathbb{E}(|f_k|\mathbf{1}_A) \leq \sup_k \mathbb{E}(|f_k|\mathbf{1}_{|f_k|\geq n}) + n\mathbb{P}(A).$$

Il suffit alors de choisir n assez grand d'abord, pour que le second sup ne dépasse pas $\frac{\varepsilon}{2}$, puis une fois fixé, de prendre $\delta = \frac{\varepsilon}{2n}$. ■

Pour conclure voici un critère pratique d'équi-intégrabilité :

Equi-intégrabilité et espaces $\mathcal{L}^{1+\varepsilon}$. *Soit $\varepsilon > 0$. Supposons que la suite (X_n) soit dans $\mathcal{L}^{1+\varepsilon}$ et que $\sup_n \mathbb{E}(|X_n|^{1+\varepsilon}) < \infty$. Alors la suite est équi-intégrable.*

Preuve. On remarque que $\int_{|X_n|\geq\alpha} \alpha^\varepsilon |X_n| \leq \int |X_n|^{1+\varepsilon} = \mathbb{E}(|X_n|^{1+\varepsilon})$. Et donc

$$\sup_n \int_{|X_n|\geq\alpha} |X_n| \leq \frac{1}{\alpha^\varepsilon} \sup_n \mathbb{E}(|X_n|^{1+\varepsilon}) \to_{\alpha\to+\infty} 0.$$

■

6.2. Suites stationnaires de v.a. et systèmes dynamiques.

Soit $(\Omega, \mathcal{B}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n\geq 0}$ une suite de v.a. réelles. Nous dirons qu'elle est **stationnaire** si pour chaque $p \geq 1$, et $n \geq 0$,

$$\mathbb{P}_{(\mathbf{X_0},\dots,\mathbf{X_{p-1}})} = \mathbb{P}_{(\mathbf{X_n},\dots,\mathbf{X_{n+p-1}})}.$$

Une suite stationnaire (X_n) est dite **intégrable** si $X_0 \in \mathcal{L}^1(\mathbb{P})$. Puisque $\mathbb{P}_{X_0} = \mathbb{P}_{X_n}$, il s'ensuit que pour une telle suite,

$$\mathbb{E}(\mathbf{X_0}) = \mathbb{E}(\mathbf{X_n}),\ \mathbf{n} \geq \mathbf{0}.$$

Un **système dynamique** est un quadruplet $(\Omega, \mathcal{B}, \mathbb{P}, T)$ où $T : \Omega \to \Omega$ est mesurable et la condition suivante est satisfaite :

$$\mathbb{P}_{\mathbf{T}} = \mathbb{P}.$$

On dit alors que $\mathbb{P}$ est **T-invariante**. Si $f \in \mathcal{L}^1(\mathbb{P})$, alors le théorème de transfert impose que

$$\mathbf{f} \circ \mathbf{T} \in \mathcal{L}^{\mathbf{1}}(\mathbb{P}) \text{ et } \mathbb{E}(\mathbf{f} \circ \mathbf{T}) = \mathbb{E}(\mathbf{f}).$$

Une conséquence simple de la T-invariance de $\mathbb{P}$ est alors aussi que $(\mathbf{f} \circ \mathbf{T^n})_{\mathbf{n}\geq\mathbf{0}}$ est une suite stationnaire intégrable.

Exemple. *Posons $(\Omega, \mathcal{B}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), m)$ et notons $T\omega = 2\omega \mod 1$. On obtient un système dynamique $(\Omega, \mathcal{B}, \mathbb{P}, T)$ (vérifier que $m_T = m$), et si $f \in \mathcal{L}^1(m)$, $(f \circ T^n)_{n\geq 0}$ est une suite stationnaire.*

Réciproque. *Etant donnée $(\tilde{\Omega}, \tilde{\mathcal{B}}, \tilde{\mathbb{P}})$ et $(\tilde{X}_n)_{n\geq 0}$ une suite stationnaire, il existe un système dynamique $(\Omega, \mathcal{B}, \mathbb{P}, T)$ et une fonction mesurable $X : \Omega \to \mathbb{R}$ tels que*

$$\mathbb{P}_{(\tilde{\mathbf{X}}_\mathbf{0},\dots,\tilde{\mathbf{X}}_\mathbf{n})} = \mathbb{P}_{(\mathbf{X},\dots,\mathbf{X}\circ\mathbf{T^n})},\ \mathbf{n} \geq \mathbf{0}.$$

Preuve. Considérons la suite d'espaces probabilisés $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_{(\tilde{X}_0,\dots,\tilde{X}_{n-1})})$. Notons encore $\psi_n : \mathbb{R}^{n+1} \to \mathbb{R}^n$ et $\pi_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^n$ les projections canoniques. Alors les conditions du théorème d'extension de Kolmogorov sont satisfaites, et donc il existe une unique probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{B}) = (\mathbb{R}^{\mathbb{N}}, \mathcal{T})$ telle que

$$\mathbb{P}_{\pi_n} = \mathbb{P}_{(\tilde{X}_0, \ldots, \tilde{X}_{n-1})}, \ n \geq 1.$$

Soit $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ définie par

$$\mathbf{T}(\omega_{\mathbf{0}}, \omega_{\mathbf{1}}, \ldots) = (\omega_{\mathbf{1}}, \omega_{\mathbf{2}}, \ldots),$$

qu'on appelle le shift ou décalage à gauche.

Si P est un pavé de $\mathbb{R}^n$, alors $T^{-1}(\pi_n^{-1} P) = \pi_{n+1}^{-1}(\mathbb{R} \times P)$, et donc

$$\begin{array}{rcll}
\mathbb{P}(T^{-1}(\pi_n^{-1}P)) & = & \mathbb{P}(\pi_{n+1}^{-1}(\mathbb{R} \times P)) & \\
& = & \mathbb{P}_{(\tilde{X}_0, \ldots, \tilde{X}_n)}(\mathbb{R} \times P) & (\mathbb{P}_{\pi_{n+1}} = \mathbb{P}_{(\tilde{X}_0, \ldots, \tilde{X}_n)}) \\
& = & \mathbb{P}_{(\tilde{X}_1, \ldots, \tilde{X}_n)}(P) & (\text{car } \tilde{X}_0 \in \mathbb{R}) \\
& = & \mathbb{P}_{(\tilde{X}_0, \ldots, \tilde{X}_{n-1})}(P) & (\text{stationnarité}) \\
& = & \mathbb{P}(\pi_n^{-1}(P)). & (\mathbb{P}_{\pi_n} = \mathbb{P}_{(\tilde{X}_0, \ldots, \tilde{X}_{n-1})})
\end{array}$$

Puisque les $\pi_n^{-1}(P)$, P pavé de $\mathbb{R}^n$, $n \geq 1$, engendrent $\mathcal{T}$, il s'ensuit que $\mathbb{P}_{\mathbf{T}} = \mathbb{P}$.

Notons $X_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, $n \geq 0$, les v.a. définies par

$$\mathbf{X_0}(\omega_{\mathbf{0}}, \omega_{\mathbf{1}}, \ldots) = \omega_{\mathbf{0}}, \ \text{ et } \mathbf{X_n} = \mathbf{X_0} \circ \mathbf{T^n}, \ \mathbf{n} \geq \mathbf{0}.$$

C'est une suite stationnaire. Il nous reste à vérifier que

$$\mathbb{P}_{(\tilde{X}_0, \ldots, \tilde{X}_{n-1})} = \mathbb{P}_{(X_0, \ldots, X_{n-1})}$$

pour chaque $n \geq 1$. Soit $P = \prod_{i=0}^{n-1} A_i$ un pavé de $\mathbb{R}^n$. Alors

$$\mathbb{P}_{(\mathbf{X_0}, \ldots, \mathbf{X_{n-1}})}(\mathbf{P}) = \mathbb{P}(\cap_{\mathbf{i=0}}^{\mathbf{n-1}} \{\omega_{\mathbf{i}} \in \mathbf{A_i}\}) = \mathbb{P}(\pi_{\mathbf{n}}^{-\mathbf{1}}(\mathbf{P})) = \mathbb{P}_{(\tilde{\mathbf{X}}_{\mathbf{0}}, \ldots, \tilde{\mathbf{X}}_{\mathbf{n-1}})}(\mathbf{P}).$$

■

On appelle souvent l'espace $(\mathbb{R}^{\mathbb{N}}, \mathcal{T}, \mathbb{P})$ construit ci-dessus à partir de la suite stationnaire $(\tilde{X}_n)$ **l'espace des trajectoires du processus** $(\tilde{X}_n)$.

Par la suite, nous ne considérerons donc plus qu'un système dynamique $(\Omega, \mathcal{B}, \mathbb{P}, T)$ et une fonction mesurable $f : \Omega \to \mathbb{R}$, ce qui est plus commode.

6.3. Ergodicité.

Le système dynamique $(\Omega, \mathcal{B}, \mathbb{P}, T)$ est dit **ergodique** si pour tout $A \in \mathcal{B}$,

$$\mathbf{A} = \mathbf{T^{-1}A} \Rightarrow \mathbb{P}(\mathbf{A}) = \mathbf{0} \textbf{ ou } \mathbb{P}(\mathbf{A}) = \mathbf{1},$$

autrement dit les ensembles mesurables invariants ($A = T^{-1}A$) sont de mesure nulle ou pleine.

LEMME - FONCTIONS INVARIANTES ET ERGODICITÉ. *Supposons* $(\Omega, \mathcal{B}, \mathbb{P}, T)$ *ergodique, et soit* $X : \Omega \to \mathbb{R}$ *une variable aléatoire* T*-invariante, i.e.* $\mathbb{P}$*-p.s.,* $X = X \circ T$*. Alors* X *est* $\mathbb{P}$*-p.s. constante.*

PREUVE. Puisque $X = X \circ T$, si $a \in \mathbb{R}$, il s'ensuit que l'ensemble $\{X \leq a\}$ est invariant. Par ergodicité, nous obtenons donc que quel que soit $a \in \mathbb{R}$, $F_X(a) = \mathbb{P}(X \leq a) = 0$ ou 1.

Comme F_X ne prend que les valeurs 0 et 1, de par ses propriétés caractéristiques, il existe un unique a_0 tel que

$$\mathbf{a < a_0 \Rightarrow F_X(a) = 0} \text{ et } \mathbf{F_X(a_0) = 1}.$$

On en conclut que $X = a_0$ $\mathbb{P}$-p.s.. ■

6.4. Théorème ergodique (maximal).

L'histoire de ce théorème remonte au début du XX$^{\text{ième}}$ siècle. Il est dû à Birkhoff, qui a montré la convergence presque sûre à peu près au moment où Von Neumann le démontrait pour la convergence dans $\mathcal{L}^2$.

La preuve originale de Birkhoff a été améliorée, et dans les années 1980 de nouvelles preuves, utilisant des techniques de l'analyse dite "non standard" sont apparues, initiées par le travail de Kamae.

La preuve que nous en présentons est dûe à Petersen (1999), issue d'améliorations successives apportées à l'approche de Kamae par Ornstein, Katznelson, Weiss, Keane.

Le théorème ergodique s'accompagne toujours d'une inégalité maximale, préliminaire.

THÉORÈME ERGODIQUE MAXIMAL. *Soit* $(\Omega, \mathcal{B}, \mathbb{P}, T)$ *un système dynamique, et* $f \in \mathcal{L}^1(\mathbb{P})$*. Soit* $\lambda \in \mathcal{L}^1(\mathbb{P})$ T*-invariante. Notons*

$$\mathbf{A_k f = \frac{1}{k} \sum_{j<k} f \circ T^j,\ f_N^* = \max_{1 \leq k \leq N} A_k f,\ f^* = \sup_{N \geq 1} f_N^* = \sup_{k \geq 1} A_k f}.$$

Alors

$$\int_{\{\mathbf{f^* > \lambda}\}} (\mathbf{f} - \lambda) \mathbf{d}\mathbb{P} \geq \mathbf{0}.$$

PREUVE. Notons $E = \{f^* > \lambda\}$, et $E_N = \{f_N^* > \lambda\}$. Remarquons que puisque $\lambda = \lambda \circ T$,

$$\mathbf{A_k}\lambda = \lambda,\ \text{et donc } \mathbf{A_k f} - \lambda = \mathbf{A_k(f} - \lambda).$$

<u>*Supposons d'abord* $f \in \mathcal{L}^\infty(\mathbb{P})$</u> : soit $m \gg N$, et notons

$$\Sigma_{\mathbf{m}} = \sum_{\mathbf{j<m}} \left[(\mathbf{f} - \lambda)\mathbf{1_{E_N}}\right] \circ \mathbf{T^j}.$$

Soit $\omega \in \Omega$ et considérons la suite dans $\{0,1\}^m$:

$$s_m(\omega) := (\mathbf{1}_{E_N}(\omega), \mathbf{1}_{E_N}(T\omega), \dots, \mathbf{1}_{E_N}(T^{m-1}\omega)).$$

Nous pouvons à priori découper cette suite en tronçons successifs de “0” et de “1”. Nous la lisons de droite à gauche, comme d'habitude.

Si $s_m(\omega)$ ne comporte pas de “1”, alors $\Sigma_m(\omega) = 0$.

Sinon, soit j le premier indice tel que $T^j\omega \in E_N$. Par définition de E_N, il existe $1 \leq k(T^j\omega) \leq N$ tel que $\sum_{i<k(T^j\omega)}(f-\lambda)(T^iT^j\omega) > 0$.

Trois cas peuvent alors se présenter.

Le premier correspond au fait que

$$(\mathbf{s_m}(\omega)_\mathbf{j}, \dots, \mathbf{s_m}(\omega)_{\mathbf{j+k(T^j}\omega)\mathbf{-1}}) = (\mathbf{1}, \dots, \mathbf{1}).$$

Alors de la positivité de $\sum_{i<k(T^j\omega)}(f-\lambda)(T^iT^j\omega)$ nous concluons que

$$\Sigma_m(\omega) > \Sigma_{m-j-k(T^j\omega)}(T^{j+k(T^j\omega)}\omega).$$

Le second est celui où $j + k(T^j\omega) < m$ mais

$$(\mathbf{s_m}(\omega)_\mathbf{j}, \dots, \mathbf{s_m}(\omega)_{\mathbf{j+k(T^j}\omega)\mathbf{-1}}) = (\mathbf{1}, \dots, \mathbf{1}, \mathbf{0}, ?, \dots, ?),$$

où “?” désigne 0 ou 1. Si $0 \leq t < k(T^j\omega)$ est tel que $s_m(\omega)_{j+t} = 0$, alors $(f-\lambda)(T^{j+t}\omega) \leq 0$, et donc

$$\sum_{k<k(T^j\omega)} (f-\lambda)\mathbf{1}_{E_N}(T^kT^j\omega) \geq \sum_{k<k(T^j\omega)} (f-\lambda)(T^kT^j\omega) > 0,$$

dont il découle que

$$\Sigma_m(\omega) > \Sigma_{m-j-k(T^j\omega)}(T^{j+k(T^j\omega)}\omega).$$

Le troisième cas est celui où $\mathbf{j} + \mathbf{k}(\mathbf{T^j}\omega) \geq \mathbf{m}$: alors on s'arrête. Et puisque $|f-\lambda| \leq |f|_\infty + |\lambda|$, et $k(T^j\omega) \leq N$,

$$\Sigma_m(\omega) \geq -N(|f|_\infty + |\lambda|).$$

Sinon on continue, dans les deux premiers cas, en repartant de $\Sigma_{m-j-k(T^j\omega)}(T^{j+k(T^j\omega)}\omega)$, et ainsi de suite, jusqu'à ce que l'on rencontre le troisième cas, ce qui est obligatoire. Au bout du compte, par une suite finie de minorations, on obtient

$$\Sigma_\mathbf{m} \geq -\mathbf{N}(|\mathbf{f}|_\infty + |\lambda|).$$

Par T-invariance, on a $\mathbb{E}(\Sigma_m) = m\int_{E_N}(f-\lambda)d\mathbb{P}$, et donc en divisant par m et en intégrant aussi le terme de droite ci-dessus, on obtient

$$\int_{E_N}(f-\lambda)d\mathbb{P} \geq \frac{N}{m}(\|f\|_\infty + |\lambda|) \underset{m\to\infty}{\longrightarrow} O.$$

Et donc

$$\int_{\mathbf{E_N}}(\mathbf{f}-\lambda)\mathbf{d}\mathbb{P} \geq \mathbf{0}.$$

Ensuite, puisque $\cup_N \uparrow E_N = E$, en dominant par $|f|+|\lambda| \in \mathcal{L}^1(\mathbb{P})$, et en appliquant le théorème de convergence dominée à la suite $((f-\lambda)\mathbf{1}_{E_N})$, on obtient l'inégalité maximale pour $f \in \mathcal{L}^\infty(\mathbb{P})$.

Seconde étape : passage à $f \in \mathcal{L}^1(\mathbb{P})$: on va tronquer f en posant $\phi_s = f\mathbf{1}_{|f|\leq s}$ pour $s \geq 1$. On a $\phi_s \to f$ simplement et par convergence dominée on a $\mathbb{E}(\phi_s) \to \mathbb{E}(f)$. Notons

$$\mathbf{f_N^{*,s}} = \max_{\mathbf{1}\leq\mathbf{k}\leq\mathbf{N}} \mathbf{A_k}(\phi_\mathbf{s}),\ \text{et } \mathbf{E_N^s} = \{\mathbf{f_N^{*,s}} > \lambda\}.$$

Puisque $\{\phi_s \neq f\} \downarrow \emptyset$, il s'ensuit que

$$\mathbb{P}(E_N \Delta E_N^s) \to 0),\ \text{et } f_N^{*,s} \underset{\mathbb{P}-p.s.}{\overset{\mathcal{L}^1(\mathbb{P})}{\longrightarrow}} f_N^*,$$

et donc à nouveau par convergence dominée, et le résultat établi dans le cas $\mathcal{L}^\infty(\mathbb{P})$, on obtient

$$0 \leq \int_{\{f_N^{*,s}>\lambda\}}(\phi_s-\lambda)d\mathbb{P} \to \int_{E_N}(f-\lambda)d\mathbb{P}.$$

Il suffit pour conclure de faire $N \to \infty$ en utilisant encore une fois la convergence dominée. ■

6.5. Le théorème ergodique.

Théorème ergodique. *Soit $(\Omega, \mathcal{B}, \mathbb{P}, T)$ un système dynamique et $f \in \mathcal{L}^1(\mathbb{P})$. Alors*

$\mathbf{A_k f}$ *CVPS vers une v.a. intégrable de même espérance que f.*

Preuve. Notons

$$\bar{\mathbf{A}} = \bar{\mathbf{A}}(\mathbf{f}) = \limsup_{\mathbf{k}} \mathbf{A_k f},\ \text{et } \underline{\mathbf{A}} = \underline{\mathbf{A}}(\mathbf{f}) = \liminf_{\mathbf{k}} \mathbf{A_k f}.$$

Montrons que $\bar{A}$ et $\underline{A}$ sont T-invariantes : soit $\omega \in \Omega$ et soit $n_k(\omega) \to \infty$ telle que $\bar{A}(\omega) = \lim_k A_{n_k(\omega)}f(\omega)$. Remarquons que

$$A_n f(T\omega) = \frac{n+1}{n}(A_{n+1}f(\omega) - \frac{f(\omega)}{n}),$$

et donc que $\lim_k A_{n_k(\omega)} f(T\omega) \to \bar{A}(\omega)$. Puisque la lim sup est la plus grande valeur d'adhérence, il s'ensuit que $\bar{A}(T\omega) \geq \bar{A}(\omega)$. Réciproquement, on utilise le même argument, et le fait que

$$A_{n+1} f(\omega) = \frac{n}{n+1}(A_n f(T\omega) + \frac{f(\omega)}{n+1}).$$

Pour montrer que $\underline{A} = \underline{A} \circ T$, on procède de même.

Supposons tout d'abord que $f \geq 0$: soit $n \geq 1$ et posons $\lambda_n = \min\{n, \bar{A}\} - \frac{1}{n}$. Alors λ_n est T-invariante, intégrable, et $\lambda_n < \bar{A}(f)$. De plus,

$$\limsup_k A_k f = \lim_n \downarrow \sup_{k \geq n} A_k f \leq \sup_{k \geq 1} A_k(f) = f^*,$$

et donc $\{f^* > \lambda_n\} = \Omega$. Ainsi le théorème ergodique maximal s'applique et montre que

$$\mathbb{E}(\mathbf{f}) \geq \mathbb{E}(\lambda_{\mathbf{n}}) = \int_{\bar{\mathbf{A}} \leq \mathbf{n}} \bar{\mathbf{A}} \mathbf{d}\mathbb{P} - \frac{\mathbf{1}}{\mathbf{n}}.$$

Puisque $\frac{1}{n} \to 0$, et que $\mathbb{E}(\min\{\bar{A}, n\}) \uparrow \mathbb{E}(\bar{A})$, il s'ensuit que $\bar{A} \in \mathcal{L}^1(\mathbb{P})$ et que

$$\int \bar{\mathbf{A}} \mathbf{d}\mathbb{P} \leq \int \mathbf{f} \mathbf{d}\mathbb{P}.$$

Ensuite décomposons $f = f^+ - f^-$: alors comme $0 \leq \bar{A}(f)^+ \leq \bar{A}(f^+)$, $0 \leq \bar{A}(f)^- \leq \bar{A}(f^-)$, le cas $f \geq 0$ précédent montre que $\bar{A}(f)^+$ et $\bar{A}(f)^-$ sont intégrables. Donc $\bar{A} \in \mathcal{L}^1(\mathbb{P})$. De plus $\bar{A}$ est T-invariante, et le même argument que ci-dessus montre que $\Omega = \{f^* > \bar{A} - \varepsilon\}$, quel que soit $\varepsilon > 0$. En prenant $\lambda_\varepsilon = \bar{A} - \varepsilon$, qui est T-invariante et intégrable, on a

$$\int f d\mathbb{P} \geq \int \lambda_\varepsilon d\mathbb{P}.$$

En outre $\int \lambda_\varepsilon d\mathbb{P} \uparrow_{\varepsilon \to 0^+} \int \bar{A} d\mathbb{P}$, et donc

$$\mathbf{f} \in \mathcal{L}^{\mathbf{1}}(\mathbb{P}) \Rightarrow \int \mathbf{f} \mathbf{d}\mathbb{P} \geq \int \bar{\mathbf{A}} \mathbf{d}\mathbb{P}.$$

Appliquons cela à $-f$ pour obtenir, puisque $\bar{A}(-f) = -\underline{A}(f)$,

$$-\int \underline{A} d\mathbb{P} \leq -\int f d\mathbb{P}.$$

De ces deux inégalités il résulte que

$$\int \bar{A} \leq \int f \leq \int \underline{A},$$

et donc $\bar{A} - \underline{A} \geq 0$, est intégrable, et $\int(\bar{A} - \underline{A}) \leq 0$. Donc $\underline{A} = \bar{A}$ $\mathbb{P}$-p.s.. Et donc $(A_k f)$ converge presque sûrement vers une v.a. $\bar{A}$ intégrable.

Il reste à montrer que $\mathbb{E}(\bar{A}) = \mathbb{E}(f)$. Remarquons que $\mathbb{E}(A_k f) = \mathbb{E}(f)$, par T-invariance de $\mathbb{P}$. Nous allons montrer que la suite $(A_k f)$ est uniformément intégrable, ce qui montrera que $\mathbb{E}(\bar{A}) = \mathbb{E}(f)$.

Soit $\lambda \in \mathbb{R}$ et posons $E_\lambda = \{f^* > \lambda\}$. Alors Par le théorème ergodique maximal, on a

$$\lambda \mathbb{P}(E_\lambda) \leq \mathbb{E}(|f|).$$

En appliquant ceci à $-f$, et en posant $G_\lambda = \{\sup_k |A_k f| > \lambda\}$, il vient

$$\lambda \mathbb{P}(G_\lambda) \leq 2\mathbb{E}(|f|).$$

Donc pour λ et $\alpha > O$,

$$\begin{aligned}
\int_{|A_k f| > \lambda} |A_k f| d\mathbb{P} &\leq \tfrac{1}{k} \textstyle\sum_{j<k} \int_{G_\lambda} |f| \circ T^j d\mathbb{P} \\
&\leq \tfrac{1}{n} \textstyle\sum_{j<k} \left(\int_{|f|\circ T^j > \alpha} |f| \circ T^j d\mathbb{P} + \alpha \mathbb{P}(G_\lambda) \right) \\
&= \int_{|f| > \alpha} |f| d\mathbb{P} + \alpha \mathbb{P}(G_\lambda) \\
&\leq \int_{|f| > \alpha} + 2\tfrac{\alpha}{\lambda} \mathbb{E}(|f|).
\end{aligned}$$

Choisir $\alpha = \sqrt{\lambda}$ et faire $\lambda \to \infty$ pour conclure. ■

Le cas ergodique. *Si $\mathbb{P}$ est ergodique, et f est intégrable, alors*

$$A_k f \to \mathbb{E}(f)\ \mathbb{P} - p.s..$$

Preuve. Notons $\bar{f} = \lim_k A_k f = \bar{A}(f)$. Alors $\bar{f}$ est T-invariante, et par ergodicité, il s'ensuit que $\bar{f} = Cte$ $\mathbb{P}$-p.s.. Mais comme $\mathbb{E}(\bar{f}) = \mathbb{E}(f)$, la valeur de la constante ne peut être que $\mathbb{E}(f)$. ■

6.6. Le théorème ergodique entraîne la loi forte i.i.d..

La loi forte d'Etemadi est valable pour une suite indépendante $(\tilde{X}_n)_{n\geq 0}$ de v.a. identiquement distribuées intégrables.

Petit Lemme. *Une suite $(Z_n)_{n\geq 0}$ converge $\mathbb{P}$-p.s. vers 0 si et seulement si*

$$\forall \varepsilon > \mathbf{0},\ \mathbb{P}(\limsup_{\mathbf{n}} \{|\mathbf{Z_n}| \geq \varepsilon\}) = \mathbf{0}.$$

Preuve. Notons $A(\varepsilon) = \limsup_n \{|Z_n| \geq \varepsilon\}$.

$\underline{\Rightarrow}$: pour presque tout $\omega \in \Omega$, et tout $\varepsilon > 0$, il existe $n(\omega)$ tel que $n \geq n(\omega) \Rightarrow |Z_n(\omega)| < \varepsilon$, et donc pour presque tout ω, pour tout $\varepsilon > 0$, ω n'appartient qu'a un nombre au plus fini des ensembles $\{|Z_n| \geq \varepsilon\}$. Il s'ensuit que $A(\varepsilon)^c$ contient l'ensemble $\{Z_n \to 0\}$, et donc a lui-même une mesure 1.
$\underline{\Leftarrow}$: posons $B = \cap_{p\geq 1} A(\frac{1}{p})^c$. Alors $\mathbb{P}(B) = 1$. Soit $\omega \in B$, et $\varepsilon > 0$. Alors soit $p \geq 1$ tel que $\frac{1}{p} < \varepsilon$. Comme $\omega \in A(\frac{1}{p})^c$, il s'ensuit qu'au-delà d'un certain rang, disons $n(\omega)$, $|Z_n(\omega)| < \frac{1}{p} < \varepsilon$. Donc $B \subset \{Z_n \to 0\}$. ■

Théorème Ergodique $\Rightarrow$ Loi Forte $\mathcal{L}^1$ i.i.d.. *Comme le titre l'indique !*

Preuve. Plaçons nous sur l'espace des trajectoires du processus supposé i.i.d. intégrable $(\tilde{X}_n)_{n\geq 0}$, à l'origine défini sur l'espace $(\tilde{\Omega}, \tilde{\mathcal{B}}, \tilde{\mathbb{P}})$. La suite obtenue $(X \circ T^n)_{n\geq 0}$ est i.i.d. dans $\mathcal{L}^1(\mathbb{P})$ car pour tout $n \geq 1$,

$$\mathbb{P}_{(\mathbf{X},\dots,\mathbf{X}\circ\mathbf{T^n})} = \mathbb{P}_{\mathbf{X}}^{\otimes^{\mathbf{n}}}.$$

Et donc $\mathbb{P} = \mathbb{P}_X^{\otimes^{\mathbb{N}}}$. La loi $0-1$ de Kolmogorov affirme que

$$\mathcal{T}_\infty := \cap_{\mathbf{n}} \sigma(\mathbf{X} \circ \mathbf{T^n}, \mathbf{X} \circ \mathbf{T^{n+1}}, \dots) \textbf{ est triviale,}$$

c'est à dire qu'elle ne contient que des évènements certains ou impossibles.

Soit $A \in \mathcal{T}$ T-invariant : alors pour chaque n, $A = T^{-n}A \in \sigma(X \circ T^n, \dots)$, et donc $A \in \mathcal{T}_\infty$, et donc $\mathbb{P}(A) = 0$ ou 1. Donc T est ergodique.

Alors par le théorème ergodique,

$$\frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{k<N}} \mathbf{X} \circ \mathbf{T^k} \to \mathbb{E}(\mathbf{X}) = \mathbb{E}(\tilde{\mathbf{X}}_{\mathbf{0}})\ \mathbb{P} - \mathbf{p.s.}.$$

Notons $S_N = \frac{1}{N}\sum_{k<N} X \circ T^k$ et $\tilde{S}_N = \frac{1}{N}\sum_{k<N} \tilde{X}_k$.

Nous voulons montrer que $\tilde{\mathbb{P}}(\tilde{S}_N - \mathbb{E}(\tilde{X}_0) \to 0) = 1$. Via notre petit Lemme, montrons que pour tout $\varepsilon > 0$,

$$\tilde{\mathbb{P}}(\limsup_n \{|\tilde{S}_n - \mathbb{E}(\tilde{X}_0)| \geq \varepsilon\}) = 0,$$

ce qui suffira. Notons $\tilde{A}(\varepsilon) = \limsup_n \{|\tilde{S}_n - \mathbb{E}(\tilde{X}_0)| \geq \varepsilon\}$. Alors

$$\tilde{A}(\varepsilon)^c = \cup_n \uparrow \underbrace{(\cap_{p\geq n} \downarrow \underbrace{\{\max_{n\leq t\leq p} \{|\tilde{S}_t - \mathbb{E}(\tilde{X}_0)|\} < \varepsilon\})}_{\tilde{C}(n,p,\varepsilon)}}_{\tilde{B}(n,\varepsilon)}.$$

Nous introduisons, de façon parfaitement similaire, les ensembles $A(\varepsilon)^c$, $B(n,\varepsilon)$, et $C(n,p,\varepsilon)$, correspondant à la suite $(X \circ T^n)_{n\geq 0}$ sur notre espace de trajectoires (enlever les "$\tilde{}$" partout).

Notons, pour $n \geq 1, p \geq n$,

$$\begin{array}{cccc} m_{n,p} & : & \mathbb{R}^p & \longrightarrow & \mathbb{R} \\ & & (x_i)_{i<p} & \mapsto & \max_{n \leq t \leq p} |\frac{1}{t}\sum_{j<t} x_j - \mathbb{E}(\tilde{X}_0)| \end{array}$$

Alors $m_{n,p}$ est borélienne, et par construction, si $\varepsilon > 0$,

$$\tilde{\mathbb{P}}(\tilde{C}(n,p,\varepsilon)) = \mathbb{P}_{(\tilde{X}_0,\dots,\tilde{X}_{p-1})}(m_{n,p} < \varepsilon) = \mathbb{P}_{(X,\dots,X\circ T^{p-1})}(m_{n,p} < \varepsilon) = \mathbb{P}(C(n,p,\varepsilon))$$

(les lois $\mathbb{P}_{(\tilde{X}_0,\dots,\tilde{X}_{p-1})}$ et $\mathbb{P}_{(X,\dots,X\circ T^{p-1})}$ sont identiques). Et donc

$$\begin{array}{rcl} \tilde{\mathbb{P}}(\tilde{A}(\varepsilon)^c) & = & \lim_n \uparrow (\lim_p \downarrow \tilde{\mathbb{P}}(\tilde{C}(n,p,\varepsilon))) \\ & = & \lim_n \uparrow (\lim_p \downarrow \mathbb{P}(C(n,p,\varepsilon))) \\ & = & \lim_n \uparrow B(\varepsilon, n) \\ & = & \mathbb{P}(A(\varepsilon)^c) = 1 \text{ !!!!} \end{array}$$

$\blacksquare$

Exercice : on note $(\epsilon_i(x))_{i\geq 1}$ la suite des chiffres de $\{0,1,\dots,9\}$ du nombre $x \in [0,1[$ qui s'écrit en base 10 : $x = 0,\epsilon_1(x)\epsilon_2(x)\dots$.

Est-ce que, si $Tx = 10x \mod 1$,

$$\lim_N \frac{1}{N}\#\{1 \leq i \leq N : \epsilon_i(x)\epsilon_{i+1}(x)\epsilon_{i+2}(x)\epsilon_{i+3}(x) = 1789\} \textit{ converge } \mathbb{P}-p.s. \text{ ?}$$

Est-ce que la loi des grands nombres permet d'obtenir ce résultat ?

On dit qu'un nombre x est **normal en base** 10 si quelle que soit la suite finie $W = W_0W_1\dots W_{w-1} \in \{0,\dots,9\}^w$,

$$\lim_N \frac{1}{N}\#\{1 \leq i \leq N : \epsilon_i(x)\dots\epsilon_{i+w-1}(x) = W\} = \frac{1}{10^w}.$$

Montrer que m presque tout x de $[0,1[$ est normal en base 10.

S'ils le sont presque tous, le problème parfois peut-être d'en trouver un précisément et qui le soit. Montrer que le nombre suivant (dû à Champernown) l'est :

$$x_0 = 0,0\,1\,2\,3\,4\,5\,6\,7\,8\,9\,10\,11\,12\,13\,14\,15\,16\,17\,18\,19\,20\dots.$$

Exercice : reprenons l'espace des lancés d'une pièce d'équilibrage $0 < p < 1$, à savoir $(\Omega,\mathcal{B},\mathbb{P}) = (\{0,1\}^{\mathbb{N}},\mathcal{T},\mathbb{P})$, et $\mathbb{P}^{(p)} := \mathbb{P}_p^{\otimes\mathbb{N}}$. Soit A l'évènement défini par

$$A : \text{“les quatre premiers lancés donnent 0001”}.$$

Montrer que $\mathbb{P}$-p.s., si $\omega = (\omega_i)_{i\geq 0} \in \{0,1\}^{\mathbb{N}}$, la fréquence d'apparitions de l'évènement A dans ω est égale à $p(1-p)^3$.

Exercice : on partitionne l'intervalle $[0,1]$ en n intervalles disjoints, de longueurs $p_1, \ldots, p_n$. On appelle *entropie de la partition* le nombre h défini par

$$h = -\sum_{i=1}^{n} p_i \log p_i.$$

Soit $(X_i)_{i\geq 1}$ une suite i.i.d. de loi commune la loi uniforme sur $[0,1]$, et soit $Z_m(i)$ le nombre des v.a.r. $X_1, \ldots, X_m$ qui sont dans le $i^{\text{ième}}$ intervalle de la partition. Montrer que

$$R_m = \prod_{i=1}^{m} p_i^{Z_m(i)}$$

satisfait à $m^{-1} \log R_m \to -h$ $\mathbb{P}$-p.s..

Exercice : mettons que les catastrophes se produisent aux temps $T_1, \ldots, T_n, \ldots$, où $T_i = X_1 + \ldots + X_n$, et la suite $(X_i)_{i\geq 1}$ est une suite i.i.d. de var positives.

Soit $N(t) = \max\{n : T_n \leq t\}$, le nombre de catastrophes déjà produites au temps t. Montrer que si $\mathbb{E}(X_1) < \infty$, alors $N(t) \to \infty$ et $N(t)/t \to 1/\mathbb{E}(X_1)$ $\mathbb{P}$-p.s..

CONVERGENCE FAIBLE

7.1. Convergence faible, convergence en probabilité.

Une suite de probabilités boréliennes (μ_n) sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ **converge faiblement vers** μ, une autre probabilité borélienne sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, si pour toute f bornée continue sur $\mathbb{R}$, on a $\mathbb{E}_{\mu_n}(f) \to \mathbb{E}_\mu(f)$. On note alors

$$\mu_n \Rightarrow \mu.$$

Pour $A \subset \mathbb{R}$, on note ∂A sa frontière, i.e. l'ensemble des points adhérents à la fois à A et à A^c : $\partial A = \bar{A} \cap \overline{A^c}$.

THÉORÈME 1. *Les propriétés suivantes sont équivalentes :*
(1) *:* $\mu_n \Rightarrow \mu$*;*
(2) *:* $\mu_n(A) \to \mu(A)$ *pour tout* $A \in \mathcal{B}(\mathbb{R})$ *tel que* $\mu(\partial A) = 0$*;*
(3) *:* $\mu_n(]-\infty, x]) \to \mu(]-\infty, x])$ *pour tout* x *tel que* $\mu(\{x\}) = 0$*;*
(4) *: (***Skorohod***) il existe un espace de probabilités* $(\Omega, \mathcal{B}, \mathbb{P})$ *et des variables aléatoires* $Y, Y_1, Y_2, \ldots$ *sur cet espace, tels que* $Y_n \to Y$ $\mathbb{P}$*-p.s. et* $\mathbb{P}_{Y_n} = \mu_n$*, et* $\mathbb{P}_Y = \mu$*.*

PREUVE.
$\underline{(2) \Rightarrow (3)}$: évident.
$\underline{(1) \Rightarrow (3)}$: soit $\varepsilon > 0$ et soit f_ε linéaire par trois morceaux, $f_\varepsilon : \mathbb{R} \to \mathbb{R}$, continue, telle que $f_\varepsilon(t) = 1$ si $t \leq x$, $f_\varepsilon(t) = 0$ si $t \geq x + \varepsilon$. Alors $\mathbf{1}_{]-\infty,x]} \leq f_\varepsilon \leq \mathbf{1}_{]-\infty,x+\varepsilon]} \leq \mathbf{1}$, et d'après (1), il vient

$$\limsup_n \mu_n(]-\infty, x]) \leq \limsup_n \mathbb{E}_{\mu_n}(f_\varepsilon) = \mathbb{E}_\mu(f_\varepsilon) \leq \mu(]-\infty, x+\varepsilon]),$$

et donc

$$\limsup_n \mu_n(]-\infty, x]) \leq \lim_{\varepsilon \to 0^+} \downarrow \mu(]-\infty, x+\varepsilon]) = \mu(]-\infty, x]).$$

Pour l'inégalité contraire, on considère $f_{-\varepsilon}$ telle que $\mathbf{1}_{]-\infty,x-\varepsilon]} \leq f_{-\varepsilon} \leq \mathbf{1}_{]-\infty,x]}$ pour conclure à

$$\liminf_n \mu_n(]-\infty, x]) \geq \liminf_n \mathbb{E}_{\mu_n}(f_{-\varepsilon}) = \mathbb{E}_\mu(f_{-\varepsilon}) \geq \mu(]-\infty, x-\varepsilon]).$$

Ensuite, $\mu(\{x\}) = 0$ impose $\lim_{\varepsilon \to 0^+} \uparrow \mu(]-\infty, x-\varepsilon]) = \mu(]-\infty, x])$, et la conclusion s'ensuit.
$\underline{(3) \Rightarrow (4)}$: posons $F_n(x) = \mu_n(]-\infty, x])$ et $F(x) = \mu(]-\infty, x])$. Puis posons $(\Omega, \mathcal{B}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), m)$ (mesure de Lebesgue). Notons ensuite

$$Y_n(\omega) = \inf\{x : F_n(x) \geq \omega\} \qquad \text{et} \qquad Y(\omega) = \inf\{x : F(x) \geq \omega\}.$$

On a

$$\begin{aligned} F_{Y_n}(t) &= m(Y_n \leq t) \\ &= m(\{\omega \in [0,1] : \inf\{x : F_n(x) \geq \omega\} \leq t\}) \\ &= m([0, \mu_n(]-\infty, t])) = F_n(t), \end{aligned}$$

et de même $F_Y(t) = F(t)$. Donc $\mathbb{P}_{Y_n} = \mu_n$ et $\mathbb{P}_Y = \mu$.

Supposons que $0 < \omega < 1$. Soit $\varepsilon > 0$, et choisissons x tel que $Y(\omega) - \varepsilon < x < Y(\omega)$, et tel que $\mu(\{x\}) = 0$. Alors $F(x) < \omega$; puisque $F_n(x) \to F(x)$, il s'ensuit que pour n assez grand, $Y(\omega) - \varepsilon < x < Y_n(\omega)$, et donc $\liminf_n Y_n(\omega) \geq Y(\omega)$. Si $\omega < \omega'$ et si $\varepsilon > 0$, choisissons y tel que $Y(\omega') < y < Y(\omega') + \varepsilon$ et $\mu(\{y\}) = 0$. Alors $\omega < \omega' \leq F(Y(\omega')) \leq F(y)$, et donc, pour n grand, $\omega \leq F_n(y)$, et donc $Y_n(\omega) \leq y < Y(\omega') + \varepsilon$. Donc $\limsup_n Y_n(\omega) \leq Y(\omega')$ si $\omega < \omega'$. Et donc $Y_n(\omega) \to Y(\omega)$ si Y est continue en ω.

Mais Y est croissante, et donc n'a qu'un nombre au plus dénombrable de discontinuités. Et donc $m(\{Y_n \to Y\}) = 1$.

$\underline{(4) \Rightarrow (1), (2)}$: si f est borélienne et notons $Disc(f) = \{x : f \text{ discontinue en } x\}$. Alors $Disc(f) \in \mathcal{B}(\mathbb{R})$. Si f est telle que $\mu(\{\text{ discontinuités de f}\}) = 0$, alors $\mathbb{P}$-p.s., $Y \notin Disc(f)$. Et donc $\mathbb{P}(f(Y_n) \to f(Y)) = 1$. Ceci redonne (1) par convergence dominée si $Disc(f) = \emptyset$ (i.e. f continue) et si f est bornée, puisque $\mathbb{E}_{\mathbb{P}}(f(Y_n)) = \mathbb{E}_{\mu_n}(f) \to \mathbb{E}_{\mathbb{P}}(f(Y)) = \mathbb{E}_{\mu}(f)$.

Si $\mu(\partial A) = 0$, alors puisque $Disc(\mathbf{1}_A) \subset \partial A$, (2) en découle de la même façon.

Donc on a $(1) \Rightarrow (3) \Rightarrow (4) \Rightarrow (2), (1)$ et $(2) \Rightarrow (3)$. ■

Soient $Y, Y_1, Y_2, \ldots$ des v.a. définies sur un espace de probabilités $(\Omega, \mathcal{B}, \mathbb{P})$. Nous dirons que (Y_n) **converge en probabilités vers** Y, si pour tout $\varepsilon > 0$, $\mathbb{P}(|Y_n - Y| > \varepsilon) \to 0$. Nous noterons

$$Y_n \xrightarrow{\mathbb{P}} Y.$$

Nous dirons que Y_n **converge en loi vers** Y si $\mathbb{P}_{Y_n}$ converge faiblement vers $\mathbb{P}_Y$, et nous noterons

$$Y_n \xrightarrow{\mathcal{L}} Y.$$

Donc $Y_n \xrightarrow{\mathcal{L}} Y \Leftrightarrow \mathbb{P}_{Y_n} \Rightarrow \mathbb{P}_Y$.

Théorème 2. *Avec les mêmes notations,*

$$(Y_n \to Y \ \mathbb{P} - p.s.) \Rightarrow (Y_n \xrightarrow{\mathbb{P}} Y) \Rightarrow (Y_n \xrightarrow{\mathcal{L}} Y).$$

Preuve. Soit $\varepsilon > 0$. Par le petit Lemme du §6.6., $\mathbb{P}(\limsup_n \{|Y_n - Y| > \varepsilon\}) = 0$.

Par le Lemme de Fatou,

$$\begin{array}{rcl} 0 & = & \mathbb{P}(\liminf_n \{|Y_n - Y| > \varepsilon\}) \\ & = & \mathbb{E}(\liminf_n \mathbf{1}_{|Y_n - Y| > \varepsilon}) \\ & \leq & \liminf_n \mathbb{P}(|Y_n - Y| > \varepsilon) \\ & \leq & \limsup_n \mathbb{P}(|Y_n - Y| > \varepsilon) \\ & \leq & \mathbb{E}(\limsup_n \mathbf{1}_{|Y_n - Y| > \varepsilon}) \\ & = & \mathbb{P}(\limsup_n \{|Y_n - Y| > \varepsilon\}) = 0. \end{array}$$

Donc la convergence presque sûre entraîne la convergence en probabilités.

Soit x tel que $\mathbb{P}_Y(\{x\}) = 0$. Soit $\varepsilon > 0$. On a

$$\mathbb{P}(Y \leq x - \varepsilon) - \mathbb{P}(|Y_n - Y| > \varepsilon) \leq \mathbb{P}(Y_n \leq x) \leq \mathbb{P}(Y \leq x + \varepsilon) + \mathbb{P}(|Y_n - Y| > \varepsilon).$$

On en déduit que pour tout $\varepsilon > 0$,

$$F_Y(x - \varepsilon) \leq \liminf_n F_{Y_n}(x) \leq \limsup_n F_{Y_n}(x) \leq F_Y(x + \varepsilon).$$

Comme $\mathbb{P}_Y(\{x\}) = 0 \Leftrightarrow F_Y$ continue en x, il s'ensuit, en faisant $\varepsilon \downarrow 0^+$, que $\lim_n F_{Y_n}(x) = F_Y(x)$. Donc la convergence en probabilités entraîne la convergence faible. ■

Convergence en loi et loi image. *Supposons $\mu_n \Rightarrow \mu$ et soit $h : \mathbb{R} \to \mathbb{R}$ telle que $\mu(Disc(h)) = 0$. Alors*

$$(\mu_n)_h \Rightarrow \mu_h.$$

Preuve. C'est une conséquence du (4) du Théorème 1 : soit $Y_n \to Y$ $\mathbb{P}$-p.s., avec $\mu_n = \mathbb{P}_{Y_n}$ et $\mu = \mathbb{P}_Y$. Puisque $\mu(Disc(h)) = 0$, alors $\mathbb{P}$-p.s., $Y \notin Disc(h)$. Et donc $\mathbb{P}$-p.s., $h(Y_n) \to h(Y)$. Mais alors par le Théorème 2, $\mathbb{P}_{h(Y_n)} \Rightarrow \mathbb{P}_{h(Y)}$. Et $(\mu_n)_h = (\mathbb{P}_{Y_n})_h = \mathbb{P}_{h(Y_n)}$, et $\mu_h = (\mathbb{P}_Y)_h = \mathbb{P}_{h(Y)}$. ■

Convergence en loi et espérance. *Supposons que $X_n \xrightarrow{\mathcal{L}} X$. Alors*

$$\mathbb{E}(|\mathbf{X}|) \leq \liminf_{\mathbf{n}} \mathbb{E}(|\mathbf{X_n}|).$$

Preuve. Par le théorème de Skorohod ((4) du théorème 1), il existe $Y, Y_1, Y_2, \ldots$ telles que $Y_n \to Y$ $\mathbb{P}$-p.s., et $\mathbb{P}_{Y_n} = \mathbb{P}_{X_n}$ et $\mathbb{P}_X = \mathbb{P}_Y$. Alors $\mathbb{P}_{|Y_n|} = \mathbb{P}_{|X_n|}$, $\mathbb{P}_{|X|} = \mathbb{P}_{|Y|}$, et par le Lemme de Fatou,

$$\mathbb{E}(|X|) = \mathbb{E}(|Y|) \leq \liminf_n \mathbb{E}(|Y_n|) = \liminf_n \mathbb{E}(|X_n|).$$

■

Exercice : supposons que $X_n \xrightarrow{\mathcal{L}} X$ et que $Y_n - X_n \xrightarrow{\mathcal{L}} 0$. Montrer qu'alors $Y_n \xrightarrow{\mathcal{L}} X$.

Convergence en loi et équi-intégrabilité. *Supposons* $X_n \xrightarrow{\mathcal{L}} X$ *et* (X_n) *équi-intégrable. Alors* X *est intégrable et* $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

Preuve. Remplaçons directement (X_n) par (Y_n) et X par Y avec le théorème de Skorohod. Alors (Y_n) est équi-intégrable et converge presque sûrement vers Y. D'après le §6.1., il s'ensuit que Y est intégrable et que $\mathbb{E}(Y_n) \to \mathbb{E}(Y)$, ce qui s'écrit encore (identité des lois) $\mathbb{E}(X_n) \to \mathbb{E}(X)$. ■

7.2. Suites tendues et théorème de sélection d'Helly.

Théorème de sélection d'Helly. *Quelle que soit la suite* (F_n) *de fonctions de répartition, il existe une sous-suite* (F_{n_k}), *une fonction croissante continue à droite* F, *telles qu'en tout point de continuité de* F, (F_{n_k}) *converge ponctuellement vers* F.

Preuve. Par extraction diagonale il existe une sous-suite (n_k) telle que $(F_{n_k}(r))$ converge pour tout rationnel r, vers une limite que nous noterons $G(r)$. Posons $F(x) = \inf\{G(r) : x < r\}$. Clairement F est croissante et continue à droite.

Si F est continue en x, et $\varepsilon > 0$, soit $y < x$ tel que $F(x)-\varepsilon < F(y)$. Soient alors $r, s \in \mathbb{Q}$ tels que $y < r < x < s$, et $G(s) < F(x)+\varepsilon$. Alors on a $F(x)-\varepsilon < G(r) \leq G(s) < F(x)+\varepsilon$, et $F_n(r) \leq F_n(x) \leq F_n(s)$. Et donc

$$F(x) - \varepsilon \leq \liminf_k F_{n_k}(x) \leq \limsup_k F_{n_k}(x) \leq F(x) + \varepsilon,$$

d'où $\lim_k F_{n_k}(x) = F(x)$. ■

Remarquons que la limite F obtenue ci-dessus satisfait $0 \leq F \leq 1$, mais que par contre elle n'est pas forcément une fonction de répartition.

Une suite (μ_n) de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite **tendue** si quel que soit $\varepsilon > 0$, il existe un intervalle $]a, b]$ tel que $\mu_n([a, b]) \geq 1 - \varepsilon$, quel que soit n.

Théorème de sélection pour les suites tendues. *Une CNS pour que d'une suite* (μ_n) *de probabilités sur* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, *et que de l'une quelconque de ses sous-suites* (μ_{n_k}), *il soit possible d'extraire une seconde fois une sous-suite* $(\mu_{n_{k(j)}})$, *et qu'il existe une mesure de probabilités* μ *sur* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *telle que* $\mu_{n_{k(j)}} \xrightarrow[j]{\mathcal{L}} \mu$, *est que la suite* (μ_n) *soit tendue.*

Preuve. C'est surtout la CS qui nous sera utile.
<u>CS</u> : supposons (μ_n) tendue. Notons F_n la fonction de répartition de μ_n. Par le théorème de selction d'Helly, si (F_{n_k}) est une extraite, il en existe une extraite, $(F_{n_{k(j)}})$, et une $0 \leq F \leq 1$ croissante continue à droite, telles que $F_{n_{k(j)}} \to F$ aux points de continuité de F.

Si (μ_n) est tendue, si $\varepsilon > 0$, il existe $a < b$ tels que $F_n(b) - F_n(a) \geq 1 - \varepsilon$ pour tout n. Quitte à augmenter b et diminuer a, on peut supposer que a et b sont deux points de continuité de F. Et donc puisque $0 \leq F \leq 1$ et que F croît, il faut que $\lim_{-\infty} F = 0$ et que $\lim_{+\infty} F = 1$. Donc F est une fonction de répartition, et il existe μ une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dont F soit la fonction de répartition, d'après le Lemme du §4.2.. Par le (3) du Théorème 1, §7.1., il s'ensuit que $\mu_{n_{k(j)}} \Rightarrow_j \mu$.

<u>CN</u> : supposons que (μ_n) ne soit pas tendue, i.e. qu'il existe $\varepsilon > 0$ tel que quel que soit l'intervalle $]a, b]$, il existe n avec $\mu_n(]a, b]) \leq 1 - \varepsilon$. Alors pour chaque k soit n_k un entier tel que $\mu_{n_k}(]-k, k]) \leq 1 - \varepsilon$. Supposons qu'une extraite $(\mu_{n_{k(j)}})$ converge faiblement vers une mesure de probabilité μ. Puisque $\mu(\mathbb{R}) = 1$, il existe $a < b$ tels que $\mu(]a, b]) > 1 - \varepsilon$. Quitte à augmenter b et diminuer a, on peut supposer que $\mu(\{a\}) = \mu(\{b\}) = 0$. Donc $\mu_{n_{k(j)}}(]a, b]) \to_j \mu(]a, b])$. Mais il existe j_0 tel que si $j \geq j_0$, alors $]a, b] \subset]-k(j), k(j)]$ et donc

$$1 - \varepsilon < \mu(]a, b]) = \limsup_j \mu_{n_{k(j)}}(]a, b]) \leq \limsup_j \mu_{n_{k(j)}}(]-k(j), k(j)]) \leq 1 - \varepsilon,$$

une contradiction. ■

Corollaire. *Si (μ_n) est tendue, et si toute sous-suite de (μ_n) convergeant faiblement converge toujours vers la même limite μ, alors $\mu_n \Rightarrow \mu$.*

Preuve. (μ_n) a des sous-suites qui convergent faiblement, puisqu'elle est tendue, d'après le théorème précédent.

Supposons que $\mu_n \not\Rightarrow \mu$. Alors il existe x tel que $\mu(\{x\}) = 0$ et $\varepsilon > 0$ et une extraite (n_k) tels que quel que soit k, $|\mu_{n_k}(]-\infty, x]) - \mu(]-\infty, x]| \geq \varepsilon$. Et par le théorème qui précède, il existe une seconde extraite $(\mu_{n_{k(j)}})$ qui converge faiblement. Mais par hypothèse cela doit être vers μ. Et c'est clairement impossible, d'où une contradiction. ■

7.3. Moments et dérivées des fonctions caractéristiques.

Rappelons que la fonction caractéristique de μ une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou d'une v.a. X est

$$\phi_X(t) = \phi_{\mathbb{P}_X}(t) = \mathbb{E}(e^{itX}) = \int e^{itx} d\mathbb{P}_X(x).$$

Une fonction caractéristique est **uniformément continue** : en effet,

$$|\phi_X(t+h) - \phi_X(t)| \leq \mathbb{E}(|e^{itX}(e^{ihX} - 1)|) = \mathbb{E}(|e^{ihX} - 1|) \to_{h \to 0} 0,$$

par convergence dominée, puisque $0 = \lim_{h\to 0} |e^{ihX} - 1| \leq 2$.

Poursuivons par quelques développements de Taylor : par intégration par parties,

$$\int_0^x (x-s)^n e^{is} ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1} \int_0^x (x-s)^{n+1} e^{is} ds, \qquad \text{(a)}$$

dont il découle par induction que

$$e^{ix} = \sum_{k\leq n} \frac{(ix)^k}{k!} + \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds. \tag{b}$$

En remplaçant n par $n-1$ dans (a), et en déduisant la valeur de l'intégrale de droite dans (a), puis en injectant dans (b), il vient

$$e^{ix} = \sum_{k\leq n} \frac{(ix)^k}{k!} + \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is}-1) ds. \tag{c}$$

En estimant les intégrales dans (b) et (c) et en séparant les cas $x \geq 0$ et $x < 0$, il vient

$$\left| e^{ix} - \sum_{k\leq n} \frac{(ix)^k}{k!} \right| \leq \min\left\{ \frac{|x|^{n+1}|}{(n+1)!}, \frac{2|x|^n}{n!} \right\}. \tag{d}$$

Remplaçons x par une v.a. X dans (d). Il s'ensuit que si $\mathbb{E}(|X|^n) < \infty$ (on dit alors que X **a un moment d'ordre** n),

$$\left| \phi_X(t) - \sum_{k\leq n} \frac{(it)^k}{k!}\mathbb{E}(X^k) \right| \leq \mathbb{E}\left[\min\left\{ \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right\} \right]. \tag{e}$$

Et donc si t est tel que $\lim_n \frac{|t|^n \mathbb{E}(|X|^n)}{n!} = 0$, alors

$$\phi_X(t) = \sum_{k\geq 0} \frac{(it)^k}{k!}\mathbb{E}(X^k). \tag{f}$$

Si la série (f) converge sur un voisinnage de 0, alors ϕ_X permet de retrouver les moments de X :

$$\phi_{\mathbf{X}}^{(\mathbf{k})}(\mathbf{0}) = \mathbf{i}^{\mathbf{k}}\mathbb{E}(\mathbf{X}^{\mathbf{k}}).$$

Par contre X n'a pas forcément de moments de tous ordres (ex : densité $Cte \times \frac{1}{x^3+1}\mathbf{1}_{[0,+\infty[}$). **Toutefois, si $\mathbb{E}(|X|^k) < \infty$, alors $\phi_X^{(k)}$ est uniformément continue, et**

$$\phi_{\mathbf{X}}^{(\mathbf{k})}(\mathbf{t}) = \mathbb{E}((\mathbf{iX})^{\mathbf{k}}\mathbf{e}^{\mathbf{itX}}).$$

Preuve. Observons que

$$\frac{\phi_X(t+h)-\phi_X(t)}{h} - \mathbb{E}(iXe^{itX}) = \mathbb{E}\left[e^{itX}\frac{e^{itX}-1-ihX}{h} \right].$$

D'après (d) pour $n=1$, on a $|e^{itx}-(1+ix)| \leq \min\{\frac{1}{2}x^2, 2|x|\}$, et donc par convergence dominée (par $2|X|$), on en déduit $\phi'_X(t) = \mathbb{E}(iXe^{itX})$, dont l'uniforme continuité se démontre sur le modèle de la preuve de celle de ϕ_X.

Si le moment d'ordre k existe, de proche en proche, on obtient les formules recherchées jusqu'à l'ordre k. ■

L'estimée suivante sera essentielle dans la preuve du théorème limite centrale.

Proposition. *Si* $\mathbb{E}(|\mathbf{X}|^{\mathbf{n}}) < \infty$*, alors*

$$\phi_{\mathbf{X}}(\mathbf{t}) = \sum_{\mathbf{k}\leq\mathbf{n}} \frac{(\mathbf{it})^{\mathbf{k}}}{\mathbf{k}!}\mathbb{E}(\mathbf{X}^{\mathbf{k}}) + \mathbf{o}(\mathbf{t}^{\mathbf{n}}) \textbf{ si t voisin de } \mathbf{0},$$

et en particulier si $n = 2$*,*

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbf{1} + \mathbf{it}\mathbb{E}(\mathbf{X}) - \frac{\mathbf{t}^{\mathbf{2}}}{\mathbf{2}}\mathbb{E}(\mathbf{X}^{\mathbf{2}}) + \mathbf{o}(\mathbf{t}^{\mathbf{2}}). \tag{g}$$

Preuve. Par (d), si nous posons $\Delta_n(t, X) = e^{itX} - \sum_{k\leq n} \frac{(it)^k X^k}{k!}$, on a pour $0 < |t| \leq 1$,

$$\begin{cases} |\frac{1}{t^n}\Delta(t, X)| \leq \frac{2}{n!}|X|^n \in \mathcal{L}^1; \\ 0 \leq \liminf_{t\to 0} |\frac{1}{t^n}\Delta(t, X)| \leq \limsup_{t\to 0} |\frac{1}{t^n}\Delta(t, X)| \leq \lim_{t\to 0} |X|^{n+1}\frac{|t|}{(n+1)!} = 0 \ \mathbb{P} - p.s.. \end{cases}$$

Ensuite, $|\mathbb{E}(\frac{1}{t^n}\Delta(t, X))| \leq \mathbb{E}(\frac{1}{|t|^n}|\Delta(t, X)|)$, et donc par convergence dominée et écrabouillage, il s'en suit que $\lim_0 \frac{1}{t^n}\mathbb{E}(\Delta(t, X)) = 0$, et donc $\mathbb{E}(\Delta(t, X)) = o(t^n)$ pour t voisin de 0. ■

7.4. Indépendance et Lemme de Riemann-Lebesgue.

Fonction caractéristique de la somme de deux v.a. indépendantes. *Si* $X \perp\!\!\!\perp Y$*, alors*

$$\phi_{\mathbf{X}+\mathbf{Y}}(\mathbf{t}) = \phi_{\mathbf{X}}(\mathbf{t})\phi_{\mathbf{Y}}(\mathbf{t}).$$

Notons aussi que $\phi_{\mathbf{aX}+\mathbf{b}}(\mathbf{t}) = \mathbf{e}^{\mathbf{itb}}\phi_{\mathbf{X}}(\mathbf{at})$.

Preuve. Puisque $X \perp\!\!\!\perp Y$, $e^{itX} \perp\!\!\!\perp e^{itY}$, et donc $\phi_{X+Y}(t) = \mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX}e^{itY}) = \mathbb{E}(e^{itX})\mathbb{E}(e^{itY})$. ■

Lemme de Riemann-Lebesgue. *Si* μ *est une probabilité sur* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *absolument continue relativement à la mesure de Lebesgue* m *(i.e.* μ *a une densité), alors*

$$\lim_{|\mathbf{t}|\to\infty} \phi_\mu(\mathbf{t}) = \mathbf{0}.$$

Preuve. D'après le cours de Licence d'Intégration, les fonctions $\mathcal{C}^\infty$ à support compact sont denses dans $L^1(m)$. Soit donc (ψ_n) une suite de telles fonctions telles que si $f = d\mu \,/\, dm$, $|f - \psi_n|_1 = \int |f - \psi_n| dm \to_n 0$. Notons que $\int |e^{itx}f(x) - \psi_n(x)e^{itx}|dx = |f - \psi_n|_1$.

Par intégration par parties, ψ_n étant à support compact (donc nulle en $\pm\infty$),

$$\int \psi_n(x)e^{itx}dx = -\frac{1}{it}\int \psi_n'(x)e^{itx}dx \to_{|t|\to\infty} 0,$$

par convergence dominée, du fait que ψ_n' est à support compact aussi.

Soit $\varepsilon > 0$, et n tel que $|f-\psi_n|_1 < \varepsilon$. Alors

$$\limsup_{|t|\to\infty} |\int f(x)e^{itx}dx| \leq |f-\psi_n|_1 + \limsup_{|t|\to\infty} |\int \psi_n(x)e^{itx}dx| < \varepsilon,$$

et donc $\lim_{|t|\to\infty}\int f(x)e^{itx}dx = \lim_{|t|\to\infty}\phi_\mu(t) = 0$. ■

7.5. Théorème d'inversion et unicité.

Lemme préliminaire. *On a*

$$\lim_{T\to+\infty}\int_0^T \frac{\sin(\theta t)}{t}dt = \begin{cases} \frac{\pi}{2} \ \text{si } \theta > 0; \\ 0 \ \text{si } \theta = 0; \\ -\frac{\pi}{2} \ \text{si } \theta < 0. \end{cases}$$

Preuve. Posons $sgn(\theta) = 1, 0$ ou -1 selon que $\theta > 0, = 0$, ou < 0. Alors si $S(T,\theta) = \int_0^T \frac{\sin(\theta t)}{t}dt$, $S(T,\theta) = sgn(\theta)S(T|\theta|, 1)$.

Puis on remarque que, par dérivation relativement à T,

$$\int_0^T e^{-ux}\sin x dx = \frac{1}{1+u^2}\left[1 - e^{-uT}(u\sin T + \cos T)\right].$$

Ensuite, $\int_0^T \left(\int_0^{+\infty} |e^{-ux}\sin x|du\right)dx = \int_0^T \frac{|\sin x|}{x}dx \leq T < +\infty$, et donc par Fubini,

$$\begin{array}{rcl} S(T,1) & = & \int_0^T \frac{\sin x}{x}dx \\ & = & \int_0^T \sin x \left(\int_0^\infty e^{-ux}du\right)dx \\ & = & \int_0^\infty \left[\int_0^T e^{-ux}\sin x dx\right]du \\ & = & \int_0^\infty \frac{du}{1+u^2} - \int_0^\infty \frac{e^{-uT}}{1+u^2}(u\sin T + \cos T)du \\ & \to_{T\to\infty} & \frac{\pi}{2} \text{ (par convergence dominée appropriée).} \end{array}$$

■

Théorème d'inversion et d'unicité. *Soit μ une mesure de probabilités sur* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Alors*

$$\mu(]\mathbf{a}, \mathbf{b}[) + \frac{\mathbf{1}}{\mathbf{2}}(\mu(\{\mathbf{a}\}) + \mu(\{\mathbf{b}\})) = \lim_{\mathbf{T}\to+\infty} \frac{\mathbf{1}}{\mathbf{2}\pi} \int_{-\mathbf{T}}^{\mathbf{T}} \frac{\mathbf{e^{-ita} - e^{-itb}}}{\mathbf{it}} \phi_\mu(\mathbf{t})\mathbf{dt}.$$

Il en découle que si ν est une autre mesure de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *et si $\phi_\nu = \phi_\mu$, alors $\nu = \mu$.* **Donc la fonction caractéristique est bien caractéristique de la loi**.

Preuve. Si la formule d'inversion est vraie, et si $\phi_\nu = \phi_\mu$, alors $\nu(]a, b]) = \mu(]a, b])$ pour tout intervalle $]a, b]$ tel que $\mu(\{a\}) = \mu(\{b\}) = \nu(\{a\}) = \nu(\{b\}) = 0$. Mais la classe de ces intervalles constitue un π-système générateur de $\mathcal{B}(\mathbb{R})$, et donc $\mu = \nu$.

Notons $I_T = \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it} \phi_\mu(t)dt$. Par Fubini (il n'y a pas de problème en $t = 0$) $(m([-T, T]) < +\infty)$, on a

$$I_T = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left[\int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt \right] d\mu(x).$$

En utilisant la formule de De Moivre pour les exponentielles complexes, et du fait que le cos est pair, sa contribution dans l'intégrale disparaît, et il vient, via le Lemme préliminaire,

$$I_T = \int_{-\infty}^{+\infty} \left[\frac{sgn(x-a)}{\pi} S(T|x-a|, 1) - \frac{sgn(x-b)}{\pi} S(T|x-b|, 1) \right] d\mu(x).$$

L'intégrande ci-dessus est bornée et converge lorsque $T \to \infty$ vers la fonction

$$\psi_{a,b}(x) = \begin{cases} 0 \text{ si } x < a; \\ \frac{1}{2} \text{ si } x = a; \\ 1 \text{ si } a < x < b; \\ \frac{1}{2} \text{ si } x = b; \\ 0 \text{ si } b < x. \end{cases}$$

Donc $\lim_{T\to\infty} I_T = \int \psi_{a,b}(x)d\mu(x) = \frac{1}{2}(\mu(\{a\}) + \mu(\{b\})) + \mu(]a, b[)$. ■

Corollaire. *Supposons que $\phi_\mu \in \mathcal{L}^1(m)$. Alors μ a une densité continue.*

Preuve. On suppose $\int_{\infty}^{+\infty} |\phi_\mu(t)|dt < \infty$. On a d'apr'es (d) pour $n = 0$,

$$\left| \frac{e^{ita} - e^{itb}}{it} \right| \leq |b - a|,$$

et donc par la formule d'inversion, si $\mu(\{a\}) = \mu(\{b\}) = 0$, $\mu([a,b]) \leq \frac{(b-a)}{2\pi}|\phi_\mu|_1$, et donc μ est sans charge ponctuelle (diffuse). Soit F la fonction de répartition de μ. Alors

$$\frac{F(x+h)-F(x)}{h} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{-itx} - e^{-it(x+h)}}{ith}\phi_\mu(t)dt,$$

où l'intégrande est dominé par $|\phi_\mu(t)|$ et converge vers $e^{itx}\phi_\mu(t)$ si $h \to 0$. Et donc F est dérivable et sa dérivée est

$$\left(\frac{\mathbf{d}\mu}{\mathbf{dx}}(\mathbf{x}) =\right) \qquad \mathbf{f}(\mathbf{x}) = \frac{\mathbf{1}}{\mathbf{2}\pi}\int_{-\infty}^{+\infty} \mathbf{e}^{-\mathbf{itx}}\phi_\mu(\mathbf{t})\mathbf{dt}.$$

Clairement f est continue, et c'est la densité de F, par le théorème fondamental du calcul intégral. ■

Exercice : "*une formule d'inversion de la transformée de Laplace*".

a) Soit $\theta, \lambda > 0$ et supposons que la loi de la variable X_λ soit une loi de Poisson de paramètre $\lambda\theta$, $\mathcal{L}(X_\lambda) = \mathcal{P}(\lambda\theta)$. Montrer que $\frac{X_\lambda - \lambda\theta}{\lambda} \xrightarrow{\mathbb{P}} 0$ si $\lambda \to \infty$. En déduire que

$$\lim_{\lambda\to\infty} e^{-\lambda\theta}\sum_{k\leq\lambda x}\frac{(\lambda\theta)^k}{k!} = \begin{cases} 0 \text{ si } \theta > x; \\ 1 \text{ si } \theta < x. \end{cases}$$

b) Soit $\mathbb{P}$ une loi de probabilité sur $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$, et notons $L(t) := \int_0^\infty e^{-tx}d\mathbb{P}(x)$ sa transformée de Laplace. Montrer que $L : \mathbb{R}^+ \to \mathbb{R}^+$ est indéfiniment dérivable, et que

$$\lim_{\lambda\to :infty}\sum_{k\leq\lambda x}\frac{(-1)^k}{k!}\lambda^k L^{(k)}(\lambda) = F_{\mathbb{P}}(x),$$

en tout point de continuité x de $F_{\mathbb{P}}$.

7.6. Théorème de continuité.

Théorème de continuité de Lévy. *Soient $\mu, \mu_1, \mu_2, \ldots$ des mesures de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors une CNS pour que $\mu_n \Rightarrow \mu$ est que pour chaque $t \in \mathbb{R}$, $\phi_{\mu_n}(t) \to \phi_\mu(t)$.*

Preuve. Supposons que $\mu_n \Rightarrow \mu$. Alors comme pour chaque $t \in \mathbb{R}$, $x \mapsto e^{itx}$ est continue bornée, il s'ensuit, par définition de la convergence faible, que $\phi_{\mu_n}(t) = \mathbb{E}_{\mu_n}(e^{itx}) \to_n \mathbb{E}_\mu(e^{itx}) = \phi_\mu(t)$.

Réciproquement, par Fubini,

$$\left\{\begin{array}{rcl} \frac{1}{u}\int_{-u}^{u}(1-\phi_{\mu_n}(t))dt & = & \int_{-\infty}^{\infty}\left[\frac{1}{u}\int_{-u}^{u}(1-e^{itx})dt\right]d\mu_n(x) \\ & = & 2\int_{-\infty}^{\infty}\left(1-\frac{\sin(ux)}{ux}\right)d\mu_n(x) \\ & \geq & 2\int_{|x|\geq\frac{2}{u}}\left(1-\frac{1}{|ux|}\right)d\mu_n(x) \\ & \geq & \mu_n(|x|\geq\frac{2}{u}) \end{array}\right. \qquad \text{(h)}$$

(remarquons que la première intégrale est réelle).

Puisque ϕ_μ est continue et $\phi_\mu(0) = 1$, pour $\varepsilon > 0$ donné, il existe u tel que $\frac{1}{u}\int_{-u}^{u}(1 - \phi_\mu(t))dt < \varepsilon$. Puisque $\phi_{\mu_n} \to \phi_\mu$ simplement, par convergence dominée, il existe n_0 tel que $n \geq n_0 \Rightarrow \frac{1}{u}\int_{-u}^{u}(1 - \phi_{\mu_n}(t))dt < 2\varepsilon$.

Si $a = \frac{2}{u}$, et au besoin en augmentant a, il s'en suit que, même pour les $n < n_0$, on aura

$$\mu_n(|x| \geq a) < 2\varepsilon,$$

et donc la suite (μ_n) est tendue.

Soit alors μ_{n_k} une extraite convergeant faiblement vers ν une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors par la CN que nous avons déjà démontrée, $\phi_\nu = \lim \phi_{\mu_{n_k}} = \phi_\mu$, et donc $\nu = \mu$. Donc toutes les extraites de (μ_n) qui convergent faiblement, convergent vers μ. Et donc $\mu_n \Rightarrow \mu$. ■

Corollaire 1. *Soit (μ_n) une suite de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et supposons que la suite (ϕ_{μ_n}) converge simplement vers une fonction g. Alors si g est continue en 0, il existe une probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $\mu_n \Rightarrow \mu$ et que $\phi_\mu = g$.*

Preuve. Un calcul similaire à (h) et le raisonnement qui le suit dans la preuve précédente permet de montrer que (μ_n) est tendue. La conclusion est la même. ■

Corollaire 2. *Supposons (μ_n) tendue, et que la suite (ϕ_{μ_n}) converge simplement vers g. Alors il existe μ telle que $\mu_n \Rightarrow \mu$ et que $g = \phi_\mu$.*

Preuve. Si (μ_n) est tendue, elle a une extraite qui converge faiblement vers μ. Mais alors par convergence simple et le théorème de continuité, $\phi_\mu = g$.

Ensuite toutes les extraites qui convergent faiblement doivent par le même argument converger vers une mesure ayant ϕ_μ comme fonction caractéristique. Et donc μ est la seule limite possible. Et donc $\mu_n \Rightarrow \mu$. ■

7.7. Le théorème limite centrale.

TLC. *Soit (X_n) une suite i.i.d. dans $\mathcal{L}^2(\mathbb{P})$ ayant une moyenne c et une variance $\sigma^2 > 0$. Notons $S_N = \sum_{k=1}^{N} X_k$. Alors*

$$\frac{\mathbf{S_N - Nc}}{\sigma\sqrt{\mathbf{n}}} \xrightarrow{\mathcal{L}} \mathbf{Y},$$

où Y a pour loi la loi normale centrée réduite $\mathcal{N}(0,1)$, de densité $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

Preuve. Notons $Y_N = \frac{S_N - Nc}{\sigma\sqrt{n}}$. Quitte à remplacer X_k par $(X_k - c)\ /\ \sigma$, on peut supposer que $c = 0$ et $\sigma = 1$. Notons $\phi_X = \phi_{X_k}$.

Nous allons montrer que $\phi_{Y_N} \to \phi_Y$ puis conclure en invoquant le théorème de continuité. Calculons $\phi_Y(t)$ d'abord. Grâce à l'exponentielle, ϕ_Y est dérivable, et

$$\phi'_Y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ixe^{itx} e^{-\frac{x^2}{2}} dx = \ldots = -t\phi_Y(t),$$

en intégrant par parties. Donc $\log \phi_Y(t) = -\frac{t^2}{2} + Cte$, et $\log \phi_Y(0) = \log 1 = 0$. Donc $\phi_Y(t) = e^{-\frac{t^2}{2}}$.

Ensuite par indépendance, on a $\phi_{Y_N}(t) = \phi_X(\frac{t}{\sqrt{n}})^n$. D'après (e) pour $n = 2$, on a

$$\phi_{Y_N}(t) - e^{-\frac{t^2}{2}} = \left(1 - \frac{t^2}{2n} + o(\frac{1}{n})\right)^n - \left(e^{-\frac{t^2}{2n}}\right)^n.$$

Si $z_1, \ldots, z_n, y_1, \ldots, y_n$ sont des nombres complexes de module plus petit que 1, il est facile de montrer que

$$\left| \prod_{i=1}^n z_i - \prod_{i=1}^n y_i \right| \le \sum_{i=1}^n |z_i - y_i|.$$

Ceci s'applique à l'expression précédente de $|\phi_{Y_N}(t) - e^{-\frac{t^2}{2}}|$ pour conclure que $\phi_{Y_N}(t) - e^{-\frac{t^2}{2}} \to 0$. ■

Puisque la loi normale ne charge pas les points, il s'en suit que si $x \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{\mathbf{S_N - Nc}}{\sigma\sqrt{\mathbf{n}}} \le \mathbf{x}\right) \to \int_{-\infty}^{\mathbf{x}} \frac{\mathbf{e}^{-\frac{\mathbf{t^2}}{\mathbf{2}}}}{\sqrt{\mathbf{2}\pi}} \mathbf{dt}.$$

Citons sans preuve le

Théorème de Berry-Esseen. *Sous les hypothèses du TLC, si en plus* $\mathbb{E}(|X_1|^3) = \rho < \infty$*, alors si* $x \in \mathbb{R}$*,*

$$\left| \mathbb{P}\left(\frac{\mathbf{S_N - Nc}}{\sigma\sqrt{\mathbf{n}}} \le \mathbf{x}\right) - \int_{-\infty}^{\mathbf{x}} \frac{\mathbf{e}^{-\frac{\mathbf{t^2}}{\mathbf{2}}}}{\sqrt{\mathbf{2}\pi}} \mathbf{dt} \right| \le \frac{\mathbf{3}\rho}{\sqrt{\mathbf{n}\sigma^{\mathbf{3}}}}.$$

EXERCICES

I : que X ait une loi a) $\mathcal{P}(\lambda)$ ou b) $\Gamma(1, \lambda)$ alors $Y_\lambda = (X - \mathbb{E}(X))var(X)^{-1/2}$ converge en loi vers $\mathcal{N}(0, 1)$.

Montrer que

$$\lim_n e^{-n} \left(1 + n + \frac{n^2}{2!} + \ldots + \frac{n^n}{n!}\right) = \frac{1}{2}.$$

II : trouver une suite de fonctions caractéristiques qui converge partout ponctuellement mais telle que la limite ne soit pas une fonction caractéristique (indication : lois uniformes sur $[-n,+n]$).

III : montrer que si $x \geq 0$, et $n \to \infty$, alors

$$\sum_{|k-n/2|\leq x\sqrt{n}/2} C_n^k \sim 2^n \int_{-x}^{x} e^{-u^2/2}du/\sqrt{2\pi};$$
$$\sum_{|k-n|\leq x\sqrt{n}} n^k/k! \sim e^n \int_{-x}^{x} e^{-u^2/2}du/\sqrt{2\pi}.$$

IV : soit $(X_i)_{i\geq 1}$ une suite iid de var de Cauchy. Si $S_n = X_1 + \ldots + X_n$, trouver $\mathbb{P}(S_n > na)$ $(a > 0)$.

V : soit $(X_i)_{i\geq 1}$ une suite iid de var équidistribuées sur $\{-1,+1\}$. Montrer qu'on a la convergence en loi

$$\sqrt{\frac{3}{n^3}}\sum_{k=1}^{n} kX_k \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

VI : **Formule de Stirling**.

Soit $(X_i)_{i\geq 1}$ une suite i.i.d. de loi commune la loi $\mathcal{P}(1)$. Notons $S_n = X_1 + \ldots + X_n$.

a) : montrer que $\mathbb{E}\left[\left(\frac{S_n-n}{\sqrt{n}}\right)^-\right] = e^{-n}\sum_{k=0}^{n}\left(\frac{n-k}{\sqrt{n}}\right)\frac{n^k}{k!} = \frac{n^{n+\frac{1}{2}}e^{-n}}{n!}$.

b) : montrer que $\left(\frac{S_n-n}{\sqrt{n}}\right)^- \xrightarrow{\mathcal{L}} N^-$, où N^- désigne la partie négative d'une v.a. N de loi $\mathcal{N}(0,1)$.

c) : montrer que $\mathbb{E}\left[\left(\frac{S_n-n}{\sqrt{n}}\right)^-\right] \to_n \mathbb{E}(N^-) = \frac{1}{\sqrt{2\pi}}$.

d) : conclure que $n! \sim \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}$.

Chapitre 8
Théorèmes limites dans $\mathbb{R}^k$

8.1. Convergence faible et suites tendues.

Soient μ et $\mu_n,\ n \geq 1$, des mesures de probabilités sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Soient X et $X_n,\ n \geq 1$, des vecteurs aléatoires à valeurs dans $\mathbb{R}^k$. Soient enfin F et $F_n,\ n \geq 1$, des fonctions de répartition de lois de probabilités, ou de vecteurs aléatoires, sur $\mathbb{R}^k$.

Nous dirons que F_n **converge faiblement vers** F**, noté** $F_n \Rightarrow F$, si $F_n(x) \to F(x)$ en tout point de continuité x de F.

Nous dirons que μ_n **converge faiblement vers** μ**, et nous noterons** $\mu_n \Rightarrow \mu$, si les fonctions de répartition associées convergent faiblement.

Nous dirons que X_n **converge en loi vers** X**, noté** $X_n \xrightarrow{\mathcal{L}} X$, si $\mathbb{P}_{X_n} \Rightarrow \mathbb{P}_X$.

Comme pour la droite, nous caractérisons la convergence faible :

Conditions équivalentes pour la convergence faible dans $\mathbb{R}^k$. *Avec les mêmes notations,* $\mu_n \Rightarrow \mu$ *si et seulement si l'une des conditions suivantes est satisfaite :*

$$\begin{array}{rcl} (i) & : & \lim_n \mathbb{E}_{\mu_n}(f) = \mathbb{E}_\mu(f) \text{ pour toute } f : \mathbb{R}^k \to \mathbb{R} \text{ continue bornée;} \\ (ii) & : & \limsup_n \mu_n(C) \leq \mu(C) \text{ pour tout } C \subset \mathbb{R}^k \text{ fermé;} \\ (iii) & : & \liminf_n \mu_n(G) \geq \mu(G) \text{ pour tout } G \subset \mathbb{R}^k \text{ ouvert;} \\ (iv) & : & \lim_n \mu_n(A) = \mu(A) \text{ pour tout } A \in \mathcal{B}(\mathbb{R}^k) \text{ de frontière } \mu\text{-négligeable.} \end{array}$$

Preuve. Montrons d'abord que les quatre conditions exprimées sont équivalentes entre elles.

<u>$(i) \Rightarrow (ii)$</u> : supposons $C \neq \emptyset$ et posons $dist(x, C) = \inf\{|x - y| : y \in C\}$. On a $x \mapsto dist(x, C)$ continue. Posons, pour $j > 0$,

$$\phi_j(t) = \begin{cases} 1 & \text{si } t \leq 0; \\ 1 - jt & \text{si } 0 \leq t \leq \frac{1}{j}; \\ 0 & \text{si } \frac{1}{j} \leq t. \end{cases}$$

Puis notons $f_j(x) = \phi_j(dist(x, C))$: f_j est continue bornée par 1, et $\lim_{j \to +\infty} f_j(x) = \mathbf{1}_C(x)$ car C est fermé. Par (i), on a $\limsup_n \mu_n(C) \leq \lim_n \mathbb{E}_{\mu_n}(f_j) = \mathbb{E}_\mu(f_j)$, et par convergence dominée, $\lim_{+\infty} \mathbb{E}_\mu(f_j) = \mu(C)$.

<u>$(ii) \Leftrightarrow (iii)$</u> : poser $C = \mathbb{R}^k \setminus G$.

<u>$\big((ii) \wedge (iii)\big) \Rightarrow (iv)$</u> : en combinant les deux, on a

$$\mu(\overset{\circ}{A}) \leq \liminf_n \mu_n(\overset{\circ}{A}) \leq \liminf_n \mu_n(A) \leq \limsup_n \mu_n(A) \leq \limsup_n \mu_n(\bar{A}) \leq \mu(\bar{A}),$$

et puisque $\mu(\partial A) = 0$, on a $\mu(\overset{\circ}{A}) = \mu(\bar{A})$.

<u>$(iv) \Rightarrow (i)$</u> : supposons f continue et $|f| \leq K$ (f bornée). Etant donné $\varepsilon > 0$, choisissons $\alpha_0 < \alpha_1 < \ldots < \alpha_l$ tels que $\alpha_0 < -K < K < \alpha_l$ et que $\alpha_{i+1} - \alpha_i < \varepsilon$, et que $\mu(\{f = \alpha_i\}) = 0$ pour chaque i.

Notons alors $A_i = \{\alpha_{i-1} < f \leq \alpha_i\}$. Alors par continuité de f,

$$\{\alpha_{i-1} < f < \alpha_i\} \subset \overset{\circ}{A}_i \subset \bar{A}_i \subset \{\alpha_{i-1} \leq f \leq \alpha_i\},$$

et donc $\partial A_i \subset \{f = \alpha_{i-1}\} \cup \{f = \alpha_i\}$. Et donc $\mu(\partial A_i) = 0$. D'autre part, que ν désigne μ_n ou μ,

$$\left| \int f d\nu - \sum_{i=1}^{l} \alpha_i \nu(A_i) \right| \leq \varepsilon, \text{ et } \sum_{i=1}^{l} \alpha_i \mu_n(A_i) \to \sum_{i=1}^{l} \alpha_i \mu(A_i),$$

d'après (iv). Puisque ε est arbitraire, on en déduit (i).

Ainsi nous avons montré $(i) \Rightarrow (ii) \Leftrightarrow (iii) \Rightarrow (iv) \Rightarrow (i)$, et reste donc à montrer par exemple que $(iv) \Rightarrow (\mu_n \Rightarrow \mu)$ et que $(\mu_n \Rightarrow \mu) \Rightarrow (iii)$.

<u>$(iv) \Rightarrow (\mu_n \Rightarrow \mu)$</u> : soit $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$, et soient F_n et F les fonctions de répartition respectives de μ_n et μ. On a donc $\mu_n(S_x) = F_n(x)$ et $\mu(S_x) = F(x)$ si

$$S_x = \{y \in \mathbb{R}^k : y_i \leq x_i,\ 1 \leq i \leq k\}.$$

La continuité de F en x équivaut à la condition $\mu(\partial S_x) = 0$, et donc $(iv) \Rightarrow (F_n \Rightarrow F)$.

<u>$(\mu_n \Rightarrow \mu) \Rightarrow (iii)$</u> : comme seul un ensemble au plus dénombrable d'hyperplans parallèles peuvent éventuellement ne pas être μ-négligeables, il existe un ensemble D dense dans $\mathbb{R}$ tel que pour chaque i, et chaque $t \in D$, $\mu(\{x : x_i = t\}) = 0$.

Soit $\mathcal{A}$ la classe des pavés $A = \prod_{i=1}^{k}]a_i, b_i]$ avec $a_i, b_i \in D$. Par construction, on a $\mu_n(A) \to \mu(A)$, puisque par hypothèse, $F_n \to F$ là où F est continue, et que tant $\mu_n(A)$ que $\mu(A)$ s'expriment à l'aide des valeurs de F_n et F en des points de continuité de F.

Et donc si $B \in \tilde{\mathcal{A}}$, la classe des unions finies disjointes d'éléments de $\mathcal{A}$, $\mu_n(B) \to \mu(B)$. Maintenant par densité de D, si $G \subset \mathbb{R}^k$ est ouvert, il s'écrit comme union croissante d'éléments de $\tilde{\mathcal{A}}$, disons $G = \cup_n \uparrow B_n$, et alors

$$\liminf_n \mu_n(G) \geq \liminf_n \mu_n(\cup_{p \leq q} B_p) = \mu(\cup_{p \leq q} B_p),$$

pour tout q, et donc $\mu(G) = \sup_q \mu(\cup_{p \leq q} B_p) \leq \liminf_n \mu_n(G)$. ∎

Transfert et convergence faible dans $\mathbb{R}^k$. *Soit $h : \mathbb{R}^k \to \mathbb{R}^j$ une application. Alors l'ensemble D_h de ses points de discontinuité est mesurable. Supposons h borélienne, et d'autre part que $\mu_n \Rightarrow \mu$ dans $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Alors si $\mu(D_h) = 0$, on a*

$$(\mu_n)_h \Rightarrow \mu_h.$$

Preuve. Montrons d'abord que $D_h \in \mathcal{B}(\mathbb{R}^k)$: soit

$$A(\varepsilon, \delta) = \{x \in \mathbb{R}^k : \exists y, z \in \mathbb{R}^k : |x - y| < \delta \ \wedge \ |x - z| < \delta \ \wedge \ |h(y) - h(z)| > \varepsilon\}.$$

Alors que h soit mesurable ou pas, $A(\varepsilon, \delta)$ est ouvert et

$$D_h = \cup_{\varepsilon \in \mathbb{Q}^+} \cap_{\delta \in \mathbb{Q}^+} A(\varepsilon, \delta) \in \mathcal{B}(\mathbb{R}^k).$$

Soit $C \subset \mathbb{R}^j$ fermé. Alors $\overline{h^{-1}C} \subset D_h \cup h^{-1}C$. Si $\mu_n \Rightarrow \mu$, par (ii),

$$\limsup_n \mu_n(h^{-1}C) \leq \limsup_n \mu_n(\overline{h^{-1}C}) \leq \mu(\overline{h^{-1}C}) \leq \mu(D_h \cup h^{-1}C) = \mu(h^{-1}C),$$

et donc encore par (ii), $\left(\mu_n\right)_h \Rightarrow \mu_h$. ∎

Notons que dans le cas de la droite, nous avions utilisé, pour montrer le résultat correspondant, le théorème de Skorohod : il reste valable pour $\mathbb{R}^k$, mais sa preuve est plus complexe. Citons le sans preuve :

Skorohod version $\mathbb{R}^k$. *Si X_n, $n \geq 1$ et X sont des vecteurs aléatoires à valeurs dans $\mathbb{R}^k$, et si $X_n \xrightarrow{\mathcal{L}} X$, alors il existe un espace probabilisé $(\Omega, \mathcal{B}, \mathbb{P})$, une suite de vecteurs aléatoires $Y_n : \Omega \to \mathbb{R}^k$ et un autre vecteur aléatoire $Y : \Omega \to \mathbb{R}^k$, tels que $Y_n \to Y$ $\mathbb{P}$-p.s., et qu'en outre*

$$\mathbb{P}_{Y_n} = \mathbb{P}_{X_n} \text{ et } \mathbb{P}_Y = \mathbb{P}_X.$$

Suites tendues dans $\mathbb{R}^k$. *Une suite de probabilités $(\mu_n)_{n\geq 1}$ sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est* **tendue** *si quel que soit $\varepsilon > 0$, il existe un rectangle $A \in \mathcal{B}(\mathbb{R}^k)$ tel que*

$$\inf_n \mu_n(A) \geq 1 - \varepsilon.$$

Une adaptation de la preuve correspondante sur la droite donne le résultat suivant :
Pseudo-pré-compacité des suites tendues *: toute suite tendue a une extraite qui converge faiblement.*

Caractérisation des suites tendues convergentes *: si une suite tendue est telle que toutes ses extraites qui convergent, convergent vers la même limite, alors elle converge faiblement, vers la limite commune.*

Preuve de la seconde affirmation. Pour l'essentiel on commence par reporter le théorème de selection d'Helly. La différence est que les fonctions "*croissantes*" ne sont plus suffisantes, modulo le fait qu'elles aient les bonnes limites en $\pm\infty$, pour caractériser les fonctions de répartition des vecteurs aléatoires. Il faut une condition supplémentaire, qui est à vérifier - en plus, dite de *positivité sur les rectangles* : par exemple dans $\mathbb{R}^2$, pour le rectangle $A =]a,b]\times]c,d]$, il faut

$$\Delta_A F = F(b,d) - F(b,c) - F(a,d) + F(a,c) \geq 0.$$

Il n'y a pas de difficulté spécifique et nous préférons passer... ■

8.2. Fonctions caractéristiques.

Soient $t = (t_1, \ldots, t_k) \in \mathbb{R}^k$ et $X = (X_1, \ldots, X_k)$ un vecteur aléatoire. Alors on pose

$$< t, X >= \sum_{i=1}^{k} t_i X_i.$$

La **fonction caractéristique de X (ou de sa loi $\mathbb{P}_X$) est définie par**

$$\phi_X(t) = \mathbb{E}(e^{i<t,X>}) = \int_{\mathbb{R}^k} e^{i<t,x>} d\mu(x) = \phi_\mu(t) \text{ si } \mathbb{P}_X = \mu.$$

Pour l'essentiel, les propriétés de ϕ_X sont identiques à celles du cas de la droite ($k = 1$). Citons notamment l'uniforme continuité, et surtout la formule d'inversion :

Formule d'inversion dans $\mathbb{R}^k$. *Si $A = \prod_{i=1}^{k}]a_i, b_i]$ est un rectangle borné, et si $\mu(\partial A) = 0$, alors*

$$\mu(A) = \lim_{T\to\infty} \frac{1}{(2\pi)^k} \int_{[-T,T]^k} \prod_{u=1}^{k} \frac{e^{-it_u a_u} - e^{-it_u b_u}}{it_u} \phi_\mu(t) dt,$$

où $dt = dt_1 \ldots dt_k$.

Et donc la fonction caractéristique caractérise bien la loi.

Preuve. On pose $B_T = [-T,T]^k$, et comme dans le cas $k = 1$, en posant

$$I_T = \frac{1}{(2\pi)^k} \int_{[-T,T]^k} \prod_{u=1}^{k} \frac{e^{-it_u a_u} - e^{-it_u b_u}}{it_u} \phi_\mu(t) dt,$$

on a par Fubini

$$\begin{array}{rcl} I_T & = & \int_{\mathbb{R}^k} \prod_{u=1}^k \left[\frac{sgn(x_u - a_u)}{\pi} S(T|x_u - a_u|) - \frac{sgn(x_u - b_u)}{\pi} S(T|x_u - b_u|) \right] d\mu(x) \\ & \underset{T\to+\infty}{\to} & \int_{\mathbb{R}^k} \prod_{u=1}^k \psi_{a_u,b_u}(x_u). \end{array}$$

La conclusion s'en suit. Remarquons qu'ici nous pourrions affiner notre énoncer en regardant de plus près les valeurs éventuelles de $\mu(\partial A)$.

Le fait que la fonction caractéristique caractérise la loi se montre comme pour la droite, une fois la formule d'inversion établie. ■

Pour $t \in \mathbb{R}^k$, notons $h_t : \mathbb{R}^k \to \mathbb{R}$ définie par $h_t(x) = < t, x >$.

Corollaire : caractérisation de μ sur les demi-plans. *Une mesure de probabilités μ sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est entièrement déterminée par ses valeurs $\mu(\{x :< t, x >\leq \alpha\})$, $t \in \mathbb{R}^k$, $\alpha \in \mathbb{R}$ (l'ensemble $\{x :< t, x >\leq \alpha\}$ est ce que l'on appelle un demi-plan dans $\mathbb{R}^k$).*

Preuve. On a $\mu(\{x :< t, x >\leq \alpha\}) = \mu_{h_t}(]-\infty, \alpha])$. Par changement de variables (transfert),

$$\phi_{\mu_{h_t}}(s) = \phi_\mu(st_1, \dots, st_k),$$

et donc la donnée de μ sur les demi-plans détermine ϕ_μ, qui à son tour détermine μ. ■

Théorème de continuité dans $\mathbb{R}^k$. *Soient μ_n, $n \geq 1$ et μ des probabilités sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Alors*

$$\mu_n \Rightarrow \mu \Leftrightarrow \phi_{\mu_n} \to \phi_\mu \text{ simplement.}$$

Preuve. Si $\mu_n \Rightarrow \mu$, alors $\phi_{\mu_n}(t) = \mathbb{E}_{\mu_n}(e^{i<t,x>}) \to \mathbb{E}_\mu(e^{i<t,x>}) = \phi_\mu(t)$, d'après le (i) de la caractérisation de la convergence faible, puisque $x \mapsto e^{i<t,x>}$ est continue bornée.

Réciproquement, si la convergence simple a lieu, alors pour chaque $t \in \mathbb{R}^k$, et chaque $s \in \mathbb{R}$,

$$\phi_{(\mu_n)_{h_t}}(s) \to \phi_{(\mu)_{h_t}}(s),$$

et donc par le théorème de continuité sur la droite, $(\mu_n)_{h_t} \Rightarrow \mu_{h_t}$.

Choisissons $t = e_u := (0, \dots, 0, \overbrace{1}^{u^{\text{ième}}}, 0, \dots, 0)$. La suite $((\mu_n)_{h_{e_u}})_{n\geq 1}$ converge faiblement, et donc est tendue. Soit $\varepsilon > 0$. Il existe alors $a_u < b_u$ tels que

$$\inf_n (\mu_n)_{h_{e_u}}(]a_u, b_u]) \geq 1 - \frac{\varepsilon}{k}.$$

Alors en posant $A = \prod_{u=1}^k]a_u, b_u]$, on a $A^c \subset \cup_{u=1}^k \mathbb{R}^{u-1} \times]a_u, b_u]^c \times \mathbb{R}^{k-u}$, et donc

$$\begin{array}{rcl}\sup_n \mu_n(A^c) & \leq & \sum_{u=1}^k \sup_n \mu_n(\cup_{u=1}^k \mathbb{R}^{u-1} \times]a_u, b_u]^c \times \mathbb{R}^{k-u}) \\ & = & \sum_{u=1}^k \sup_n \big(\mu_n\big)_{h_{e_u}}(]a_u, b_u]^c) < k \times \frac{\varepsilon}{k} = \varepsilon,\end{array}$$

et donc la suite (μ_n) est tendue.

Si une sous-suite $(\mu_{n_i})_{i\geq 1}$ converge faiblement vers ν, alors par hypothèse de convergence simple des fonctions caractéristiques, et par le sens direct déjà traité, on a $\phi_\nu = \lim_n \phi_{\mu_n} = \phi_\mu$. Donc $\nu = \mu$. Et donc toute sous-suite convergente converge vers μ. Donc $\mu_n \Rightarrow \mu$. ■

Au passage, pour pas beaucoup plus cher, citons le

Théorème de Cramér-Wold. *Une suite de vecteurs aléatoires $X_n = (X_{n,1}, \ldots, X_{n,k})$ dans $\mathbb{R}^k$ converge en loi vers le vecteur $X = (X_1, \ldots, X_k)$ si et seulement si pour tout $t = (t_1, \ldots, t_k) \in \mathbb{R}^k$,*

$$\sum_{u=1}^k t_u X_{n,u} \xrightarrow{\mathcal{L}} \sum_{u=1}^k t_u X_u.$$

Preuve. Considérer les applications continues h_t pour la nécessité. Pour la suffisance, chaque convergence en loi sur la droite (t fixé) entraîne la convergence simple des fonctions caractéristiques (théorème de continuité sur la droite), et donc pour chaque $s \in \mathbb{R}$,

$$\mathbb{E}(e^{is(\sum_{u=1}^k t_u X_{n,u})}) \to \mathbb{E}(e^{is(\sum_{u=1}^k t_u X_u)}).$$

Faire $s = 1$ pour en déduire la convergence simple des fonctions caractéristiques dans $\mathbb{R}^k$, et conclure avec la version $\mathbb{R}^k$ du théorème de continuité. ■

8.3. Distributions normales dans $\mathbb{R}^k$.

Par le théorème d'extension de Kolmogorov, il existe un espace de probabilités sur lequel est défini un vecteur aléatoire $X = (X_1, \ldots, X_k)$ tel que les X_u soient indépendantes et suivent chacune la loi normale centrée réduite, i.e. $\mathbb{P}_{X_u}$ a la densité

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

relativement à la mesure de Lebesgue sur la droite.

Par indépendance, pour tout rectangle, $\mathbb{P}_X(A_1 \times \ldots \times A_k) = \prod \mathbb{P}_{X_u}(A_u) = \prod \int_{A_u} f(x_u)dx_u = \int_{\prod A_u} f(x_1)\ldots f(x_k)dx_1 \ldots dx_k$, et donc X a la densité

$$f_k(x) = \left(\frac{1}{2\pi}\right)^{\frac{k}{2}} e^{-\frac{|x|_2^2}{2}},$$

relativement à la mesure de Lebesgue m_k sur $\mathbb{R}^k$, où $|x|_2^2 = \sum_u x_u^2$. C'est **la densité de la loi normale centrée réduite sur** $\mathbb{R}^k$. Sa fonction caractéristique est

$$\phi_X(t) = \prod_u e^{-\frac{t_u^2}{2}} = e^{-\frac{|t|_2^2}{2}}.$$

Soit $A = (a_{uv})$ une matrice carrée d'ordre k. Posons $Y = AX$, en notant X sous la forme d'un vecteur colonne. Comme $\mathbb{E}(X_\alpha X_\beta) = \mathbf{1}_{\alpha=\beta}$, la matrice $\Sigma = (\sigma_{uv})$ des covariances de Y vérifie

$$\sigma_{uv} = \mathbb{E}(Y_u Y_v) = \sum_\alpha a_{u\alpha} a_{v\alpha},$$

et donc $\Sigma = \mathbf{AA'}$, où A' désigne la transposée de A. Et donc Σ est **symétrique définie positive** : $\sum_{u,v} \sigma_{uv} x_u x_v = |A'x|_2^2 \geq 0$.

Notons t un vecteur colonne de transposé t' (ligne) : notons $< t, x >= t'x$. Alors la fonction caractéristique de AX est

$$\phi_{AX}(t) = \mathbb{E}(e^{it'(AX)}) = \mathbb{E}(e^{i(A't)'X}) = \phi_X(A't) = e^{-\frac{|A't|_2^2}{2}} = e^{-\frac{t'\Sigma t}{2}}.$$

Nous définissons un vecteur centré Gaussien comme un vecteur dont la fonction caractéristique est de la forme ci-dessus pour une matrice Σ symétrique définie positive.

Si Σ est telle, il existe une matrice orthogonale U et une matrice diagonale D telles que

$$U'\Sigma U = D,$$

où les éléments diagonaux de D sont les valeurs propres de Σ, et donc sont positifs ou nuls. Si D_0 est la matrice diagonale dont les éléments diagonaux sont les racines carrées de ceux de D, et si $A = UD_0$, alors $\Sigma = AA'$. Et donc **pour toute matrice symétrique définie positive Σ, il existe un vecteur Gaussien centré $Y = AX$ de matrice de covariances Σ et de fonction caractéristique** $e^{-\frac{t'\Sigma t}{2}}$.

Si Σ est non-singulière, il en est de même de la matrice A construite ci-dessus. Puisque X a une densité $f(x)$, il en est alors de même de $Y = AX$, d'après la formule du Jacobien pour le changement de variables dans les intégrales multiples, et sa densité est $f(A^{-1}x)|detA|^{-1}$. Puisque $\Sigma = AA'$, $|detA| = \sqrt{det\Sigma}$. De plus, $\Sigma^{-1} = (A')^{-1}A^{-1}$, et donc $|A^{-1}x|_2^2 = x'\Sigma^{-1}x$.

Donc **si Σ est non-singulière, la loi normale associée a pour densité**

$$f_{AX}(x) = \frac{1}{(2\pi)^{\frac{k}{2}}\sqrt{det\Sigma}} e^{-\frac{x'\Sigma x}{2}}.$$

Si Σ est singulière, A aussi, et $Y = AX$ est portée par un hyper-plan de $\mathbb{R}^k$. Donc elle est singulière à la mesure de Lebesgue sur $\mathbb{R}^k$.

Un vecteur Gaussien centré est entièrement déterminé par sa matrice de covariances Σ, puisque celle-ci détermine sa fonction caractéristique. Si ses composantes sont indépendantes, Σ est diagonale.

Réciproquement, si Σ est diagonale, et le vecteur Gaussien, alors la matrice A associée à Σ est diagonale, et ses éléments diagonaux sont les racines carrées de ceux de Σ. Alors AX, dont la loi est celle du vecteur Gaussien de matrice de covariances Σ, a ses composantes indépendantes.

Donc **les composantes d'un vecteur Gaussien sont indépendantes si et seulement si elles ne sont pas corrélées**.

Soit M une matrice de dimensions $j \times k$ et soit Y un vecteur Gaussien de $\mathbb{R}^k$ de matrice de covariance Σ. Alors la fonction caractéristique du vecteur aléatoire dans $\mathbb{R}^j$ MY, est

$$\phi_{MY}(t) = e^{-\frac{1}{2}t'(M\Sigma M')t}.$$

Donc MY **suit une loi Gaussienne centrée dans $\mathbb{R}^j$ de matrice de covariances** $M\Sigma M'$.

Enfin, il existe des vecteurs Gaussiens non -centrés : ce sont les translatés des centrés. Et donc en général, **un vecteur aléatoire, s'il est Gaussien, est entièrement déterminé par son vecteur d'espérance, et sa matrice de covariances**.

8.4. Le TCL dans $\mathbb{R}^k$.

TCL DANS $\mathbb{R}^k$. *Soit $(X_n)_{n\geq 1} = ((X_{n1}, \ldots, X_{nk}))_{n\geq 1}$ une suite i.i.d. de vecteurs aléatoires dans $\mathbb{R}^k$, tels que* $\max_u \mathbb{E}(X_{nu}^2) < +\infty$.

Notons $c = (c_1, \ldots, c_k) = (\mathbb{E}(X_{n1}), \ldots, \mathbb{E}(X_{nk}))$ le vecteur d'espérances de la loi commune, et soit $\Sigma = (\sigma_{uv})$ la matrice de covariances associée $(\sigma_{uv} = \mathbb{E}\big((X_{nu}) - c_u)(\mathbb{E}(X_{nv}) - c_v)\big))$.

Alors

$$\frac{X_1 + \ldots + X_n - nc}{\sqrt{n}} \xrightarrow{\mathcal{L}} X(c, \Sigma),$$

où $X(c, \Sigma)$ désigne un vecteur Gaussien de $\mathbb{R}^k$ d'espérances c et de covariances Σ.

PREUVE. Soit Y un vecteur Gaussien centré de covariances Σ. Soit $t = (t_1, \ldots, t_k)$ et notons $Z_n = \sum_{1\leq u\leq k} t_u(X_{nu} - c_u)$, et $Z = \sum_{1\leq u\leq k} t_u Y_u$.

Par Le Théorème de Cramér-Wold, il suffit de montrer que $\frac{Z_1+\ldots+Z_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} Y$, ce qui découle directement des hypothèses et du TCL sur la droite. ■

Chapitre 9
Grandes déviations

9.1. Grandes déviations - énoncé.

Soit $(X_i)_{i\geq 1}$ une suite de v.a.r. i.i.d. intégrables. Supposons que $\mathbb{E}(X_1) < 0$. Alors la loi forte nous dit que si $Y_n = X_1 + \ldots + X_n$, $\frac{1}{n}Y_N \to \mathbb{E}(X_1) < 0$ $\mathbb{P}$-p.s., et donc que $\mathbb{P}(Y_n \geq 0) \to 0$, puisque la convergence presque sûre entraîne la convergence en probabilités.

Théorème de Chernoff (Grandes déviations). *Sous les hypothèses précédentes, supposons en plus que X_1 soit simple, et que $\mathbb{P}(X_1 > 0) > 0$.*

Notons $M : t \mapsto \mathbb{E}(e^{tX_1})$ la fonction génératrice des moments de X_1. Alors M est convexe, $\mathcal{C}^\infty$, et $\inf_t M(t) = M(\tau) = \rho$ pour un certain $0 < \tau$ avec $0 < \rho < 1$.

Alors

$$\frac{1}{n}\ln\left(\mathbb{P}(Y_n \geq 0)\right) = \ln\rho + O(\frac{1}{\sqrt{n}}).$$

Nous développons la preuve de ce résultat dans les deux sections suivantes.

9.2. Préliminaires.

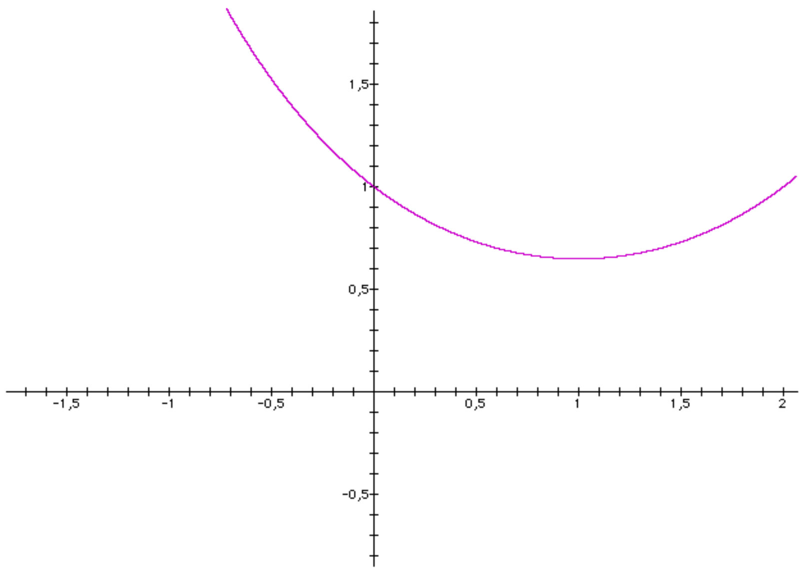
Soit Y une v.a.r. simple, prenant un nombre fini de valeurs y_j, avec les probabilités $p_j = \mathbb{P}(Y = y_j)$, telle que $\mathbb{E}(Y) < 0$ et $\mathbb{P}(Y > 0) > 0$. Rappelons que $\mathbb{E}(Y) = \sum_j p_j y_j$.

Notons $M_Y(t) = \mathbb{E}(e^{tY}) = \sum_j p_j e^{ty_j}$. La condition $\mathbb{P}(Y > 0) > 0$ signifie exactement qu'il existe un j_0 au moins tel que $y_{j_0} > 0$ et $p_{j_0} > 0$, puisque $\mathbb{P}(Y > 0) = \sum_j p_j \mathbf{1}_{]0,+\infty[}(y_j)$.

Alors $M_Y : t \mapsto \sum_j p_j e^{ty_j}$ est une combinaison linéaire à coefficients positifs des fonctions $t \mapsto e^{ty_j}$. Elle est clairement $\mathcal{C}^\infty$, et

$$\begin{cases} M_Y(0) = \mathbb{E}(1) = 1; \\ M_Y'(t) = \sum_j p_j y_j e^{ty_j} \Rightarrow M_Y'(0) = \mathbb{E}(Y) < 0; \\ M_Y''(t) = \sum_j p_j y_j^2 e^{ty_j} > 0 \Rightarrow M_Y \text{ convexe et } M_Y''(O) = \mathbb{E}(Y^2) \geq y_{j_0}^2 p_{j_0} > 0; \\ M_Y^{(k)}(t) = \sum_j p_j y_j^k e^{ty_j} \Rightarrow \mathbb{E}(Y^k) = M_Y^{(k)}(0); \\ \lim_{+\infty} M_Y(t) \geq \lim_{+\infty} p_{j_0} e^{ty_{j_0}} = +\infty \Rightarrow \lim_{+\infty} M_Y(t) = +\infty. \end{cases}$$

Voici l'allure de $t \mapsto M_Y(t)$:



Donc il existe $\tau > 0$ tel que $M_Y(\tau) = \rho = \inf_t M_Y(t)$ et $0 < \rho < 1$.

Si $t \geq 0$, $Y \geq 0 \Leftrightarrow e^{tY} \geq 1$, et donc si $t \geq 0$,

$$\mathbb{P}(Y \geq 0) = \mathbb{P}(e^{tY} \geq 1) = \int_{\mathbb{P}(Y\geq 0)} \min(\mathbf{1}, e^{tY}) d\mathbb{P} \leq \int_\Omega e^{tY} d\mathbb{P} = M_Y(t),$$

et donc $\mathbb{P}(Y \geq 0) \leq M_Y(\tau) = \rho$.

Posons $\mathbb{P}(Y \geq 0) = \rho e^{-\theta}$, puisque $\rho > 0$ et $\mathbb{P}(Y \geq 0) \geq \mathbb{P}(Y > 0) > 0$. Alors $\mathbb{P}(Y \geq 0) \leq \rho \Rightarrow \theta \geq 0$.

Une variable auxilliaire : soit Z une v.a.r. simple prenant les mêmes valeurs que Y mais avec les distributions suivantes :

$$\mathbb{P}(Z = y_j) = \frac{e^{\tau y_j}}{\rho} \mathbb{P}(Y = y_j) = \frac{e^{\tau y_j}}{\rho} p_j.$$

On vérifie que $\sum_j \mathbb{P}(Z = y_j) = M_Y(\tau) \ / \ \rho = 1$! On a en outre

$$\begin{cases} M_Z(t) := \mathbb{E}(e^{tZ}) = \sum_j e^{ty_j} e^{\tau y_j} \frac{p_j}{\rho} = \frac{M_Y(t+\tau)}{\rho}; \\ \mathbb{E}(Z) = M_Z'(0) = \frac{M_Y'(\tau)}{\rho} = 0 \text{ car } M_Y \text{ est extrémale en } \tau; \\ s^2 := \mathbb{E}(Z^2) = M_Z''(0) = \frac{M_Y''(\tau)}{\rho} = \sum_j \frac{p_j}{\rho} y_j^2 e^{\tau y_j} \geq \frac{p_{j_0}}{\rho} y_{j_0}^2 > 0. \end{cases}$$

Nous avons majoré $\mathbb{P}(Y \geq 0) = \rho e^{-\theta}$. Nous allons maintenant minorer, en utilisant Z. Puis nous minorerons $\mathbb{P}(Z \geq 0)$ que nous verrons intervenir plus loin.

Notons Σ' la somme sur les indices j tels que $y_j \geq 0$. Alors $\mathbb{P}(Y \geq 0) = \Sigma' p_j$, ce qui, compte tenu de la définition de Z, peut s'écrire sous la forme

$$\mathbb{P}(Y \geq 0) = \rho(\Sigma' e^{-\tau y_j} \mathbb{P}(Z = y_j)) = \rho e^{-\theta}.$$

Notons $p = \mathbb{P}(Z \geq 0)$ $(p > 0)$. Alors $\Sigma' \frac{\mathbb{P}(Z=y_j)}{p} = 1$, et par concavité du "ln", il s'en suit que

$$\begin{aligned} -\theta &= \ln(\Sigma' e^{-\tau y_j} \tfrac{\mathbb{P}(Z=y_j)}{p}) + \ln p \\ &\geq \left(\Sigma' \tfrac{\mathbb{P}(Z=y_j)}{p}(-\tau y_j)\right) + \ln p \\ &= \ln p - \tfrac{\tau s}{p}\left(\Sigma' \tfrac{y_j}{s} \mathbb{P}(Z = y_j)\right). \end{aligned}$$

Par Cauchy-Schwarz,

$$\left(\Sigma' \frac{y_j}{s} \mathbb{P}(Z = y_j)\right) = \frac{1}{s}\mathbb{E}(Z\mathbf{1}_{Z\geq 0}) \leq \frac{1}{s}\sqrt{\mathbb{E}(Z^2)}\sqrt{\mathbb{E}(\mathbf{1})} = 1,$$

et donc on obtient

$$0 \leq \theta \leq \frac{\tau s}{\mathbb{P}(Z \geq 0)} - \ln \mathbb{P}(Z \geq 0). \qquad (\star)$$

Maintenant nous allons minorer $\mathbb{P}(Z \geq 0)$ *ce qui donnera une majoration du terme de droite dans* $(\star)$. Pour cela, utilisons le

Lemme. *Soit* W *une v.a.r. centrée dans* $\mathcal{L}^4$. *Notons* $l^2 = \mathbb{E}(W^2)$ *et supposons* $\xi^4 := \mathbb{E}(W^4) > 0$. *Alors*

$$\mathbb{P}(W \geq 0) \geq \frac{l^4}{4\xi^4}.$$

Preuve. Notons $q = \mathbb{P}(W \geq 0)$. Ecrivons $W^+ = W\mathbf{1}_{W\geq 0}$ et $W^- = -W\mathbf{1}_{W<0}$. Alors $W = W^+ - W^-$, et $l^2 = \mathbb{E}((W^+)^2) + \mathbb{E}(W^-)^2)$. Par Cauchy-Schwarz,

$$\mathbb{E}((W^+)^2) = \mathbb{E}(W^2\mathbf{1}^2_{W\geq 0}) \leq \sqrt{\mathbb{E}(W^4)}\sqrt{\mathbb{E}(\mathbf{1}^4_{W\geq 0})} = \sqrt{q}\xi^2.$$

Ensuite avec l'inégalité de Hölder d'exposants $(\frac{3}{2}, 3)$, nous avons

$$\mathbb{E}((W^-)^2) = \mathbb{E}((W^-)^{\frac{2}{3}}(W^-)^{\frac{4}{3}}) \leq \mathbb{E}(W^-)^{\frac{2}{3}}\mathbb{E}((W^-)^4)^{\frac{1}{3}} \leq \mathbb{E}(W^-)^{\frac{2}{3}}\xi^{\frac{4}{3}}.$$

Ensuite $\mathbb{E}(W) = 0$ donc $\mathbb{E}(W^-) = \mathbb{E}(W^+)$, et Hölder d'exposants $(\frac{4}{3}, 4)$ permet d'écrire

$$\mathbb{E}(W^-) = \mathbb{E}(W\mathbf{1}_{W\geq 0}) \leq \mathbb{E}(W^4)^{\frac{1}{4}}\mathbb{E}(\mathbf{1}^{\frac{4}{3}}_{W\geq 0})^{\frac{3}{4}} = \xi q^{\frac{3}{4}}.$$

Et donc

$$\mathbb{E}((W^-)^2) \leq \mathbb{E}(W^-)^{\frac{2}{3}}\xi^{\frac{4}{3}} \leq \sqrt{q}\xi^2,$$

et donc

$$l^2 = \mathbb{E}(W^-)^2) + \mathbb{E}((W^+)^2) \leq 2\sqrt{q}\xi^2.$$

■

Du Lemme et de $(\star)$ il découle que $(s = l = \mathbb{E}(Z^2)$ et $\xi^4 = \mathbb{E}(Z^4))$

$$\mathbb{P}(Y \geq 0) = \rho e^{-\theta} \text{ et } 0 \leq \theta \leq \frac{4\tau\xi^4}{s^3} - \ln\left(\frac{s^4}{4\xi^4}\right). \qquad (\star\star)$$

9.3. Fin de la preuve du Théorème de Chernoff.

Posons donc $Y_n = X_1 + \ldots + X_n$ et $M(t) = M_{X_1}(t)$. Alors Y_n est simple, $\mathbb{E}(Y_n) = n\mathbb{E}(X_1) < 0$, et $\mathbb{P}(Y_n > 0) \geq \mathbb{P}(\cap_{i=1}^n \{X_i > 0\}) = \mathbb{P}(X_1 > 0)^n > 0$, par indépendance et distribution identique. Par indépendance et distribution identique de la suite $(X_n)_{n\geq 0}$ encore, il vient

$$M_n(t) := M_{Y_n}(t) = M(t)^n, \ n \geq 1.$$

Soient $\tau_n > 0$ et $\rho_n \in]0, 1[$ les nombres tels qu'associés à Y en **9.2.**, mais ici à Y_n - qui satisfait aux hypothèses qu'il faut. Alors comme $M_n(t) = M(t)^n \geq M(\tau)^n = M_n(\tau) = \rho^n$, il vient

$$\tau_n = \tau \text{ et } \rho_n = \rho^n.$$

Notons ensuite S_n la variable correspondant à la variable Z associée à Y, mais ici à Y_n. Alors

$$M_{S_n}(t) = \frac{M_n(t + \tau_n)}{\rho_n} = \left(\frac{M(t+\tau)}{\rho}\right)^n,$$

par ce qui précède.

Puis notons Z_j la variable telle que Z, mais associée ici à X_j - qui satisfait aux hypothèses requises. Par le théorème d'extension de Kolmogorov, puisqu'aucun espace n'est désigné à priori pour supporter la variable Z, nous pouvons supposer que la suite $(Z_j)_{j\geq 1}$ est indépendante.

De fait elle sera identiquement distribuée puisque la suite $(X_j)_{j\geq 1}$ l'est, et que la loi de Z_j est entièrement déterminée par celle de X_j.

Notons donc $V_n = Z_1 + \ldots + Z_n$, pour voir. Alors par indépendance et distribution identique, et construction des Z_j, il vient que V_n est simple et que

$$M_{V_n}(t) = \mathbb{E}(e^{t(Z_1+\ldots+Z_n)}) = \prod_{i=1}^n M_{Z_i}(t) = M_{Z_1}(t)^n = \left(\frac{M(t+\tau)}{\rho}\right)^n = M_{S_n}(t) \ !!$$

Quel est l'intérêt ? Calculer les moments de S_n pour appliquer $(\star\star)$ à Y_n :

$$\begin{aligned}
&\mathbb{E}(S_n) = M'_{S_n}(0) = M'_{V_n}(0) = \mathbb{E}(V_n) = n\mathbb{E}(Z_1) = 0;\\
&\mathbb{E}(S_n^2) = M''_{S_n}(0) = M''_{V_n}(0) = \mathbb{E}(V_n^2) = n\mathbb{E}(Z_1^2) =: n\sigma^2;\\
&\mathbb{E}(S_n^4) = M^{(4)}_{S_n}(0) = M^{(4)}_{V_n}(0) = \mathbb{E}(V_n^4) = n\mathbb{E}(Z_1^4) + 3n(n-1)\sigma^4.
\end{aligned}$$

Donc si nous notons $l_n^2 = \mathbb{E}(S_n^2)$ et $\xi_n^4 = \mathbb{E}(S_n^4)$, il vient $l_n^4 \ / \ (4\xi_n^4) \to 1 \ / \ 12 > 0$, et donc il existe un $\alpha > 0$ tel que pour chaque $n \geq 1$, $l_n^4 \ / \ (4\xi_n^4) \geq \alpha > 0$. Donc nous pouvons encore majorer uniformément en n pour Y_n dans $(\star\star)$: en effet, $\mathbb{P}(Y_n \geq 0) = \rho^n e^{-\theta_n}$ pour θ_n satisfaisant $(\star\star)$ pour Y_n, et alors on a

$$\mathbb{P}(Y_n \geq 0) = \rho^n e^{-\theta_n} \text{ et } 0 \leq \theta_n \leq \frac{\tau\sigma\sqrt{n}}{\alpha} - \ln\alpha.$$

En prenant la formule ci-dessus, en passant au “ln”, et en divisant par n, il vient

$$\frac{1}{n}\mathbb{P}(Y_n \geq 0) - \ln \rho = \frac{\theta_n}{n} = O\left(\frac{1}{\sqrt{n}}\right).$$

■

9.4. Application au pile ou face équilibré.
Supposons que la suite $(X_i)_{i\geq 1}$ soit i.i.d. de loi commune la loi telle que

$$\mathbb{P}(X_1 = \pm 1) = \frac{1}{2}.$$

Si $Y_n = X_1 + \ldots + X_n$, on a $\frac{Y_n}{n} \to 0$ $\mathbb{P}$-p.s., et donc en probabilité. Soit $0 < \beta$: alors $\mathbb{P}(Y_n \geq n\beta)$ mesure la “*déviation*” de Y_n relativement à son espérance, qui est nulle.

Nous pouvons ici appliquer le théorème de Chernoff pour avoir une idée de la taille de cette déviation, asymptotiquement.

En effet, si $\beta > 1$, $\{Y_n \geq n\beta\} = \emptyset$, et la taille est claire : $\mathbb{P}(Y_n \geq n\beta) = 0$. Si $\beta = 1$, alors $\{Y_n = 1\} = \cap_{i=1}^n \{X_i = 1\}$ et donc $\mathbb{P}(Y_n \geq n) = \frac{1}{2^n}$, et la taille est bien connue.

Ensuite si $0 < \beta < 1$, alors la suite $(X_i - \beta)_{i\geq 1}$ satisfait aux hypothèses du théorème de Chernoff, et alors

$$\frac{1}{n}\ln \mathbb{P}(Y_n \geq n\beta) = \ln \rho + O\left(\frac{1}{\sqrt{n}}\right),$$

et il n'y a plus qu'à calculer $\rho = \min_t M_{X_1-\beta}(t)$. Ceci n'est pas compliqué : $M_{X_1-\beta}(t) = \frac{e^{-t\beta}}{2}(e^t + e^{-t}) = e^{-t\beta}\mathrm{ch}t$.

Pour avoir ρ, on trouve $\tau > 0$ tel que $M'_{X_1-\beta}(\tau) = 0$ (alors ρ vaudra $M_{X_1-\beta}(\tau)$) : soit

$$-\beta e^{-t\beta}\mathrm{ch}(t) + e^{-t\beta}\mathrm{sh}t = 0,$$

soit $\mathrm{th}t = \beta$, et donc $\tau = \mathrm{Argth}\beta = \frac{1}{2}\ln\left(\frac{1+\beta}{1-\beta}\right)$.

Alors un calcul pas trop difficile montre que $\rho = M_{X_1-\beta}(\tau) = \frac{(1-\beta)^{\frac{(\beta-1)}{2}}}{(1+\beta)^{\frac{(\beta+1)}{2}}}$. Et donc

$$\frac{1}{n}\ln \mathbb{P}(Y_n \geq n\beta) \to -\frac{1}{2}\Big((1+\beta)\ln(1+\beta) + (1-\beta)\ln(1-\beta)\Big).$$

Sujet de reflexion : qu'en est-il du théorème de Chernoff si $X_i = \tilde{X}_i - \gamma$, avec X_i qui suit une loi de Poisson de paramètre $\lambda > 0$, et en supposant $\gamma > \lambda$? Est-ce que la preuve du théorème s'adapte ?

Espérances et lois conditionnelles

Nous avons vu rapidement au Chapitre 3 les probabilités conditionnelles “naïves”. Nous verrons ici une version bien plus élaborée, qui l'étend, et qui est nécessaire au traitement de la théorie des martingales. Tout ce qui suit repose sur le théorème de Radon-Nikodym.

10.1. Espérance conditionnelle.

Théorème - Définition. *Soit X une variable aléatoire réelle intégrable sur $(\Omega, \mathcal{B}, \mathbb{P})$. Soit $\mathcal{C} \subset \mathcal{B}$ une σ-algèbre. Alors il existe une et une seule variable aléatoire réelle Y sur $(\Omega, \mathcal{B}, \mathbb{P})$ possédant les deux propriétés suivantes :*

$$\begin{cases} (i) & : \quad Y \text{ est } \mathcal{C}\text{-mesurable et intégrable;} \\ (ii) & : \quad \text{quel que soit } C \in \mathcal{C},\ \int_C Y d\mathbb{P} = \int_C X d\mathbb{P}. \end{cases}$$

On appelle Y l'espérance conditionnelle de X sachant $\mathcal{C}$, et on la note

$$\mathbb{E}(X|\mathcal{C}).$$

Preuve. Supposons d'abord $X \geq 0$. Définissons alors une mesure ν sur $\mathcal{C}$ en posant $\nu(C) = \int_C X d\mathbb{P}$. C'est une mesure finie positive, du fait de la positivité et de l'intégrabilité de X.

De plus ν est absolument continue relativement à $\mathbb{P}$. Donc par le théorème de Radon-Nykodim, il existe une unique fonction $\mathcal{C}$-mesurable f telle que

$$\nu(C) = \int_C f d\mathbb{P}, \qquad C \in \mathcal{C}.$$

La fonction f satisfait donc aux conditions annoncées $(i) - (ii)$. Donc voilà pour l'existence.

De l'unicité, supposons que f et g satisfassent $(i) - (ii)$. Alors $f - g$ est $\mathcal{C}$-mesurable et quel que soit $C \in \mathcal{C}$,

$$\int_C (f - g) d\mathbb{P} = 0.$$

Il suffit ensuite d'appliquer le Lemme Préliminaire 2 page 13, avec $\Delta = \{0\}$, pour en déduire que $f - g = 0$ $\mathbb{P}$-p.s..

Pour conclure lorsque X n'est pas $\mathbb{P}$-p.s. positive, on écrit $X^+ = \max\{0, X\}$, $X^- = \min\{0, X\}$, et on obtient facilement que

$$\mathbb{E}(X|\mathcal{C}) = \mathbb{E}(X^+|\mathcal{C}) + \mathbb{E}(X^-|\mathcal{C}).$$

∎

DÉFINITION. *Etant données deux variables aléatoires réelles X, Y sur $(\Omega, \mathcal{B}, \mathbb{P})$, avec X intégrable, on appelle espérance conditionnelle de X sachant Y l'espérance conditionnelle de X sachant la tribu engendrée par Y. On la note $\mathbb{E}(X|Y)$. Si nous notons cette tribu $\mathcal{B}(Y)$, il s'ensuit donc que*

$$\mathbb{E}(X|Y) = \mathbb{E}(X|\mathcal{B}(Y)).$$

Illustrons ces définitions par un exemple important :
La cas discret : on suppose ici que $\mathcal{C} = \sigma((B_i)_{i\in I})$, où $(B_i)_{i\in I}$ est une collection finie ou dénombrable d'ensembles mesurables, et qui partitionne Ω. Alors si X est intégrable, on vérifie (sans peine ?) que

$$\mathbb{E}(X|\mathcal{C}) = \sum_{i\in I} \frac{\int_{B_i} X d\mathbb{P}}{\mathbb{P}(B_i)} \mathbf{1}_{B_i}.$$

PROPRIÉTÉS DES ESPÉRANCES CONDITIONNELLES. *Supposons que $\mathcal{P}$ soit un π-système générateur de $\mathcal{C}$, et que Ω soit une union finie ou dénombrable d'éléments de $\mathcal{P}$. Alors une fonction $f : \Omega \to \mathbb{R}$ $\mathcal{C}$-mesurable est une version de $\mathbb{E}(X|\mathcal{C})$ si pour chaque $P \in \mathcal{P}$,*

$$\int_P f d\mathbb{P} = \int_P X d\mathbb{P}.$$

Supposons maintenant que X, Y, Y_n sont intégrables. Alors

$$\begin{array}{rcl}
(i) & : & si\ X = a\ \mathbb{P}\text{-}p.s.,\ \mathbb{E}(X|\mathcal{C}) = a\ \mathbb{P}\text{-}p.s.; \\
(ii) & : & \mathbb{E}(aX + bY|\mathcal{C}) = a\mathbb{E}(X|\mathcal{C}) + b\mathbb{E}(Y|\mathcal{C}); \\
(iii) & : & si\ X \le Y\ \mathbb{P}\text{-}p.s.,\ alors\ \mathbb{E}(X|\mathcal{C}) \le \mathbb{E}(Y|\mathcal{C}); \\
(iv) & : & |\mathbb{E}(X|\mathcal{C})| \le \mathbb{E}(|X||\mathcal{C}); \\
(v) & : & si\ \lim_n Y_n = Y\ \mathbb{P}\text{-}p.s.,\ et\ si\ \forall n,\ |Y_n| \le X, \\
 & & alors\ \lim_n \mathbb{E}(Y_n|\mathcal{C}) = \mathbb{E}(Y|\mathcal{C})\ \mathbb{P}\text{-}p.s..
\end{array}$$

PREUVE. Pour $(i-iv)$, il suffit de vérifier les conditions caractéristiques des espérances conditionnelles relativement à $\mathcal{C}$, sauf pour (iii) où un raisonnement par l'absurde fait l'affaire.

Pour (v), poser $Z_n = \sup_{k\ge n} |Y_k - Y|$. Alors $Z_n \downarrow 0$ $\mathbb{P}$-p.s.. Et $|\mathbb{E}(Y_n|\mathcal{C}) - \mathbb{E}(Y|\mathcal{C})| \le \mathbb{E}(Z_n|\mathcal{C})$. Il suffit donc de montrer que $\mathbb{E}(Z_n|\mathcal{C}) \downarrow 0$ $\mathbb{P}$-p.s.. Par décroissance, il existe Z, $\mathcal{C}$-mesurable, intégrable, telle que $Z = \lim_n \downarrow \mathbb{E}(Z_n|\mathcal{C})$. En outre $Z \ge 0$. Il nous suffit donc de montrer que $\mathbb{E}(Z) = 0$. Or

$$\mathbb{E}(Z) = \int \lim_n \downarrow \mathbb{E}(Z_n|\mathcal{C}) d\mathbb{P} \le \liminf_n \int \mathbb{E}(Z_n|\mathcal{C}) d\mathbb{P} = \liminf_n \mathbb{E}(Z_n) = 0,$$

par convergence dominée. ■

Proposition. *Si X est $\mathcal{C}$-mesurable, et si Y et XY sont intégrables, alors*

$$\mathbb{E}(XY|\mathcal{C}) = X\mathbb{E}(Y|\mathcal{C}).$$

Preuve. Il suffit de considérer l'application $\mathcal{C}$-mesurable $X\mathbb{E}(Y|\mathcal{C})$ et de vérifier qu'elle vérifie les conditions caractéristiques (i) et (ii) de l'espérance conditionnelle.

Pour cela traiter d'abord de $X = \mathbf{1}_C$ avec $C \in \mathcal{C}$, puis de X simple $\mathcal{C}$-mesurable, puis de $X \geq 0$ etc etc... ■

Proposition. *Supposons que deux sous-σ-algèbres de $\mathcal{B}$ satisfassent $\mathcal{C}_1 \subset \mathcal{C}_2$. Alors si X est intégrable,*

$$\mathbb{E}(\mathbb{E}(X|\mathcal{C}_2)|\mathcal{C}_1) = \mathbb{E}(X|\mathcal{C}_1).$$

Preuve. $\mathbb{E}(X|\mathcal{C}_1)$ est $\mathcal{C}_1$-mesurable donc $\mathcal{C}_2$-mesurable. Et pour chaque $C \in \mathcal{C}_1$, $C \in \mathcal{C}_2$, donc

$$\int_C \mathbb{E}(X|\mathcal{C})d\mathbb{P} = \int_C X d\mathbb{P} = \int_C \mathbb{E}(X|\mathcal{C}_2)d\mathbb{P} = \int_C \mathbb{E}(\mathbb{E}(X|\mathcal{C}_2)|\mathcal{C}_1)d\mathbb{P}.$$

■

10.2. Probabilités ou lois conditionnelles.

Avec les mêmes notations, appelons probabilité de A sachant $\mathcal{C}$ la quantité notée $\mathbb{P}(A|\mathcal{C})$ et définie par

$$\mathbb{P}(A|\mathcal{C}) = \mathbb{E}(\mathbf{1}_A|\mathcal{C}),\ A \in \mathcal{B}.$$

Probabilités conditionnelles et π-systèmes générateurs. *Soit $\mathcal{P}$ un π-système générateur de $\mathcal{C}$, et supposons que Ω soit une union finie ou dénombrable d'éléments de $\mathcal{P}$. Soit enfin $A \in \mathcal{B}$. Alors une fonction mesurable $f : \Omega \to \mathbb{R}$ est une version de $\mathbb{P}(A|\mathcal{C})$ si elle est $\mathcal{C}$-mesurable, intégrable, et si pour chaque $P \in \mathcal{P}$,*

$$\int_P f d\mathbb{P} = \mathbb{P}(A \cap P).$$

Preuve. Paragraphe précédent avec $X = \mathbf{1}_A$. ■

Propriétés de base des probabilités conditionnelles. *Presque sûrement,*

$$\mathbb{P}(\emptyset|\mathcal{C}) = 0,\ \mathbb{P}(\Omega|\mathcal{C}) = 1,\ 0 \leq \mathbb{P}(A|\mathcal{C}) \leq 1,$$

et enfin si $(A_n)_{n\geq 1}$ *est une collection disjointe d'ensembles mesurables,*

$$\mathbb{P}(\cup_n A_n|\mathcal{C}) = \sum_n \mathbb{P}(A_n|\mathcal{C}).$$

On a aussi les propriétés suivantes ($\mathbb{P}$*-p.s.):*

$$\begin{cases} \mathbb{P}(B \setminus A|\mathcal{C}) = \mathbb{P}(B|\mathcal{C}) - \mathbb{P}(A|\mathcal{C}) \text{ si } A \subseteq B; \\ \mathbb{P}(\cup_n \uparrow A_n|\mathcal{C}) = \lim_n \uparrow \mathbb{P}(A_n|\mathcal{C}); \\ \mathbb{P}(\cap_n \downarrow A_n|\mathcal{C}) = \lim_n \downarrow \mathbb{P}(A_n|\mathcal{C}); \\ \mathbb{P}(A\Delta B) = 0 \Rightarrow \mathbb{P}(A|\mathcal{C}) = \mathbb{P}(B|\mathcal{C}). \end{cases}$$

Preuve. Découle des propriétés correspondantes pour les espérances conditionnelles. ■

10.3. Le lien - plus fin - entre probabilités et espérances conditionnelles.

Version uniforme de la loi conditionnelle. *Il existe une application* $\mu : \mathcal{B}(\mathbb{R})\times \Omega \to [0,1]$ *telle que*

$$\begin{cases} (i) & : \quad \text{pour chaque } \omega \in \Omega,\ \mu(\cdot,\omega) \text{ est une probabilité sur } (\mathbb{R},\mathcal{B}(\mathbb{R})); \\ (ii) & : \quad \text{pour chaque } H \in \mathcal{B}(\mathbb{R}),\ \mu(H,\cdot) \text{ est une version de } \mathbb{P}(X \in H|\mathcal{C}) =: \mathbb{P}_X(H|\mathcal{C}). \end{cases}$$

On appelle $\mu(\cdot,\omega)$ *la loi conditionnelle de* X *sachant* $\mathcal{C}$*, ou sachant* Y *si dans le contexte* $\sigma(Y) = \mathcal{C}$*. Nous la noterons parfois* $\mathbb{P}_X(\cdot|\mathcal{C})$ *ou* $\mathbb{P}_X(\cdot|Y)$*.*

Preuve. Pour chaque rationnel r, notons $F(r,\omega)$ une version de $\mathbb{P}(X \leq r|\mathcal{C})(\omega)$. Si $r \leq s$, alors

$$F(r,\omega) \leq F(s,\omega)$$

sauf peut-être pour $\omega \in A_{r,s}$ avec $A_{r,s} \in \mathcal{C}$ et $\mathbb{P}(A_{r,s}) = 0$. Aussi,

$$F(r,\omega) = \lim_n F(r + \frac{1}{n},\omega),$$

sauf peut-être pour $\omega \in B_r$ avec $B_r \in \mathcal{C}$ et $\mathbb{P}(B_r) = 0$. Enfin, sauf peut-être pour $\omega \in C$ avec $C \in \mathcal{C}$ et $\mathbb{P}(C) = 0$,

$$\lim_{-\infty} F(r,\omega) = 0 \qquad \text{et} \qquad \lim_{\infty} F(r,\omega) = 1.$$

Alors si $E = \cup_{r,s\in\mathbb{Q}}(C \cup B_r \cup A_{r,s})$, on a $E \in \mathcal{C}$ et $\mathbb{P}(E) = 0$. Si $\omega \notin E$, on étend $F(\cdot, \omega)$ à $\mathbb{R}$ en posant

$$F(x, \omega) = \inf\{F(r, \omega) : x < r\}.$$

Et si $\omega \in E$, on pose $F(x, \omega) = F_0(x)$ où F_0 est une fonction de répartition, arbitrairement choisie.

Alors pour tout ω, $F(\cdot, \omega)$ est une fonction de répartition, d'une probabilité borélienne $\mu(\cdot, \omega)$. Et ainsi on a (i).

Pour (ii), remarquons d'abord que la classe $\mathcal{H}$ des $H \in \mathcal{B}(\mathbb{R})$ pour lesquels $\mu(H, \cdot)$ est $\mathcal{C}$-mesurable est un λ-système contenant les ensembles $]-\infty, r]$, $r \in \mathbb{Q}$. Et donc $\mu(H, \cdot)$ est $\mathcal{C}$-mesurable pour $H \in \sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R})$.

Pour conclure, si $H =]-\infty, r]$, $r \in \mathbb{Q}$, alors, pour $C \in \mathcal{C}$, nous avons

$$\psi_C(H) := \int_C \mu(H, \omega) d\mathbb{P}(\omega) = \mathbb{P}((X \in H) \cap C) =: \gamma_C(H),$$

et donc, ci-dessus, en fixant C, nous avons deux mesures positives finies qui coïncident sur tous les $]-\infty, r]$, $r \in \mathbb{Q}$. Elles sont donc égales ($\psi_C = \gamma_C$), pour chaque $C \in \mathcal{C}$. Donc pour tout $H \in \mathcal{B}(\mathbb{R})$, $\mu(H, \cdot)$ est une version de $\mathbb{P}_X(H|\mathcal{C})$. ■

Exemple : supposons que $\mathbb{P}_{(X,Y)}$ ait une densité f relativement à m_2. Posons $g(x, y) = f(x, y) \;/\; \int_{\mathbb{R}} f(t, y)dt$ si $\int_{\mathbb{R}} f(t, y)dt \neq 0$, et posons alors $\mu(H, y) = \int_H g(x, y)dx$ ($d\mu(\cdot, y) \;/\; dm(x) = g(x, y)$). Si $\int_{\mathbb{R}} f(t, y)dt = 0$, choisissons pour $\mu(\cdot, y)$ une mesure de probabilité borélienne arbitraire sur la droite. Alors $\mu(H, Y(\omega))$ est une version de la loi conditionnelle de X sachant Y : en effet, pour $E, F \in \mathcal{B}(\mathbb{R})$,

$$\int_{\mathbb{R}\times E} \mu(F, Y(\omega)) d\mathbb{P}_{(X,Y)}(\omega) = \mathbb{P}_{(X,Y)}(F \times E),$$

et donc $\int_{Y\in E} \mu(F, Y(\omega)) d\mathbb{P}(\omega) = \mathbb{P}((X, Y) \in F \times E)$. Et donc $\mu(F, Y)$ est une version de $\mathbb{P}(X \in F|Y)$.

Le lien entre l'espérance conditionnelle et la loi conditionnelle est le suivant :

L'espérance conditionnelle est l'intégrale sous la loi conditionnelle. *Si $\phi : \mathbb{R} \to \mathbb{R}$ est borélienne, et si $\phi(X) \in \mathcal{L}^1$, alors*

$$\mathbb{E}(\phi(X)|\mathcal{C}) = \int_{\mathbb{R}} \phi(x) d\mathbb{P}_X(x|\mathcal{C}).$$

Preuve. Si $\phi = \mathbf{1}_A$, alors $\mathbb{E}(\phi(X)|\mathcal{C}) = \mathbb{P}(X \in A|\mathcal{C}) = \int \mathbf{1}_A(x) d\mathbb{P}_X(x|\mathcal{C})$, par construction et définition de $\mathbb{P}_X(A|\mathcal{C})$.

Par linéarité, l'égalité a lieu pour ϕ simple. Si ϕ est positive intégrable, ça passe par convergence monotone et convergence dominée dans les espérances conditionnelles.

Pour ϕ de signe arbitraire, on décompose en parties positives et négatives. ■

COROLLAIRE : INÉGALITÉ DE JENSEN CONDITIONNELLE. *Si ϕ est convexe, et si $\phi(X)$ et X sont intégrables, alors*

$$\phi(\mathbb{E}(X|\mathcal{C})) \leq \mathbb{E}(\phi(X)|\mathcal{C}) \ \mathbb{P} - p.s..$$

PREUVE. Appliquer Jensen classique et le résultat précédent, c'est à dire Jensen à la loi conditionnelle.

Exercice : montrer que si $X \perp\!\!\!\perp Y$, et X intégrable, alors $\mathbb{E}(X|Y) = \mathbb{E}(X)$ $\mathbb{P}$-p.s..

Martingales

11.1. Martingales, sous-martingales.

Définition. *Soit* $(X_i)_{i\geq 1}$ *une suite de variables aléatoires définies sur un espace de probabilités* $(\Omega, \mathcal{B}, \mathbb{P})$ *et soit* $(\mathcal{F}_i)_{i\geq 1}$ *une suite de sous-*σ*-algèbres de* $\mathcal{B}$*. La suite* $(X_n, \mathcal{F}_n)_{n\geq 1}$ *est une martingale (resp. une sous-martingale) si*

$$\begin{cases} (i) & : \quad \mathcal{F}_n \subset \mathcal{F}_{n+1}; \\ (ii) & : \quad X_n \textit{ est } \mathcal{F}_n\textit{-mesurable}; \\ (iii) & : \quad \mathbb{E}(|X_n|) < \infty; \\ (iv) & : \quad \mathbb{P}\textit{-p.s.}, \ \mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n \\ & \qquad \textit{(resp. } \mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n\textit{).} \end{cases}$$

Remarquons que si la définition ci-dessus tient pour les $\mathcal{F}_n$, elle tient forcément pour les $\sigma(X_1, \ldots, X_n)$, car en posant $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$, on a $\mathcal{G}_n \subset \mathcal{F}_n$, et donc

$$\mathbb{E}(X_{n+1}|\mathcal{G}_n) = \mathbb{E}(\mathbb{E}(X_{n+1}|\mathcal{F}_n)|\mathcal{G}_n) = \mathbb{E}(X_n|\mathcal{G}_n) = X_n.$$

Remarquons qu'il découle de la croissance des $\mathcal{F}_n$ que pour chaque $k \geq 1$, dans le cas d'une martingale,

$$\mathbb{E}(X_{n+k}|\mathcal{F}_n) = X_n.$$

Exemple 1 : supposons que la suite $(Y_i)_{i\geq 1}$ soit i.i.d. intégrable centrée. Alors la suite $(X_n) = (Y_1 + \ldots + Y_n)$ est une martingale relativement à la filtration $(\mathcal{F}_n) = (\sigma(X_1, \ldots, X_n))$.

Si l'on suppose $Y_i \geq 0$ au lieu de $\mathbb{E}(Y_i) = 0$, on obtient une sous-martingale par le même procédé.

Exemple 2 : soit $(\Omega, \mathcal{B}, \mathbb{P})$ un espace probabilisé, et ν une mesure signée sur $(\Omega, \mathcal{B})$. Supposons que $(\mathcal{F}_n)_{n\geq 1}$ soit une suite croissante de sous-σ-algèbres de $\mathcal{B}$, et que pour chaque n, $\nu_{|\mathcal{F}_n} \ll \mathbb{P}_{|\mathcal{F}_n}$. Soit $X_n = d\nu \big/ d\mathbb{P}$ (restreintes à $\mathcal{F}_n$). Alors $(X_n, \mathcal{F}_n)_{n\geq 1}$ est une martingale.

Exemple 3 : soit Z intégrable sur $(\Omega, \mathcal{B}, \mathbb{P})$, et $(\mathcal{F}_n)_{n\geq 1}$ une suite croissante de sous-σ-algèbres. Notons $X_n = \mathbb{E}(Z|\mathcal{F}_n)$. Alors $(X_n, \mathcal{F}_n)$ est une martingale.

Exemple 4 : si $(X_n, \mathcal{F}_n)_{n\geq 1}$ est une martingale, alors $(|X_n|, \mathcal{F}_n)_{n\geq 1}$ est une sous-martingale.

Théorème - fonctions de martingales. *Si* $(X_n, \mathcal{F}_n)_{n\geq 1}$ *est une martingale (resp. sous-martingale), si* ϕ *est convexe (resp. convexe croissante), et si* $\phi(X_n)$ *est intégrable, alors* $(\phi(X_n), \mathcal{F}_n)_{n\geq 1}$ *est une sous-martingale.*

Preuve. Dans le cas sous-martingale, $X_n \leq \mathbb{E}(X_{n+1}|\mathcal{F}_n)$, et ϕ étant non décroissante, il s'ensuit que $\phi(X_n) \leq \phi(\mathbb{E}(X_{n+1}|\mathcal{F}_n))$, puis par application du lemme de Jensen conditionnel, pour lequel les conditions sont réunies, on en déduit $\phi(X_n) \leq \mathbb{E}(\phi(X_{n+1})|\mathcal{F}_n)$.

Le reste de la preuve n'est qu'une vérification à partir des définitions. ■

11.2. Temps d'arrêts.

Soit $(\Omega, \mathcal{B}, \mathbb{P})$ un espace de probabilités, et $(\mathcal{F}_n)_{n\geq 1}$ une suite croissante de sous-σ-algèbres de $\mathcal{B}$. Une variable aléatoire $\tau : \Omega \to \mathbb{N}^\star$ (prenant des valeurs entières positives) est un temps d'arrêt pour la filtration $(\mathcal{F}_n)_{n\geq 1}$ si pour chaque $n \geq 1$,

$$\{\tau = n\} \in \mathcal{F}_n \qquad (\Leftrightarrow \{\tau \leq n\} \in \mathcal{F}_n).$$

Nous définissons la sous-σ-algèbre $\mathcal{F}_\tau$ par

$$\mathcal{F}_\tau = \{A \in \mathcal{B} : A \cap \{\tau \leq k\} \in \mathcal{F}_k,\ 1 \leq k < \infty\}.$$

On peut dans la définition de $\mathcal{F}_\tau$ remplacer $\{\tau \leq k\}$ par $\{\tau = k\}$ sans que cela n'occasionne de modification de $\mathcal{F}_\tau$. Remarquons que τ est $\mathcal{F}_\tau$-mesurable.

Supposons que τ_1 et τ_2 soient deux temps d'arrêts et que $\tau_1 \leq \tau_2$. Si $A \in \mathcal{F}_{\tau_1}$, alors $A \cap \{\tau_1 \leq k\} \in \mathcal{F}_k$, et donc $A \cap \{\tau_2 \leq k\} = A \cap \{\tau_1 \leq k\} \cap \{\tau_2 \leq k\} \in \mathcal{F}_k$: et donc $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.

Théorème des temps d'arrêts. *Soit $X_1, \ldots, X_n$ une (sous-)martingale relativement à la filtration finie $\mathcal{F}_1 \subset \ldots \subset \mathcal{F}_n$, et soient τ_1 et τ_2 deux temps d'arrêts tels que $1 \leq \tau_1 \leq \tau_2 \leq n$.*

Alors X_{τ_1}, X_{τ_2} est une (sous-)martingale relativement à la filtration $\mathcal{F}_{\tau_1}, \mathcal{F}_{\tau_2}$.

En particulier $\mathbb{E}(X_{\tau_1}) \leq \mathbb{E}(X_{\tau_2})$.

Preuve. On a $|X_{\tau_i}| \leq \sum_{k=1}^n |X_k|$, et donc X_{τ_i} est intégrable, $i = 1, 2$. Remarquons aussi que X_{τ_i} est $\mathcal{F}_{\tau_i}$-intégrable.

Considérons le cas "sous-"martingale. On veut montrer que $\mathbb{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) \geq X_{\tau_1}$, ce qui revient à montrer que quel que soit $A \in \mathcal{F}_{\tau_1}$,

$$\int_A (X_{\tau_2} - X_{\tau_1}) d\mathbb{P} \geq 0,$$

puisque les deux variables $\mathbb{E}(X_{\tau_2}|\mathcal{F}_{\tau_1})$ et X_{τ_1} sont $\mathcal{F}_{\tau_1}$-mesurables.

Mais $A \in \mathcal{F}_{\tau_1} \Rightarrow A \cap \{\tau_1 < k \leq \tau_2\} = (A \cap \{\tau_1 \leq k-1\}) \cap \{\tau_2 \leq k-1\}^c \in \mathcal{F}_{k-1}$. Si nous posons $\Delta_k = X_k - X_{k-1}$, il s'ensuit que

$$\begin{aligned} \int_A (X_{\tau_2} - X_{\tau_1}) d\mathbb{P} &= \int_A \sum_{k=1}^n \mathbf{1}_{\tau_1 < k \leq \tau_2} \Delta_k d\mathbb{P} \\ &= \sum_{k=1}^n \int_{A \cap \{\tau_1 < k \leq \tau_2\}} \Delta_k d\mathbb{P} \geq 0, \end{aligned}$$

par propriété de sous-martingale. ■

11.3. Traversées ascendantes.

L'inégalité de base de la théorie pour la convergence presque sûre est celle des "**up-crossings**", ou "**traversées ascendantes**" : soit $[\alpha, \beta]$ un intervalle fermé, et soient $X_1, \ldots, X_n$ des variables aléatoires. Définissons inductivement $\tau_1, \tau_2, \ldots$ comme suit;

- τ_1 est le plus petit j, $1 \leq j \leq n$, tel que $X_j \leq \alpha$, et vaut n s'il n'existe pas de tel j;
- τ_k est, pour k pair, le plus petit j tel que $\tau_{k-1} < j \leq n$ et $X_j \geq \beta$, et vaut n s'il n'y a pas de tel j;
- τ_k, pour k impair plus grand que 1, est le plus petit j tel que $\tau_{k-1} < j \leq n$ et $X_j \leq \alpha$, et vaut n s'il n'y a pas de tel j.

Le nombre U de traversées ascendantes (en abrégé t.a.**) est le plus grand i tel que $X_{\tau_{2i-1}} \leq \alpha < \beta \leq X_{\tau_{2i}}$.**

Théorème des traversées. *Si $X_1, \ldots, X_n$ est une sous-martingale, et si U désigne le nombre de t.a. de $[\alpha, \beta]$, alors*

$$\mathbb{E}(U) \leq \frac{\mathbb{E}((X_n - \alpha)^+) - \mathbb{E}((X_1 - \alpha)^+)}{\beta - \alpha}.$$

Preuve. Posons $Y_k = \max\{0, X_k - \alpha\}$ et $\theta = \beta - \alpha$. Alors $Y_1, \ldots, Y_n$ est une sous-martingale. Les τ_k restent identiques si dans leur définition nous remplaçons la condition "$X_j \leq \alpha$" par "$Y_j = 0$", et "$X_j \geq \beta$" par "$Y_j \geq \theta$", et donc U désigne aussi bien le nombre de t.a. de la sous-martingale $Y_1, \ldots, Y_n$ de l'intervalle $[0, \theta]$. Si k est pair et si τ_{k-1} est un temps d'arrêt, alors pour $j < n$,

$$\{\tau_k = j\} = \cup_{i=1}^{j-1}\{\tau_{k-1} = i, Y_{i+1} < \theta, \ldots, Y_{j-1} < \theta, Y_j \geq \theta\} \in \mathcal{F}_j,$$

et de plus $\{\tau_k = n\} = \{\tau_k \leq n-1\}^c \in \mathcal{F}_n$. Et donc dans ce cas τ_k est aussi un temps d'arrêt.

Si k est impair, un raisonnement similaire montre que si τ_{k-1} est un temps d'arrêt, alors τ_k en est un aussi.

Et donc les τ_k sont des temps d'arrêts si τ_1 en est un. Mais ceci est clair d'après la définition.

La suite de temps d'arrêts $\tau_1, \ldots, \tau_n$ est strictement croissante tant qu'elle n'a pas atteint n, à la suite de quoi elle vaut constamment n. En particulier $\tau_n \equiv n$. Et donc

$$Y_n = Y_{\tau_n} \geq Y_{\tau_n} - Y_{\tau_1} = \sum_{k=2}^{n}(Y_{\tau_k} - Y_{\tau_{k-1}}) = \Sigma_p + \Sigma_i,$$

où Σ_p est la somme sur les indices k pairs, et Σ_i, sur les impairs. D'après le Théorème des temps d'arrêts, $\mathbb{E}(\Sigma_i) \geq 0$ et $\mathbb{E}(Y_1) \leq \mathbb{E}(Y_{\tau_1})$, et donc

$$\mathbb{E}(Y_n) \geq \mathbb{E}(Y_n) - \mathbb{E}(Y_1) \geq \mathbb{E}(Y_n) - \mathbb{E}(Y_{\tau_1}) \geq ee(\Sigma_p).$$

Si $Y_{\tau_{2i-1}} = 0 < \theta \leq Y_{\tau_{2i}}$, alors la différence $Y_{\tau_{2i}} - Y_{\tau_{2i-1}}$ apparaît dans Σ_p, et vaut au moins θ. Puisque ceci se produit U fois, on a $\Sigma_p \geq \theta U$, et donc

$$\mathbb{E}(Y_n) - \mathbb{E}(Y_1) \geq \theta \mathbb{E}(U).$$

■

11.4. Convergence presque sûre.

Le théorème de convergence des martingales, dû à Doob, revêt différentes forme, dont voici l'une :

Convergence des martingales. *Soit* $X_1, \dots, X_n, \dots$ *une sous-martingale, et supposons que* $K = \sup_n \mathbb{E}(X_n^+) < \infty$. *Alors il existe une variable aléatoire intégrable* X*, telle que* $\mathbb{E}(X^+) \leq K$*, et que*

$$X_n \to X \ \mathbb{P} - p.s..$$

Preuve. Fixons $\alpha < \beta$, et notons U_n le nombre de t.a. de $X_1, \dots, X_n$ de $[\alpha, \beta]$. Par le théorème des traversées ascendantes, (U_n) est une suite de variables aléatoires prenant des valeurs entières, croissante, et telle que

$$\mathbb{E}(U_n) \leq \frac{K + |\alpha|}{\beta - \alpha}, \ n \geq 1.$$

Par convergence monotone, il s'ensuit donc que la suite (U_n) converge vers $\sup_n U_n$, qui est intégrable et donc $\mathbb{P}$-p.s. finie.

Notons $X_* = \liminf X_n$ et $X^* = \limsup X_n$. Si $X_* < \alpha < \beta < X^*$, U_n doit tendre vers l'infini, ce qui n'est pas, presque sûrement. Et donc pour tout $\alpha < \beta$,

$$\mathbb{P}(X_* < \alpha < \beta < X^*) = 0.$$

Mais d'autre part,

$$\{X_* \neq X^*\} = \bigcup_{\substack{\alpha, \beta \in \mathbb{Q}, \\ \alpha < \beta}} \{X_* < \alpha < \beta < X^*\},$$

et donc $\mathbb{P}(X_* = X^*) = 1$. Notons $X = \lim_n X_n$ $(= X_* = X^*)$.

Pour conclure, par le Lemme de Fatou,

$$\mathbb{E}(X^+) \leq \liminf_n \mathbb{E}(X_n^+) \leq K < \infty,$$

donc $X^+ \in \mathcal{L}^1$. Par la propriété des sous-martingales,

$$\mathbb{E}(X_n^-) = \mathbb{E}(X_n^+) - \mathbb{E}(X_n) \leq \mathbb{E}(X_n^+) - \mathbb{E}(X_1),$$

et par le lemme de Fatou,

$$\mathbb{E}(X^-) \leq \liminf_n \mathbb{E}(X_n^-) \leq K - \mathbb{E}(X_1^-) < +\infty.$$

Donc $X \in \mathcal{L}^1$. ■

Un exemple important de martingale est celui de $X_n = \mathbb{E}(Z|\mathcal{F}_n)$ avec Z fixée intégrable et $(\mathcal{F}_n)$ une filtration croissante (Exemple 3).

LEMME. *Si Z est intégrable et si la suite $(\mathcal{F}_n)$ est une suite arbitraire de sous-σ-algèbres, alors la suite $(\mathbb{E}(Z|\mathcal{F}_n))$ est équi-intégrable.*

PREUVE. On a $|\mathbb{E}(Z|\mathcal{F}_n)| \leq \mathbb{E}(|Z||\mathcal{F}_n)$, et donc nous pouvons supposer $Z \geq 0$. Notons $A_{\alpha,n} = \{\mathbb{E}(Z|\mathcal{F}_n) \geq \alpha\}$. Puisque $A_{\alpha,n} \in \mathcal{F}_n$, il vient

$$\int_{A_{\alpha,n}} \mathbb{E}(Z|\mathcal{F}_n) d\mathbb{P} = \int_{A_{\alpha,n}} Z d\mathbb{P}.$$

Il suffit donc, pour $\varepsilon > 0$ donné, de trouver un $\alpha > 0$ tel que cette dernière intégrale ne dépasse pas ε, indépendamment de n.

Mais $A \mapsto \int_A Z d\mathbb{P}$ définit une mesure positive finie $\nu_Z \ll \mathbb{P}$, et par caractérisation de l'absolue continuité, il s'ensuit qu'il existe $\delta > 0$ tel que quel que soit A, $\mathbb{P}(A) < \delta \Rightarrow \nu_Z(A) < \varepsilon$.

Il nous suffit donc de trouver $\alpha > 0$ tel que pour tout n, $\mathbb{P}(A_{\alpha,n}) < \delta$. Cela ira si α est assez grand, car par l'inégalité de Markov, on a

$$\mathbb{P}(\mathbb{E}(Z|\mathcal{F}_n) \geq \alpha) \leq \frac{1}{\alpha}\mathbb{E}(\mathbb{E}(Z|\mathcal{F}_n)) = \frac{1}{\alpha}\mathbb{E}(Z).$$

■

Supposons maintenant que nous ayons une filtration, i.e. que $\mathcal{F}_1 \subset \ldots \subset \mathcal{F}_n \subset \ldots$. Notons $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. **Nous notons** $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, ce qui signifie simplement que $(\mathcal{F}_n)$ croît pour l'inclusion, et que $\mathcal{F}_\infty$ est engendrée par les $\mathcal{F}_n$.

CONVERGENCE PRESQUE SÛRE DES ESPÉRANCES CONDITIONNELLES. *Si $Z \in \mathcal{L}^1$ et si $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, alors*

$$\mathbb{E}(Z|\mathcal{F}_n) \to \mathbb{E}(Z|\mathcal{F}_\infty)\ \mathbb{P} - p.s..$$

PREUVE. Pour appliquer le théorème de convergence p.s., montrons d'abord que $\sup_n \mathbb{E}(|\mathbb{E}(Z|\mathcal{F}_n)|) = K < \infty$. Cela découle de ce que

$$\sup_n \mathbb{E}(|\mathbb{E}(Z|\mathcal{F}_n)|) \leq \sup_n \mathbb{E}(\mathbb{E}(|Z||\mathcal{F}_n)) = \mathbb{E}(|Z|) = K < \infty.$$

Soit donc $X \in \mathcal{L}^1$ telle que $\mathbb{E}(Z|\mathcal{F}_n) \to X$ $\mathbb{P}$-p.s.. Alors X est limite d'une suite de variables $\mathcal{F}_\infty$-mesurables, donc l'est elle-même.

Par intégrabilité uniforme, quel que soit $A \in \cup_k \mathcal{F}_k$, on a

$$\int_A X d\mathbb{P} = \lim_n \int_A \mathbb{E}(Z|\mathcal{F}_n) d\mathbb{P} = \lim_n \int_A Z d\mathbb{P} = \int_A Z d\mathbb{P},$$

et puisque $\cup_k \mathcal{F}_k$ est un π-système générateur de $\mathcal{F}_\infty$, il s'en suit que $X = \mathbb{E}(Z|\mathcal{F}_\infty)$. ■

Corollaire : loi $0-1$ de Lévy. *Si $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, et si $A \in \mathcal{F}_\infty$, alors*

$$\mathbb{E}(\mathbf{1}_A|\mathcal{F}_n) \to \mathbf{1}_A \ \mathbb{P}-p.s.$$

Exercice : montrer que la loi $0-1$ de Kolmogorov découle de la loi $0-1$ de Lévy.

Le cas des dérivées de Radon-Nikodym. *Supposons que $(\Omega, \mathcal{B}, \mathbb{P})$ soit un espace probabilisé, que ν soit une mesure positive finie sur $(\Omega, \mathcal{B})$, et que $\mathcal{F}_n \uparrow \mathcal{F}_\infty \subset \mathcal{B}$. Supposons que restreinte à chaque $\mathcal{F}_n$, $\nu \ll \mathbb{P}$, et notons $X_n = d\nu \,/\, d\mathbb{P}$ la dérivée correspondante. Alors il existe $X \in \mathcal{L}^1$ telle que*

$$X_n \to X \ \mathbb{P}-p.s.,$$

et

- *si $\nu \ll \mathbb{P}$ sur $\mathcal{F}_\infty$, alors $X = d\nu \,/\, d\mathbb{P}$ relativement à $\mathcal{F}_\infty$;*
- *si $\nu \perp \mathbb{P}$ sur $\mathcal{F}_\infty$, alors $X \equiv 0$.*

Preuve. On a $X_n \geq 0$ et $\mathbb{E}(|X_n|) = \mathbb{E}(X_n) = \nu(\Omega) = K < \infty$. De plus, la suite $(X_n, \mathcal{F}_n)$ constitue une martingale, et donc le théorème de convergence p.s. s'applique. Soit X la limite correspondante, $\mathcal{F}_\infty$-mesurable.

Si $\nu \ll \mathbb{P}$ sur $\mathcal{F}_\infty$, notons $Z = d\nu \,/\, d\mathbb{P}$ relativement à $\mathcal{F}_\infty$. Alors pour chaque n et chaque $A \in \mathcal{F}_n$,

$$\int_A Z d\mathbb{P} = \nu(A) = \int_A X_n d\mathbb{P},$$

et donc pour chaque n, $X_n = \mathbb{E}(Z|\mathcal{F}_n)$. Et donc par le théorème précédent, $X = \mathbb{E}(Z|\mathcal{F}_\infty)$, mais puisque Z est déjà $\mathcal{F}_\infty$-mesurable, il vient

$$X = Z.$$

Supposons sinon que $\nu \perp \mathbb{P}$ sur $\mathcal{F}_\infty$. Il existe donc $S \in \mathcal{F}_\infty$ tel que $\nu(S) = 0$ et $\mathbb{P}(S) = 1$. Par le lemme de Fatou, $\int_A X d\mathbb{P} \leq \liminf_n \int_A X_n d\mathbb{P} = \nu(A)$ si A est dans l'algèbre $\cup_k \mathcal{F}_k$.

Il s'ensuit par le théorème des classes monotones (Chap. 1) que $\int_A X d\mathbb{P} \leq \nu(A)$ reste vrai pour tout $A \in \mathcal{F}_\infty$, et donc en particulier si $A = S$. Et puisque $X \geq 0$, il s'ensuit que $\mathbb{E}(X) \leq 0$ entraîne que $X = 0$ $\mathbb{P}$-p.s.. ■

11.5. Slutsky, intégrabilité uniforme et convergence dans L^1.

D'abord commençons par un théorème utile :

Théorème de Slutsky. *Si $V_n \xrightarrow{\mathcal{L}} V$ et $W_n \xrightarrow{\mathcal{L}} 0$, alors $V_n + W_n \xrightarrow{\mathcal{L}} V$*

Si $X_n \xrightarrow{\mathcal{L}} X$ et si $Y_n \xrightarrow{\mathcal{L}} a$, alors $X_n Y_n \xrightarrow{\mathcal{L}} aX$.

Preuve. Observons d'abord que $\mathbf{W_n} \xrightarrow{\mathcal{L}} \mathbf{0} \Rightarrow \mathbf{W_n} \xrightarrow{\mathbb{P}} \mathbf{0}$. En effet, puisque F_0 n'a qu'un point de discontinuité en 0, on a, si $\varepsilon > 0$, $\mathbb{P}(|W_n| \geq \varepsilon) = F_{Y_n}(-\varepsilon) + \mathbb{P}(Y_n \geq \varepsilon) \leq F_{Y_n}(-\varepsilon) + 1 - F_{Y_n}(\frac{\varepsilon}{2}) \to 0$, et donc par écrabouillage, $\mathbb{P}(|Y_n| \geq \varepsilon) \to 0$.

Soit $v \in \mathbb{R}$, et soient $w' < v < w''$ tels que $\mathbb{P}(V = w') = \mathbb{P}(V = w'') = 0$. Si $w' < v - \varepsilon < v < v + \varepsilon < w''$, alors

$$\mathbb{P}(V_n \leq w') - \mathbb{P}(|W_n| \geq \varepsilon) \leq \mathbb{P}(V_n + W_n \leq v) \leq \mathbb{P}(V_n \leq w'') + \mathbb{P}(|W_n| \geq \varepsilon),$$

ce qui, puisque $V_n \xrightarrow{\mathcal{L}} V$ et $W_n \xrightarrow{\mathbb{P}} 0$, donne à la limite

$$\mathbb{P}(V \leq w') \leq \liminf_n \mathbb{P}(V_n + W_n \leq v) \leq \limsup_n \mathbb{P}(V_n + W_n \leq v) \leq \mathbb{P}(V \leq w'').$$

Si alors $\mathbb{P}(V = v) = 0$, F_V est continue en v, et on peut trouver des $w' < v < w''$ arbitrairement proches de v, et aux propriétés requises. Donc $V_n + W_n \xrightarrow{\mathcal{L}} V$ puisque les fonctions caractéristiques convergent aux points de continuité de celle de V.

Pour la seconde partie de cet énoncé, notons que $X_n Y_n = X_n(Y_n - a) + aX_n$, et que si $Y_n \xrightarrow{\mathcal{L}} a$, alors $Y_n - a \xrightarrow{\mathcal{L}} 0$ (exprimer que c'est une convegrence en probabilités), et enfin que si $X_n \xrightarrow{\mathcal{L}} X$, alors $aX_n \xrightarrow{\mathcal{L}} aX$ (distinguer ici le cas $a = 0$, et raisonner à l'aide des fonctions de répartitions).

Par la première partie de notre énoncé, si nous montrons que $X_n(Y_n - a) \xrightarrow{\mathcal{L}} 0$, alors la seconde partie s'en suivra. Notons donc $\delta_n = Y_n - a$, et nos hypothèses sont à présent que $\delta_n \xrightarrow{\mathcal{L}} 0$ et $X_n \xrightarrow{\mathcal{L}} X$. La conclusion recherchée est $X_n \delta_n \xrightarrow{\mathcal{L}} 0$. On calcul... soient $\varepsilon > 0$ et $\lambda > 0$ tels que $\mathbb{P}(|X| = \frac{\varepsilon}{\lambda}) = 0$:

$$\begin{array}{rcl} \limsup_n \mathbb{P}(|\delta_n X_n| \geq \varepsilon) & \leq & \limsup_n \mathbb{P}(|\delta_n| \geq \lambda) + \limsup_n \mathbb{P}(|X_n| \geq \frac{\varepsilon}{\lambda}) \\ & \leq & 0 + F_Z(-\frac{\varepsilon}{\lambda}) + 1 - F_Z(\frac{\varepsilon}{\lambda}) \end{array}$$

qui est arbitrairement petit (il suffit de choisir λ grand). Donc par écrabouillage,

$$\limsup_n \mathbb{P}(|\delta_n Z_n| \geq \varepsilon) = 0,$$

soit $\delta_n X_n \xrightarrow{\mathbb{P}} 0$. $\square$

Nous avons rencontré l'équi-intégrabilité à l'occasion du chapitre 6. Le théorème suivant montre son lien étroit avec la convergence dans L^1 :

Equi-intégrabilité et convergence $\mathcal{L}^1$. *Supposons que $X_n \xrightarrow{\mathcal{L}} X$, $X_n, X \in \mathcal{L}^1$. Alors les conditions suivantes sont équivalentes :*
(i) : $(X_n)_{n\geq 1}$ est équi-intégrable;
(ii) : $X_n \to X$ dans L^1;
(iii) : $\mathbb{E}(|X_n|) \to \mathbb{E}(|X|)$.

Preuve.

(i) ⇒ (ii) : soit $M > 0$ et définissons

$$\phi_M(x) = \begin{cases} M \text{ si } x \geq M; \\ x \text{ si } |x| \leq M; \\ -M \text{ si } x < -M. \end{cases}$$

Par l'inégalité triangulaire, on a

$$|X_n - X| \leq |X_n - \phi_M(X_n)| + |\phi_M(X_n) - \phi_M(X)| + |\phi_M(X) - X|.$$

Puisque $|\phi_M(Y) - Y| = (|Y| - M)^+ \leq |Y|\mathbf{1}_{|Y|>M}$, en passant aux espérances, il vient

$$\mathbb{E}(|X_n - X|) \leq \mathbb{E}(|\phi_M(X_n) - \phi_M(X)|) + \mathbb{E}(|X_n|\mathbf{1}_{|X_n|>M}) + \mathbb{E}(|X|\mathbf{1}_{|X|>M}).$$

On a $\phi_M(X_n) \xrightarrow{\mathcal{L}} \phi_M(X)$, puisque ϕ_M est continue. Par continuité du module et en application du Théorème de Slutsky, il vient $Y_n := |\phi_M(X_n) - \phi_M(X)| \xrightarrow{\mathcal{L}} 0$. Par Skorohod, il existe des variables Z_n de lois respectives celles des variables Y_n, et telles que $Y_n \to 0$ $\mathbb{P}$-p.s.. On a $\mathbb{P}(|Z_n| > 2M) = \mathbb{P}(|Y_n| > 2M) \leq \mathbb{P}(|\phi_M(X_n)| > M) + \mathbb{P}(|\phi_M(X)| > M) = 0$, et donc $|Y_n| \leq 2M$ $\mathbb{P}$-p.s.. Donc par convergence dominée, il vient $\mathbb{E}(Y_n) = \mathbb{E}(|\phi_M(X_n) - \phi_M(X)|) \to 0$.

Par uniforme intégrabilité, $\lim_M \sup_n \mathbb{E}(|X_n|\mathbf{1}_{|X_n|>M}) = 0$.

A ce stade on pourrait même conclure ici que $X \in \mathcal{L}^1$. En effet, par uniforme intégrabilité toujours, on a $\sup_n \mathbb{E}(|X_n|) < +\infty$, et le lemme de Fatou (dans le cas dominé) entraîne, via le Théorème de Skorohod de nouveau, que $\mathbb{E}(|X|) \leq \sup_n \mathbb{E}(|X_n|) < +\infty$. Donc $X \in \mathcal{L}^1$, et donc $\lim_M \mathbb{E}(|X|\mathbf{1}_{|X|>M}) = 0$. Mais ceci résulte plus simplement directement de nos hypothèses.

(ii) ⇒ (iii) : découle de

$$|\mathbb{E}(|X_n|) - \mathbb{E}(|X|)| \leq \mathbb{E}(||X_n| - |X||) \leq \mathbb{E}(|X_n - X|) \to 0.$$

(iii) ⇒ (i) : soit $\psi_M : \mathbb{R}^+ \to \mathbb{R}^+$ continue telle que $\psi_M(x) = x$ sur $[0, M-1]$, $\psi_M \equiv 0$ sur $[M, +\infty[$, et ψ_M linéaire sur $[M-1, M]$. Soit $\varepsilon > 0$.

Par convergence dominée, si M est assez grand, $\mathbb{E}(|X|) - \mathbb{E}(\psi_M(|X|)) \leq \frac{\varepsilon}{2}$.

D'autre part $\psi_M(|X_n|) \xrightarrow{\mathcal{L}} \psi_M(|X|)$, et donc par convergence dominée, via Skorohod encore une fois, la domination s'obtenant de façon semblable à son obtention dans le passage $(i) \Rightarrow (ii)$, $\mathbb{E}(\psi_M(|X_n|)) \to \mathbb{E}(\psi_M(|X|))$.

Donc avec (iii), on a pour $n \geq n_0$,

$$\mathbb{E}(|X_n|\mathbf{1}_{|X_n|>M}) \leq \mathbb{E}(|X_n|) - \mathbb{E}(\psi_M(|X_n|)) \leq \mathbb{E}(|X|) - \mathbb{E}(\psi_M(|X|)) + \frac{\varepsilon}{2} \leq \varepsilon.$$

Au besoin en augmentant M on obtient simultanément $\sup_{n<n_0} \mathbb{E}(|X_n|\mathbf{1}_{|X_n|>M}) \leq \varepsilon$, et donc l'équi-intégrabilité (une suite finie dans $\mathcal{L}^1$ est toujours équi-intégrable). ■

Convergence L^1 des sous-martingales. *Pour une sous-martingale, les conditions suivantes sont équivalentes :*
(i) *: elle est uniformément intégrable;*
(ii) *: elle converge presque sûrement et dans L^1;*
(iii) *: elle converge dans L^1.*

Preuve.
(i) ⇒ (ii) : l'uniforme intégrabilité entraîne que $\sup_n \mathbb{E}(|X_n|) < +\infty$, et donc la sous-martingale converge presque sûrement. Le théorème qui précède montre, par équi-intégrabilité, que sa limite presque sûre l'est aussi au sens L^1.
(ii) ⇒ (iii) : évident.
(iii) ⇒ (i) : la convergence L^1 entraîne la convergence en probabilité, et le théorème précédent permet d'en déduire l'équi-intégrabilité. ■

Intéressons nous à présent au cas plus spécifique des martingales :

Lemme. *Si (X_n) est une martingale relativement à la filtration $(\mathcal{F}_n)$ et si $X_n \to X$ dans L^1, alors*

$$X_n = \mathbb{E}(X|\mathcal{F}_n).$$

Preuve. La propriété de martingale implique que si $m > n$, $\mathbb{E}(X_m|\mathcal{F}_n) = X_n$, et donc si $A \in \mathcal{F}_n$, alors $\mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(X_m \mathbf{1}_A)$. Mais $X_n \to X$ dans L^1 entraîne que $X_n \mathbf{1}_A \to X \mathbf{1}_A$ dans L^1, et donc $\mathbb{E}(X_m \mathbf{1}_A) \to \int_A X d\mathbb{P}$. Ainsi

$$A \in \mathcal{F}_n \Rightarrow \int_A X_n d\mathbb{P} = \int_A X d\mathbb{P},$$

et donc X_n est une version de $\mathbb{E}(X|\mathcal{F}_n)$. ■

Convergence L^1 des martingales. *Si (X_n) est une martingale relativement à la filtration $(\mathcal{F}_n)$, alors les conditions suivantes sont équivalentes :*
(i) *: elle est équi-intégrable;*
(ii) *: elle converge p.s. et dans L^1;*
(iii) *: elle converge dans L^1;*
(iv) *: il existe une variable aléatoire intégrable X telle que pour chaque n, $X_n = \mathbb{E}(X|\mathcal{F}_n)$.*

Preuve. On sait déjà que $(i - iii)$ sont équivalentes entre elles. On sait aussi que (iv) entraîne (i). Et le lemme qui précède montre que $(iii) \Rightarrow (iv)$. ■

Corollaire. *Dans le §11.4., à la rubrique "Convergence presque sûre des espérances conditionnelles", il est possible de rajouter la convergence L^1 à la convergence p.s..*

Preuve. On a affaire à une martingale, équi-intégrable par le théorème qui précède. ■

11.6. Convergence L^p, $p > 1$.

Commençons par une inégalité maximale, utilisée dans les preuves de convergence L^p des martingales :

Inégalité maximale de Doob. *Si $X_1, \dots, X_n$ est une sous-martingale, et si $\alpha > 0$, alors*

$$\begin{cases} \mathbb{P}\left(\max_{i\le n} X_i^+ \ge \alpha\right) \le \frac{1}{\alpha}\mathbb{E}(X_n^+ \mathbf{1}_{\max_{i\le n} X_i^+ \ge \alpha}) \le \frac{1}{\alpha}\mathbb{E}(X_n^+); \\ \mathbb{P}\left(\max_{i\le n} X_i^+ > \alpha\right) \le \frac{1}{\alpha}\mathbb{E}(X_n^+ \mathbf{1}_{\max_{i\le n} X_i^+ > \alpha}) \le \frac{1}{\alpha}\mathbb{E}(X_n^+). \end{cases}$$

Remarquons que ceci étend l'inégalité maximale de Kolmogorov : si $S_1, S_2, \dots$ sont les sommes partielles de variables indépendantes centrées dans $\mathcal{L}^2$, alors leur suite forme une martingale, et donc la suite de leurs carrés constitue une sous-martingale. Pour les carrés, on retrouve bien l'inégalité maximale de Kolmogorov, puisque

$$\{\max_{k\le n} |S_k| \ge \alpha\} = \{\max_{k\le n} |S_k|^2 \ge \alpha^2\}.$$

Preuve. Remarquons que $\phi(x) = \max\{x, 0\} = x^+$ est convexe croissante, et donc par le théorème "*fonctions de sous-martingales*", $(X_n^+)_{n\ge 1}$ est une sous-martingale (l'intégrabilité est automatique car $X_n^+ \le |X_n|$). Nous supposons donc que $X_n \ge 0$.

Posons $\tau_2 = n$, et soit $\tau_1 = \min\{n, \inf\{k \ge 1 : X_k \ge \alpha\}\}$. On vérifie sans peine que $1 \le \tau_1 \le \tau_2 \le n$, et que τ_1 et τ_2 sont des temps d'arrêts.

On pose ensuite $M_k = \max_{i\le k} X_i$, et l'on vérifie que

$$\{M_n \ge \alpha\} \cap \{\tau_1 \le k\} = \{M_k \ge \alpha\} \in \mathcal{F}_k,$$

et donc $\{M_n \ge \alpha\} \in \mathcal{F}_{\tau_1}$. Par application du théorème des temps d'arrêts, il vient, puisque $X_{\tau_1} \ge \alpha$ sur $\{M_n \ge \alpha\}$,

$$\begin{aligned} \alpha\mathbb{P}(M_n \ge \alpha) &\le \int_{M_n\ge\alpha} X_{\tau_1} d\mathbb{P} \le \int_{M_n\ge\alpha} \mathbb{E}(X_n|\mathcal{F}_{\tau_1}) d\mathbb{P} \\ &= \int_{M_n\ge\alpha} X_n d\mathbb{P} \le \mathbb{E}(X_n), \end{aligned}$$

ce qui démontre la première inégalité.

Pour la seconde, on se ramène d'abord à nouveau au cas de $X_n \ge 0$, puis on pose $\tau_2 \equiv n$, et $\tau_1 = \min\{n, \inf\{k \ge 1 : X_k > \alpha\}\}$. On obtient à nouveau des temps d'arrêts, et $\{M_n > \alpha\} \in \mathcal{F}_{\tau_1}$. La conclusion s'en suit par un calcul similaire. ■

Remarquons que si (X_i) forme une martingale, $(|X_i|)$ constitue une sous-martingale, et donc dans ce cas l'inégalité maximale précédente peut s'écrire en remplaçant X_i^+ par $|X_i|$.

Inégalité maximale $\mathcal{L}^p$. *Soit $p > 1$ et (X_n) une sous-martingale. Notons $\bar{X}_n = \max_{1\le k\le n} X_k^+$. Alors*

$$\mathbb{E}(\bar{X}_n^p) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}((X_n^+)^p).$$

En particulier si (Y_n) est une martingale et si $Y_n^* = \max_{1\le k\le n} |Y_k|$, alors

$$\mathbb{E}((Y_n^*)^p) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}(|Y_n|^p)$$

Preuve. La seconde inégalité découle de la première en posant $X_n = |Y_n|$. Par le même argument que celui utilisé dans la preuve de l'inégalité maximale précédente, nous supposerons sans perte de généralité que $X_n \ge 0$.

Notons $\bar{X}_n \wedge M = \min\{\bar{X}_n, M\}$, avec $M > 0$. Comme $\{\bar{X}_n \wedge M > \lambda\}$ vaut soit $\{\bar{X}_n > \lambda\}$ soit $\emptyset$, cela ne modifie pas l'application de l'inégalité de Doob. On a (calcul du moment d'ordre $p \ge 1$ d'une v.a. positive)

$$\begin{array}{rcl}
\mathbb{E}((\bar{X}_n \wedge M)^p) & = & \int_0^\infty p\lambda^{p-1}\mathbb{P}(\bar{X}_n \wedge M > \lambda)d\lambda \\
& \le & \int_0^\infty p\lambda^{p-1}\left(\frac{1}{\lambda}\int X_n \mathbf{1}_{\bar{X}_n\wedge M>\lambda} d\mathbb{P}\right)d\lambda \\
& = & \int X_n \int_0^{\bar{X}_n\wedge M} p\lambda^{p-2} d\lambda d\mathbb{P} \\
& = & \frac{p}{p-1}\int X_n(\bar{X}_n \wedge M)^{p-1} d\mathbb{P}.
\end{array}$$

Posons $q = p \,/\, (p-1)$ - l'exposant conjugué de p - et appliquons l'inégalité de Hölder : on obtient

$$\mathbb{E}((\bar{X}_n \wedge M)^p) \le q\mathbb{E}(|X_n|^p)^{\frac{1}{p}}\mathbb{E}((\bar{X}_n \wedge M)^p)^{\frac{1}{q}}.$$

Arrivés là, si $\mathbb{E}((X_n)^p) = +\infty$, il n'y a rien à démontrer. Sinon, en divisant les deux côtés de l'inégalité précédente par $\mathbb{E}((\bar{X}_n \wedge M)^p)^{\frac{1}{q}}$ et en élevant à la puissance p, on en déduit

$$\mathbb{E}((\bar{X}_n \wedge M)^p) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}((X_n)^p).$$

Il suffit ensuite de laisser $M \to \infty$, et d'appliquer le théorème de convergence monotone. ∎

Convergence p.s. et L^p d'une martingale. *Si $p > 1$ et si (X_n) est une martingale, si $\sup_n \mathbb{E}(|X_n|^p) = K < \infty$, alors (X_n) converge p.s. et dans L^p.*

Preuve. On a $\mathbb{E}(X_n^+)^p \le \mathbb{E}(|X_n|^p)$ (Jensen, $p > 1$), et donc il existe $X \in \mathcal{L}^1$ telle que $X_n \to X$ p.s..

La seconde conclusion de l'inégalité maximale précédente s'applique, et entraîne que

$$\mathbb{E}\left((\sup_{1\le k\le n} |X_k|)^p\right) \le \left(\frac{p}{p-1}\right)^p K.$$

En faisant $n \to \infty$, par convergence monotone des “sup”, il s'ensuit que $\sup |X_n| \in \mathcal{L}^p$. Ensuite, on remarque que
$$|X_n - X|^p \leq (2 \sup_n |X_n|)^p,$$
et donc par convergence dominée, $\mathbb{E}(|X_n - X|^p) \to 0$. ∎

11.7. Martingales renversées.

Une suite infinie $\ldots, X_{-2}, X_{-1}$ est une martingale relativement à la filtration $\ldots \subset \mathcal{F}_{-2} \subset \mathcal{F}_{-1}$ si les définitions de martingale standard sont staisfaites pour $n \leq -1$ ou $n < -1$ si nécessaire.

Une telle suite s'appelle une **martingale renversée**.

Convergence des martingales renversées. *Pour une martingale renversée,*
$$\lim_{n\to\infty} X_{-n} = X$$
existe presque sûrement, est intégrable, et satisfait $\mathbb{E}(X) = \mathbb{E}(X_{-n})$ *pour tout* $n \geq 1$. *La convergence a lieu aussi dans* $\mathcal{L}^1$.

Preuve. La preuve est quasi identique à celle de la convergence des martingales “directes” (cf. p. 96). Remarquons que l'hypothèse de martingale renversée entraîne, d'après p. 89, et Jensen conditionnelle, que
$$\mathbb{E}(X_{-n}^+) = \mathbb{E}(\mathbb{E}(X_{-1}|\mathcal{F}_{-n})^+) \leq \mathbb{E}(\mathbb{E}(X_{-1}^+|\mathcal{F}_n)) = \mathbb{E}(X_{-1}^+) < +\infty,$$
et donc que le pendant, au cas présent, de l'hypothèse de finitude du sup des intégrales de la page 96, est satisfait.

Notons X^* et X_* les limites inférieures et supérieures de la suite $(X_{-n})_{n\geq 1}$. Notons U_n le nombre de t.a. de $X_{-n}, \ldots, X_{-1}$ de l'intervalle $[\alpha, \beta]$.

On montre que (U_n) converge en croissant vers $\sup_n U_n$ qui est intégrable positive entière. Le même argument permet de montrer que $X_* = X^*$ $\mathbb{P}$-p.s., et donc qu'il existe $X = \lim_n X_{-n}$.

Ici la propriété de martingale permet d'envisager $X_{-n} = \mathbb{E}(X_{-1}|\mathcal{F}_{-n})$, et donc d'utiliser l'équi-intégrabilité de (X_{-n}) pour conclure que puisque $X_{-n} \xrightarrow{\mathcal{L}} X$, $X \in \mathcal{L}^1$ et $X_{-n} \to X$ dans $\mathcal{L}^1$ (cf. le second énoncé du §11.5.). ∎

Corollaire. *Si* $\mathcal{F}_1 \supset \mathcal{F}_2 \supset \ldots \supset \mathcal{F}_n \supset \ldots$ *est une suite décroissante de sous-*σ*-algèbres, alors* $\mathcal{F}_0 = \cap_n \mathcal{F}_n$ *est une sous-*σ*-algèbre, et l'on note* $\mathcal{F}_n \downarrow \mathcal{F}_0$.

Si $Z \in \mathcal{L}^1$ *et si* $\mathcal{F}_n \downarrow \mathcal{F}_0$, *alors*
$$\mathbb{E}(Z|\mathcal{F}_n) \to \mathbb{E}(Z|\mathcal{F}_0)\ \mathbb{P}-p.s.\ et\ dans\ \mathcal{L}^1.$$

Preuve. Si $X_{-n} = \mathbb{E}(Z|\mathcal{F}_n)$, alors (X_{-n}) est une martingale renversée. Et donc il existe X intégrable telle que

$$X = \lim_n \mathbb{E}(Z|\mathcal{F}_n)\ \mathbb{P} - p.s..$$

Par décroissance des $\mathcal{F}_n$, X est $\mathcal{F}_n$-mesurable pour chaque n, et donc $\mathcal{F}_0$-mesurable.

Par équi-intégrabilité, si $A \in \mathcal{F}_0$,

$$\begin{array}{rcl} \int_A X d\mathbb{P} & = & \lim_n \int_A \mathbb{E}(Z|\mathcal{F}_n) d\mathbb{P} = \lim_n \int_A \mathbb{E}(\mathbb{E}(Z|\mathcal{F}_n)|\mathcal{F}_0) d\mathbb{P} \\ & = & \lim_n \int_A \mathbb{E}(Z|\mathcal{F}_0) d\mathbb{P} = \int_A \mathbb{E}(Z|\mathcal{F}_0) d\mathbb{P}, \end{array}$$

et donc X est une version de $\mathbb{E}(Z|\mathcal{F}_0)$.

La convergence $\mathcal{L}^1$ résulte de la propriété de martingale renversée. ■

Exercice : "*une nouvelle preuve de la loi forte classique*".

Soit $(X_i)_{i\geq 1}$ une suite i.i.d. dans $\mathcal{L}^1$. On pose $S_n = X_1 + \ldots + X_n$ et $Y_{-n} = \frac{S_n}{n}$, $n \geq 1$. On note aussi

$$\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, \ldots) = \sigma(S_n, X_{n+1}, \ldots).$$

Montrer que si $j, k \leq n+1$, $\mathbb{E}(X_j|\mathcal{F}_{-n-1}) = \mathbb{E}(X_k|\mathcal{F}_{-n-1})$.

En déduire que $\mathbb{E}(X_{n+1}|\mathcal{F}_{-n-1}) = \frac{S_{n+1}}{n+1}$.

En déduire que $(Y_{-n}, \mathcal{F}_{-n})_{n\geq 1}$ est une martingale renversée et qu'elle converge.

En déduire une loi forte des grands nombres, en utilisant la loi $0-1$ de Kolmogorov.

Exercice : "*une martingale qui converge p.s. mais pas dans $\mathcal{L}^1$*".

Soit $(X_n)_{n\geq 1}$ une suite i.i.d. telle que

$$\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 2) = \frac{1}{2}.$$

On note $Z_n = \prod_{k=1}^n X_k$, et on pose $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$.

a) Montrer que $(Z_n, \mathcal{F}_n)_{n\geq 1}$ est une martingale, qui converge presque sûrement vers une variable aléatoire $Z_\infty \in \mathcal{L}^1$.

b) Montrer que $Z_\infty = 0$ $\mathbb{P}$-p.s..

c) En déduire que (Z_n) ne converge pas vers Z_∞ dans $\mathcal{L}^1$.

Lois usuelles

I : Bernoulli : $\mathcal{L}(X) = B(1,p)$, Bernoulli de paramètre p, si $X \in \{0,1\}$ et $\mathbb{P}(X=1) = p$.
Espérance : p
Variance : $p(1-p)$
Fonction caractéristique : $1 - p + pe^{it}$

II : binomiale : $\mathcal{L}(X) = B(n,p)$ si $X \in \{0,\ldots,n\}$ et $\mathbb{P}(X=i) = C_n^i p^i (1-p)^{n-i}$. C'est la loi de $X_1 + \ldots + X_n$, où chaque X_i suit $B(1,p)$, et (X_i) indépendante.
Espérance : np
Variance : $np(1-p)$
Fonction caractéristique : $(1-p+pe^{it})^n$
Stabilité par convolution : $B(n,p) \star B(m,p) = B(n+m,p)$.
Limite : si $\mathcal{L}(X_n) = B(n,p_n)$ et il existe $\lambda \geq 0$ tel que $np_n \to \lambda$, alors $X_n \xrightarrow{\mathcal{L}} \mathcal{P}(\lambda)$; si X_n suit $B(n,p)$, alors $(X_n - np) \,/\, \sqrt{np(1-p)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

III : Poisson : $\mathcal{L}(X) = \mathcal{P}(\lambda)$ si $\mathbb{P}(X=k) = \lambda^k e^{-\lambda}/k!$.
Espérance : λ
Variance : λ
Fonction caractéristique : $\exp(\lambda(e^{it}-1))$
Stabilité par convolution : $\mathcal{P}(\lambda) \star \mathcal{P}(\beta) = \mathcal{P}(\lambda+\beta)$.
Limite : si $\mathcal{L}(X_\lambda) = \mathcal{P}(\lambda)$, $(X_\lambda - \lambda) \,/\, \sqrt{\lambda} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

IV : Multinomiale : $\mathcal{L}(X) = \mathcal{M}(n,p_1,\ldots,p_d)$ si $p_1 + \ldots + p_d = 1$, $0 \leq p_i \leq 1$, et si $n_1 + \ldots + n_d = n$, alors $\mathbb{P}(X = (n_1,\ldots,n_d)) = (n!) \,/\, (n_1! \ldots n_d!)$.
Espérance : $(np_1,\ldots,np_d)$
Covariance : $cov(X_i,X_j) = -np_ip_j$ si $i \neq j$
Variance : $var(X_i) = np_i(1-p_i)$
Fonction caractéristique : $(\sum_{1\leq j\leq d} p_j e^{ip_j})^n$
Si l'on dispose de n boules que l'on jette aléatoirement dans d boîtes distinctes, chaque boule ayant la probabilité p_i d'aller dans la $i^{\text{ième}}$, les nombres $(N_1,\ldots,N_d)$ de boules dans les boîtes $1,\ldots,d$ suivent une loi $\mathcal{M}(n,p_1,\ldots,p_d)$.

V : Hypergéométrique : $\mathcal{L}(X) = (N,n,p)$ si

$$\mathbb{P}(X=k) = \begin{cases} (C_{Np}^k C_{Nq}^{n-k}) \,/\, C_N^n \text{ pour } \max(0, n-N(1-p)) \leq k \leq \min(n,Np); \\ 0 \text{ sinon.} \end{cases}$$

Espérance : np
Variance : $np(1-p)(N-n) \,/\, (N-1)$
Si on tire n boules sans remise dans une urne en contenant N, une proportion p étant noires, $1-p$ blanches, le nombre de boules noires tirées suit une loi (N,n,p).

VI : Binomiale Négative : $\mathcal{L}(X)$ si $\mathbb{P}(X=k) = C_{n+k-1}^{n-1} p^n (1-p)^k$, $k \in \mathbb{N}$.

Espérance : $np \,/\, (1-p)$
Variance : $np \,/\, (1-p)^2$
Fonction caractéristique : $(p \,/\, (1-qe^{it}))^n$
Si $(X_i)_{i\geq 1}$ est i.i.d., $\mathcal{L}(X_1) = B(1,p)$, X_i représentant un succès si $X_i = 1$, le temps du $k^{\text{ième}}$ échec suit une loi binomiale négative de paramètres (n,p). Si $n = 1$, on parle aussi de loi géométrique.

VII : Uniforme continue : $\mathcal{L}(X) = \mathcal{U}_{[a,b]}$ si $\mathbb{P}_X$ a la densité $(b-a)^{-1}\mathbf{1}_{[a,b]}$.
Espérance : $(a+b)/2$
Variance : $(b-a)^2/12$
Fonction caractéristique : $(e^{itb} - e^{ita})/(it(b-a))$

VIII : Paréto : $\mathcal{L}(X)$ est Paréto de paramètre $p > 1$ si $\mathbb{P}_X$ a la densité $(p-1)x^{-p}\mathbf{1}_{[1,+\infty[}$.
Espérance : $(p-1)(p-2)$ si $p > 2$
Variance : $(p-1) \,/\, ((p-3)(p-2)^2)$ si $p > 3$

IX : Gamma : $\mathcal{L}(X) = \gamma(p,\theta)$ si $\mathbb{P}_X$ a la densité $\theta^p e^{-\theta x} x^{p-1}\Gamma(p)^{-1}\mathbf{1}_{[0,+\infty[}$, où $\Gamma(p) = \int_0^\infty e^{-\theta x} x^{p-1}$.
Espérance : p/θ
Variance : p/θ^2
Fonction caractéristique : $(\theta \,/\, (\theta - it))^p$
Stabilité par convolution : $\gamma(p,\theta) \star \gamma(q,\theta) = \gamma(p+q,\theta)$
Limite : si $(X-p)/\sqrt{p} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ si $p \to \infty$ et $\mathcal{L}(X) = \gamma(p,1)(:= \gamma(p))$.

X : Béta : $\mathcal{L}(X) = \beta(p,q)$ est Paréto de paramètre $p > 1$ si $\mathbb{P}_X$ a la densité $x^{p-1}(1-x)^{q-1}\Gamma(p+q)\Gamma(p)^{-1}\Gamma(q)^{-1}\mathbf{1}_{[0,1]}$.
Espérance : $p/(p+q)$
Variance : $pq/(p+q)^2(p+q+1)$
Stabilité : si $X \perp Y$ et si $\mathcal{L}(X) = \gamma(p)$ et $\mathcal{L}(Y) = \gamma(q)$, alors $\mathcal{L}(X/(X+Y)) = \beta(p,q)$, et $X/(X+Y)$ est indépendante de $X+Y$.

XI : Laplace : $\mathcal{L}(X)$ est Laplace si $\mathbb{P}_X$ a la densité $e^{-|x|}/2$.
Espérance : 0
Variance : 1
Fonction caratérsitique : $1/(1+t^2)$

XII : Cauchy : $\mathcal{L}(X) = \mathcal{C}(\alpha,\beta)$ si $\mathbb{P}_X$ a la densité $\beta \,/\, (\pi(\beta^2 + (x-\alpha)^2)$.
Espérance :
Variance :
Fonction caractéristique : $e^{it\alpha - \beta|t|}$

XIII : Loi Normale : $\mathcal{L}(X) = \mathcal{N}(a,A)$ si $X = (X_1,\ldots,X_n)$, si $A > 0$, et $\mathbb{P}_X$ a une densité $\exp(- < A^{-1}(x-a), x-a >) \,/\, ((2\pi)^{n/2}\sqrt{det(A)})$.
Espérance : a
Variance-covariance : A
Fonction caractéristique : $\exp(i < a,t > -\frac{1}{2} < At,t >^2)$

Stabilité : $\mathcal{N}(a, A) \star \mathcal{N}(b, B) = \mathcal{N}(a + b, A + B)$

XIV : Chi-deux : $\mathcal{L}(X) = \chi^2(d)(= \gamma(d/2, 1/2))$ (loi du chi-deux à d degrés de liberté) si si $\mathbb{P}_X$ a une densité $x^{(d/2)-1}e^{-x/2} \;/\; (2^{d/2}\Gamma(d/2))\mathbf{1}_{[0,+\infty[}$.
Espérance : d
Variance-covariance : $2d$
Fonction caractéristique : $1 \;/\; (1 - 2it)^{d/2}$
Stabilité : c'est la loi de la somme de d v.a.r. i.i.d. de loi $\mathcal{N}(0, 1)$

XV : Student : $\mathcal{L}(X) = \mathcal{S}(d)$ (loi de Student à d degrés de liberté) si si $\mathbb{P}_X$ a une densité $\Gamma((d + 1)/2)(1 + x^2/d)^{-(d+1)/2} \;/\; (\sqrt{d}\Gamma(d/2)\Gamma(1/2))$.
Espérance : 0 si $d > 1$
Variance-covariance : $d/(d - 2)$ si $d > 2$
Fonction caractéristique : si $d = 1$, c'est la loi de Cauchy $C(0, 1)$
Stabilité : c'est la loi de $Y \;/\; (\sqrt{Z/d})$ où $Y \perp Z$, $\mathcal{L}(Y) = \mathcal{N}(0, 1)$ et $\mathcal{L}(Z) = \chi^2(d)$.

XVI : exponentielle, exponentielle positive de paramètre λ : c'est une loi de densité de la forme $Ke^{-\lambda|x|}$ sur la droite, ou $Ke^{-\lambda x}\mathbf{1}_{[0,+\infty[}(x)$ pour la positive.
Ses moments existent à tous les ordres, et ses paramètres sont faciles à calculer.

Mathématiques

MM3 Probabilités

20 Novembre 2000, 9h-12h

Amphi Cassini

Polycopié et notes de cours autorisés, autres interdits.
Calculettes interdites.

Les trois exercices et le problème sont indépendants les uns des autres.
Les parties A,B et C du problème peuvent être abordées sans que les deux autres ne l'aient été.
Les réponses concises et suffisamment argumentées sont préférables.

TROIS EXERCICES.

Premier exercice.

Soit (X,Y) un couple de variables aléatoires réelles. On note m_2 la mesure de Lebesgue sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, et m la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On suppose que $\mathbb{P}_{(X,Y)} \ll m_2$, et on note

$$\frac{d\mathbb{P}_{(X,Y)}}{dm_2}(x,y) = f(x,y)$$

sa densité.

1-1 : montrer que $\mathbb{P}_X \ll m$ et que $\mathbb{P}_Y \ll m$.

1-2 : exprimer $\frac{d\mathbb{P}_X}{dm}(x)$ sous la forme d'une intégrale faisant intervenir f. Faire de même en ce qui concerne $\frac{d\mathbb{P}_Y}{dm}(y)$.

1-3 : soit $A \in \mathcal{B}(\mathbb{R}^2)$. On note $A_y = \{x \in \mathbb{R} : (x,y) \in A\}$. On rappelle que m presque sûrement en y, $A_y \in \mathcal{B}(\mathbb{R})$, et que $y \mapsto m(A_y)$ est mesurable, et que $m_2(A) = \int m(A_y)dy$.

Soient μ et ν deux mesures de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, absolument continues relativement à la mesure de Lebesgue m. Montrer qu'alors $\mu \otimes \nu \ll m_2$, et déterminer $\frac{d(\mu\otimes\nu)}{dm_2}$ en fonction de $\frac{d\mu}{dm}$ et de $\frac{d\nu}{dm}$.

1-4 : en déduire une CNS pour que $X \perp\!\!\!\perp Y$, concernant la forme algébrique de f.

1-5 : on suppose que la loi d'un couple de v.a.r. (X,Y), $\mathbb{P}_{(X,Y)}$, a une densité $f(x,y) = \frac{1}{\pi(1+x^2)} \frac{1}{\pi(1+y^2)}$ relativement à m_2. Les variables X et Y sont-elles indépendantes ?

Second exercice.

Soit (X,Y) un couple de variables aléatoires réelles tel que pour chaque $A \in \mathcal{B}(\mathbb{R}^2)$,

$$\mathbb{P}((X,Y) \in A) = \sum_{n\geq 1} \frac{1}{2^{n+1}} \mathbf{1}_A(n,n) + C \int_A (1 - x^2y^2) \mathbf{1}_{[0,1]^2}(x,y) dm_2(x,y),$$

où C est une constante positive.

2-1 : calculer C.

2-2 : déterminer $\mathbb{P}_X$, puis $\mathbb{P}_Y$ sans calcul supplémentaire.

2-3 : déterminer la décomposition de Lebesgue de $\mathbb{P}_X$ relativement à m.

2-4 : X et Y sont-elles indépendantes ?

2-5 : montrer que X est intégrable et calculer $\mathbb{E}(X)$.

Troisième exercice.

Soit $(A_n)_{n\geq 1}$ une suite d'évènements mesurables d'un espace probabilisé $(\Omega, \mathcal{B}, \mathbb{P})$, telle que $\mathbb{P}(A_1) > 0$. On note

$$N_n = \mathbf{1}_{A_1} + \ldots + \mathbf{1}_{A_n}, \text{ et } \theta_n = \frac{\sum_{j,k\leq n} \mathbb{P}(A_j \cap A_k)}{\left(\sum_{k\leq n} \mathbb{P}(A_k)\right)^2}, \quad n \geq 1.$$

3-1 : montrer que $\limsup_n A_n = \{\lim_n \uparrow N_n = +\infty\}$.

3-2 : on note $p_k = \mathbb{P}(A_k)$ et $m_n = p_1 + \ldots + p_n$. Calculer $Var(N_n)$ en fonction de θ_n et de m_n.

3-3 : soit $x \in \mathbb{R}$ tel que $m_n > x$. Montrer que $\mathbb{P}(N_n \leq x) \leq \mathbb{P}(|N_n - m_n| \geq m_n - x)$, et en déduire que $\mathbb{P}(N_n \leq x) \leq \frac{(\theta_n - 1)m_n^2}{(m_n - x)^2}$.

3-4 : on suppose désormais que $\sum_n p_n = +\infty$, et que $\liminf_n \theta_n \leq 1$. Montrer qu'alors quel que soit $x \in \mathbb{R}$, $\liminf_n \mathbb{P}(N_n \leq x) = 0$.

3-5 : montrer que $\mathbb{P}(\sup_k N_k \leq x) \leq \mathbb{P}(N_n \leq x)$. En déduire que $\mathbb{P}(\limsup_n A_n) = 1$.

3-6 : déduire de ce qui précède que le second Lemme de Borel-Cantelli est encore valable si l'on suppose les A_i deux à deux indépendants.

PROBLÈME : GRANDES DÉVIATIONS.

A : préliminaires.

Soit Y une v.a.r. **simple** prenant les valeurs y_j avec les probabilités p_j. On note $M_Y(t) = \mathbb{E}(e^{tY})$ sa fonction génératrice des moments. Nous supposons dans cette partie que

$$\mathbb{E}(Y) < 0, \qquad \mathbb{P}(Y > 0) > 0.$$

a : montrer que $M'_Y(0) < 0$, après en avoir justifié l'existence. Montrer aussi que $\lim_{+\infty} M_Y(t) = +\infty$, que M_Y est une fonction convexe de t, et que $M_Y(0) = 1$.

En déduire l'existence de $\tau > 0$ et $0 < \rho < 1$ tels que $M_Y(\tau) = \inf_t M_Y(t) = \rho$.

b : soit Z une v.a.r. simple prenant les mêmes valeurs que Y et telle que $\mathbb{P}(Z = y_j) = \frac{e^{\tau y_j}}{\rho}\mathbb{P}(Y = y_j)$. Montrer que

$$\begin{cases} M_Z(t) = \frac{M_Y(t+\tau)}{\rho}; \\ \mathbb{E}(Z) = \frac{M'_Y(\tau)}{\rho} = 0; \\ s^2 := \mathbb{E}(Z^2) = \frac{M''_Y(\tau)}{\rho} > 0. \end{cases}$$

c : montrer que pour chaque $t \geq 0$, $\mathbb{P}(Y \geq 0) \leq M(t)$, et en déduire que $\mathbb{P}(Y \geq 0) \leq \rho$.

d : notons Σ' la somme sur les indices j tels que $y_j \geq 0$. Montrer que $\mathbb{P}(Y \geq 0) = \Sigma' p_j = \rho\left(\Sigma' e^{-\tau y_j}\mathbb{P}(Z = y_j)\right)$. Poser $\mathbb{P}(Y \geq 0) = \rho e^{-\theta}$, et $\mathbb{P}(Z \geq 0) = p$. Montrer que $\theta \geq 0$. Justifier les égalités et inégalités suivantes ("ln" est concave) :

$$\begin{aligned} -\theta &= \ln\left(\Sigma' e^{-\tau y_j}\tfrac{\mathbb{P}(Z=y_j)}{p}\right) + \ln p \\ &\geq \left(\Sigma'(-\tau y_j)\tfrac{\mathbb{P}(Z=y_j)}{p}\right) + \ln p \\ &= \tfrac{-\tau s}{p}\left(\Sigma' \tfrac{y_j}{s}\mathbb{P}(Z = y_j)\right) + \ln p. \end{aligned}$$

e : utiliser l'inégalité de Cauchy-Schwarz pour montrer que $\Sigma' \frac{y_j}{s}\mathbb{P}(Z = y_j) \leq 1$.

f : montrer que

$$0 \leq \theta \leq \frac{\tau s}{\mathbb{P}(Z \geq 0)} - \ln \mathbb{P}(Z \geq 0). \qquad (\star)$$

g : soit W une v.a.r. telle que $\mathbb{E}(|W|^4) < \infty$. Supposons que $\mathbb{E}(W) = 0$, $l^2 = \mathbb{E}(W^2)$, et que $\mathbb{E}(W^4) = \xi^4 > 0$. Posons $q = \mathbb{P}(W \geq 0)$.

g1 : poser $W^+ = W\mathbf{1}_{W\geq 0}$ et $W^- = -W\mathbf{1}_{W<0}$. Montrer que $\mathbb{E}((W^+)^2) \leq \sqrt{q}\xi^2$.

g2 : avec les exposants $(\frac{3}{2}, 3)$, utiliser l'inégalité de Hölder pour montrer que $\mathbb{E}((W^-)^2) \leq \mathbb{E}(W^-)^{\frac{2}{3}}\xi^{\frac{4}{3}}$.

g3 : utiliser l'inégalité de Hölder pour les exposants $(4, \frac{4}{3})$ pour montrer que $\mathbb{E}(W^-) \leq q^{\frac{3}{4}}\xi$.

g4 : utiliser **g1-3** et $l^2 = \mathbb{E}((W^-)^2) + \mathbb{E}((W^+)^2)$, pour montrer que

$$\mathbb{P}(W \geq 0) \geq \frac{l^4}{4\xi^4}. \qquad (\star\star)$$

B : Théorème de Chernoff.

Soit $(X_n)_{n\geq 1}$ une suite i.i.d. de v.a.r. simples, telle que $\mathbb{E}(X_1) < 0$, $\mathbb{P}(X_1 > 0) > 0$. On note $M(t) = M_{X_1}(t)$. D'après **A**, il existe $\tau > 0$ et $0 < \rho < 1$ tels que $\rho = M(\tau) = \inf M(t)$. On pose $Y_n = X_1 + \ldots + X_n$, et $M_n(t) = M_{Y_n}(t)$, $n \geq 1$.

h : montrer que $\mathbb{E}(Y_n) < 0$ et $\mathbb{P}(Y_n > 0) > 0$. En déduire que $\inf M_n(t) = M(\tau)^n = \rho^n$.

i : soit S_n l'analogue de la v.a.r. Z de la partie **A**, associée ici à Y_n. Montrer que $M_{S_n}(t) = \left(\frac{M(t+\tau)}{\rho}\right)^n$.

j : soit Z_j l'analogue de la v.a.r. Z de la partie **A**, associée ici à X_j. Alors $\mathbb{E}(Z_j) = 0$, $\mathbb{E}(Z_j^2) := \sigma^2 > 0$, notons $\xi^4 = \mathbb{E}(Z_j^4) > 0$, et $M_{Z_1}(t) = \frac{M(t+\tau)}{\rho}$.

Montrer que l'on peut supposer la suite (Z_j) indépendante, et qu'en posant $V_n = Z_1 + \ldots + Z_n$, alors $M_{V_n} = M_{S_n}$.

k : en déduire que $\mathbb{E}(S_n) = 0$, $\mathbb{E}(S_n^2) = n\sigma^2$, et que $\xi_n^4 = \mathbb{E}(S_n^4) = n\xi^4 + 3n(n-1)\sigma^4$.

l : en déduire qu'il existe $\alpha > 0$ tel que pour chaque $n \geq 1$, $\frac{s_n^4}{4\xi_n^4} \geq \alpha$.

m : avec $(\star)$ et $(\star\star)$, en déduire l'existence d'un réel θ_n tel que $\mathbb{P}(Y_n \geq 0) = \rho^n e^{-\theta_n}$ et

$$0 \leq \theta_n \leq \frac{\tau\sigma\sqrt{n}}{\alpha} - \ln\alpha.$$

n : en déduire le théorème de Chernoff, à savoir

$$\left|\frac{1}{n}\ln\mathbb{P}(X_1 + \ldots + X_n \geq 0) - \ln\rho\right| = O(\frac{1}{\sqrt{n}}). \qquad (\star\star\star)$$

$\boxed{\textbf{C : application}}$

On se donne les mêmes notations que dans la partie **B**, avec en outre X_1 telle que $\mathbb{P}(X_1 = -1) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$.

o : montrer que $\mathbb{P}(Y_n \geq n\beta) \to 0$ lorsque $n \to \infty$ si $\beta > 0$ (*on pourra distinguer les cas* $\beta < 1$, $\beta = 1$, *et* $\beta > 1$).

p : montrer que pour $1 \geq \beta > 0$,

$$\lim_n \frac{1}{n}\ln\mathbb{P}(Y_n \geq n\beta) = \ln\left(\frac{(1-\beta)^{\frac{(\beta-1)}{2}}}{(1+\beta)^{\frac{(\beta+1)}{2}}}\right).$$

Mathématiques

MM3 Probabilités

26 Janvier 2001, 13h30–17h30

Polycopié et notes de cours autorisés, autres interdits. Calculettes interdites.

Les six exercices sont indépendants les uns des autres.
Une argumentation minimale est impérative.

Premier exercice.

"*Une formule d'inversion de la transformée de Laplace*".

a) Soit $\theta, \lambda > 0$ et supposons que la loi de la variable X_λ soit une loi de Poisson de paramètre $\lambda\theta$, $\mathcal{L}(X_\lambda) = \mathcal{P}(\lambda\theta)$. Montrer que $\frac{X_\lambda - \lambda\theta}{\lambda} \xrightarrow{\mathbb{P}} 0$ si $\lambda \to \infty$. En déduire que

$$\lim_{\lambda\to\infty} e^{-\lambda\theta} \sum_{k\leq\lambda x} \frac{(\lambda\theta)^k}{k!} = \begin{cases} 0 \text{ si } \theta > x; \\ 1 \text{ si } \theta < x. \end{cases}$$

b) Soit $\mathbb{P}$ une loi de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, portée par $\mathbb{R}^+$, et telle que pour chaque $n \geq 1$, $\int_0^\infty x^n d\mathbb{P}(x) < +\infty$. Notons $L(t) := \int_0^\infty e^{-tx} d\mathbb{P}(x)$ sa transformée de Laplace. Montrer que $L : [0, +\infty[\to [0, +\infty[$ est bien définie, indéfiniment dérivable, et calculer $L^{(k)}(t)$ pour $t \geq 0$ et $k \geq 0$.

c) Déduire des questions précédentes que

$$\lim_{\lambda\to\infty} \sum_{k\leq\lambda x} \frac{(-1)^k}{k!} \lambda^k L^{(k)}(\lambda) = F_{\mathbb{P}}(x),$$

en tout point de continuité x de $F_{\mathbb{P}}$.

Second exercice.

"*Espérance conditionnelle et projecteurs L^2*".

Soit $(\Omega, \mathcal{B}, \mathbb{P})$ un espace probabilisé, et soit $\mathcal{D} \subset \mathcal{B}$ une sous-σ-algèbre. On note $\mathcal{L}^2(\mathcal{D})$ l'ensemble défini par

$$\mathcal{L}^2(\mathcal{D}) = \{X \in \mathcal{L}^2(\mathcal{B}) : X \text{ est } \mathcal{D}\text{-mesurable}\}.$$

On note aussi

$$L^2(\mathcal{D}) = \{\bar{X} \in L^2(\mathcal{B}) : \exists X \in \bar{X} \cap \mathcal{L}^2(\mathcal{D})\},$$

où $\bar{X} = \{Y : X = Y \ \mathbb{P} - p.s.\}$ est la classe d'équivalence de X modulo $\mathbb{P}$.

$\boxed{\textbf{a)}}$ Montrer que $L^2(\mathcal{D})$ est un sous-espace vectoriel fermé de l'espace de Hilbert $(L^2(\mathcal{B}), \|\ \|_2)$.

$\boxed{\textbf{b)}}$ Montrer que si $X \in \mathcal{L}^2(\mathcal{B})$, alors $\mathbb{E}(X|\mathcal{D}) \in \mathcal{L}^2(\mathcal{D})$.

$\boxed{\textbf{c)}}$ Montrer que si $Z \in \mathcal{L}^2(\mathcal{D})$, alors

$$\mathbb{E}(Z(X - \mathbb{E}(X|\mathcal{D})) = 0.$$

$\boxed{\textbf{d)}}$ On rappelle que si $(H, \|\ \|)$ est un espace de Hilbert, avec un produit scalaire noté $< h, h' >$ (et donc $|h| = \sqrt{< h, h >}$) et si $K \subset H$ est un sous-espace vectoriel fermé, alors le projecteur orthogonal $p_K : H \to K$ est l'application linéaire continue définie par l'identité

$$p_K(x) = \tilde{x} \Leftrightarrow |x - \tilde{x}| = \min\{|x - k| : k \in K\} \Leftrightarrow \forall k \in K,\ < x - \tilde{x}, k >= 0.$$

On rappelle aussi que dans $L^2(\mathcal{B})$, $< \bar{X}, \bar{Y} >= \mathbb{E}(XY)$.

Montrer que

$$\overline{\mathbb{E}(X|\mathcal{D})} = p_{L^2(\mathcal{D})}(\bar{X}),$$

autrement dit que pour les variables $\mathcal{L}^2$, l'espérance conditionnelle relativement à $\mathcal{D}$ coïncide avec la projection orthogonale de $L^2(\mathcal{B})$ sur $L^2(\mathcal{D})$, modulo $\mathbb{P}$.

$\boxed{\textbf{e)}}$ Applications : avec les mêmes hypothèses, on suppose que $\mathcal{C} \subset \mathcal{D}$ est une troisième sous-σ-algèbre.

$\boxed{\textbf{e1)}}$ Montrer que

$$\mathbb{E}((X - \mathbb{E}(X|\mathcal{D}))^2) + \mathbb{E}((\mathbb{E}(X|\mathcal{C}) - \mathbb{E}(X|\mathcal{D}))^2) = \mathbb{E}((X - \mathbb{E}(X|\mathcal{C})^2)$$

(on pourra remarquer que $L^2(\mathcal{C}) \subset L^2(\mathcal{D}) \subset L^2(\mathcal{B})$).

On choisit $\mathcal{C} = \{\emptyset, \Omega\}$. On pose $var(X|\mathcal{D}) := \mathbb{E}(X^2|\mathcal{D}) - \mathbb{E}(X|\mathcal{D})^2$. Montrer que

$$var(X) = \mathbb{E}(var(X|\mathcal{D})) + var(\mathbb{E}(X|\mathcal{D})).$$

$\boxed{\textbf{e2)}}$ Montrer que si $X, Y \in \mathcal{L}^2$ sont deux variables aléatoires telles que $\mathbb{E}(Y|\mathcal{D}) = X$, et que $\mathbb{E}(X^2) = \mathbb{E}(Y^2)$, alors $X = Y$ $\mathbb{P}$-p.s..

Troisième exercice.

"*Quelques martingales ?* ".

Soit $X_1, X_2, \ldots$ une suite de variables aléatoires indépendantes et de même loi. On considère la filtration naturelle $\mathcal{G}_n = \sigma(X_1, ...X_n)$. Lesquels des processus suivants sont des martingales par rapport à $(\mathcal{G}_n, n \geq 1)$ (discuter selon les valeurs de λ et les paramètres de X_1) ?

$\boxed{\textbf{a)}}$ $S_n = X_1 + \ldots + X_n$ si $\mathbb{E}|X_1|$ est finie, $n \geq 1$.

$\boxed{\textbf{b)}}$ $X_1^2 + \ldots + X_n^2 - n\lambda$ si $\mathbb{E}|X_1|^2$ est finie et λ est réel, $n \geq 1$.

$\boxed{\textbf{c)}}$ $e^{(\alpha S_n - n\lambda)}$ si $\mathbb{E}(e^{\alpha X_1})$ est finie pour tout $\alpha \in \mathbb{R}$, $n \geq 1$.

$\boxed{\textbf{d)}}$ $Y_n = \prod_{k=1}^{n} X_k$ si $X_1 \in \mathcal{L}^\infty$, $n \geq 1$.

Quatrième exercice.

"*Marche aléatoire*".

Le temps t est un entier positif ou nul. Un individu déambule "aléatoirement" sur $\mathbb{Z}$, en partant de 0 à l'instant initial 0. On suppose qu'à l'instant n, si sa position est S_n, sa position à l'instant $n+1$ sera $S_n + X_{n+1} = S_{n+1}$, où $(X_n)_{n \geq 1}$ est une suite i.i.d. de variables telles que

$$\mathbb{P}(X_n = 1) = p, \ \mathbb{P}(X_n = -1) = 1 - p,$$

et $0 < p < 1$ est un paramètre fixé du problème. Nous nous proposons d'étudier l'ensemble

$$R = \{\omega : \#\{n : S_n(\omega) = 0\} = +\infty\},$$

c'est à dire l'ensemble des marches ou déambulations qui reviennent une infinité de fois à l'origine.

$\boxed{\textbf{a)}}$ Exprimer S_n à l'aide des X_p pour $n \geq 1$, et en déduire que $(\frac{S_n}{n})$ converge presque sûrement, vers une limite que l'on précisera.

$\boxed{\textbf{b)}}$ Montrer que $\mathbb{P}(R) = 0$ si $p \neq \frac{1}{2}$.

$\boxed{\textbf{c)}}$ On suppose désormais que $p = \frac{1}{2}$.

$\boxed{\textbf{c1)}}$ Pour tout $k \geq 0$, on pose $Z_k = (S_{2^{k+1}} - S_{2^k}) \ / \ \sqrt{2^k}$. Montrer que Z_k a la même loi que $S_{2^k} \ / \ \sqrt{2^k}$, et en déduire, à l'aide du théorème limite centrale, que pour tout réel M,

$$\sum_k \mathbb{P}(Z_k \geq M) = +\infty.$$

$\boxed{\textbf{c2)}}$ En déduire que pour tout réel M, $\mathbb{P}(\sup_k Z_k \geq M) = 1$, puis que $\mathbb{P}(\sup_k |Z_k| = +\infty) = 1$. En déduire que

$$\mathbb{P}(\sup_n \left|\frac{S_n}{\sqrt{n}}\right| = +\infty) = 1.$$

$\boxed{\textbf{c3)}}$ Utiliser la loi $0 - 1$ de Kolmogorov pour montrer que l'évènement $B^+ = \{\sup_n \frac{S_n}{\sqrt{n}} = +\infty\}$ a une probabilité nulle ou pleine. Démontrer que $\mathbb{P}(B^+) = \mathbb{P}(B^-)$, où $B^- = \{\inf_n \frac{S_n}{\sqrt{n}} = -\infty\}$. Utiliser c2) pour en conclure que

$$\mathbb{P}(B^+) = \mathbb{P}(B^-) = 1.$$

$\boxed{\textbf{c4)}}$ Conclure en montrant que $\mathbb{P}(R) = 1$, autrement dit que la marche aléatoire simple symétrique revient presque sûrement une infinité de fois en 0.

Cinquième exercice.

"*Un calcul de loi conditionnelle*".

Soit (X, Y) un couple de v.a.r., de loi sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ décrite par

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}((X,Y) \in A) = \sum_{n \geq 1} c_n \int_A x^n \mathbf{1}_{[0,1]}(x) dm \otimes d\delta_n(x,y),$$

où $(c_n)_{n\geq 1}$ est une suite de constantes positives, m désigne la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, et δ_n désigne la mesure de Dirac en n ($\delta_n(B) = \mathbf{1}_B(n),\ B \in \mathcal{B}(\mathbb{R})$).

a) Montrer que $\mathbb{P}_{(X,Y)}$ est portée par $\cup_{n\geq 1}([0,1] \times \{n\})$.

b) Exhiber une suite $(c_n)_{n\geq 1}$ de réels strictement positifs pour laquelle la loi décrite ci-dessus est effectivement une loi de probabilité.

c) Montrer que $\mathbb{P}_X \ll m$, et calculer $\frac{d\mathbb{P}_X}{dm}$. Montrer que X a des moments de tous ordres.

d) Décrire une version régulière de la loi de X sachant Y, $\mathbb{P}_X(\cdot|Y)$.

e) Calculer $\mathbb{E}(X^2|Y)$.

Sixième exercice.

"*Quelques petites questions*".

a) Soit $X_1, X_2, \ldots$ une suite de variables aléatoires indépendantes de même loi.

Montrer que $\frac{X_n}{n} \to 0$ p.s. lorsque $n \to \infty$ si et seulement si $\mathbb{E}|X_1| < \infty$.

Montrer que $\frac{X_n}{n} \xrightarrow{\mathbb{P}} 0$ lorsque $n \to \infty$.

b) Soit $Y_1, Y_2, \ldots$ une suite de variables aléatoires suivant une loi de Poisson de paramètre $\lambda > 0$.

Montrer que $\frac{Y_n}{\sqrt{n}} \to 0$ p.s. lorsque $n \to \infty$.

c) Soit $(Y_i)_{i\geq 1}$ une suite de v.a.r. i.i.d. dans $\mathcal{L}^2$, de moyenne μ et de variance σ^2. Soit N une variable aléatoire $\mathcal{L}^2$, indépendante de la suite (Y_i), et prenant des valeurs entières positives. On pose

$$X = Y_1 + \ldots + Y_N.$$

Montrer que $X \in \mathcal{L}^2$ et que

$$var(X) = \sigma^2 \mathbb{E}(N) + \mu^2 var(N).$$

d) Soient Z une v.a. $\mathcal{L}^1$ centrée et $c > 0$ tels que $\mathbb{P}(|Z| \leq c) = 1$. Soit $\varepsilon > 0$. Montrer que

$$1 - \frac{(\varepsilon + c)^2}{var(Z)} \leq \mathbb{P}(|Z| \geq \varepsilon).$$

UNIVERSITÉ de Picardie Jules Verne

Mathématiques

MM3 Probabilités

5 Septembre 2001, 9h – 13h

Polycopié et notes de cours autorisés, autres interdits. Calculettes interdites.

Une argumentation minimale est impérative.

Dans tout ce qui suit $(\Omega, \mathcal{B}, \mathbb{P})$ désigne un espace probabilisé.

Premier exercice.

"*Espérance conditionnelle*".

a) : on choisit $(\Omega, \mathcal{B}, \mathbb{P}) = ([0,1], \mathcal{B}, m)$ où m désigne la mesure de Lebesgue et $\mathcal{B}$ la tribu borélienne. On pose

$$\begin{cases} X = \mathbf{1}_{[0,\frac{1}{2}[}, \\ Y = \frac{1}{3}\mathbf{1}_{[0,\frac{1}{3}[} + \frac{1}{2}\mathbf{1}_{[\frac{1}{3},\frac{2}{3}[} + \mathbf{1}_{[\frac{2}{3},1]}. \end{cases}$$

Déterminer la loi du couple (X, Y). Déterminer $\mathbb{E}(X|Y)$ et $\mathbb{E}(Y|X)$.

b) : soit (X, Y) un couple de var tel que $X > 0$, $Y \geq 0$ et

$$\mathbb{P}((X = k \text{ et } Y \leq t) = C\frac{1 - e^{-k^2 t}}{k^2}, \qquad k > 0,\ t \in \mathbb{R}^+.$$

Calculer $\mathbb{E}(X|Y)$ et $\mathbb{E}(Y|X)$.

Second exercice.

"*Loi de Poisson - Loi géométrique*".

a) : soient X et Y deux var indépendantes de lois de Poisson de paramètres respectifs λ et β. Montrer que $\mathcal{L}(X+Y) = \mathcal{P}(\lambda + \beta)$. Puis montrer que $\mathcal{L}(X|X+Y=n)$ est binomiale, et en déterminer les paramètres.

b) : si X est une var de loi géométrique, à valeurs dans $\mathbb{N}$, montrer que quels que soient $k \geq 1$ et $n \geq 0$,

$$\mathbb{P}(X = n + k | X \geq n) = \mathbb{P}(X = k).$$

Pourquoi appelle-t-on cette propriété la propriété d'"*absence de mémoire*" ?

Existe-t-il une autre distribution sur $\mathbb{N}$ ayant cette propriété ?

Troisième exercice.

"*Convergences*".

a) : mettons que les catastrophes se produisent aux temps $T_1, \ldots, T_n, \ldots$, où $T_n = X_1 + \ldots + X_n$, et la suite $(X_i)_{i \geq 1}$ est une suite i.i.d. de var positives.

Soit $N(t) = \max\{n : T_n \leq t\}$, le nombre de catastrophes déjà produites au temps t. Montrer que si $\mathbb{E}(X_1) < \infty$, alors $N(t) \to \infty$ et $N(t)/t \to 1/\mathbb{E}(X_1)$ $\mathbb{P}$-p.s..

$\boxed{\textbf{b)}}$: soit $(u_n)_{n\geq 1}$ une suite dans $]0, 1/2[$. Soit $(X_n)_{n\geq 1}$ une suite indépendante de var telle que X_n suive la loi définie par

$$\mathbb{P}(X_n = 0) = 1 - u_n \qquad \text{et} \qquad \mathbb{P}(X_n = 1/u_n) = u_n.$$

Montrer qu'alors

$\boxed{\textbf{b1)}}$: $\mathbb{E}(X_n) = 1$;

$\boxed{\textbf{b2)}}$: $X_n \xrightarrow{\mathbb{P}} 0$ si et seulement si $u_n \to 0$;

$\boxed{\textbf{b3)}}$: $X_n \overset{p.s.}{\to} 0$ si et seulement si $\sum_{n\geq 1} u_n < \infty$.

On remarque que dans le cas b3), $n^{-1}(\sum_{k=1}^n X_k) \to 0$ $\mathbb{P}$-p.s., mais que $\mathbb{E}(X_1) = 1$.

$\boxed{\textbf{c)}}$: soit $(X_n)_{n\geq 1}$ une suite de var iid dans L^2. On pose $m = \mathbb{E}(X_1)$ et $\sigma^2 = var(X_1)$. Pour $n \geq 2$, on pose

$$Y_n = \frac{1}{n}\sum_{k=1}^n X_k \ \text{ et } \ Z_n = \frac{1}{n-1}\sum_{k=1}^n (X_k - Y_n)^2.$$

$\boxed{\textbf{c1)}}$: calculer $\mathbb{E}(Z_n)$;

$\boxed{\textbf{c2)}}$: montrer que $Z_n \overset{p.s}{\to} \sigma^2$ lorsque $n \to \infty$.

$\boxed{\textbf{d)}}$: soit $(X_i)_{i\geq 1}$ une suite i.i.d. de var à valeurs dans $\mathbb{Z}$, d'espérance finie.

On pose $S_0 = 0$ et $S_n = X_1 + \ldots + X_n$, $n \geq 1$. Montrer que si $\mathbb{E}(X_1) \neq 0$, alors $\mathbb{P}(\limsup\{S_n = 0\}) = 0$.

Quatrième exercice.

"*Martingales*".

Démontrer ou donner un contre-exemple :

$\boxed{\textbf{a)}}$: si $(M_n; n \geq 1)$ est une sous-martingale alors $(M_n^2; n \geq 1)$ est une sous-martingale.

$\boxed{\textbf{b)}}$: si $(M_n; n \geq 1)$ est une sous-martingale positive et $t > 0$, alors

$$\mathbb{P}\big(\max_{1\leq j\leq n} M_j > x\big) \leq e^{-tx} E(e^{tM_n}).$$

Cinquième exercice.

"*Loi de Cauchy*".

Soit $(X_n)_{n\geq 1}$ une suite de var iid de loi commune la loi de Cauchy de paramètre $a > 0$ (densité $c_a(x) = \frac{1}{\pi}\frac{a}{a^2+x^2}$, $x \in \mathbb{R}$).

$\boxed{\textbf{a)}}$: montrer que $\frac{1}{n}\sum_{k=1}^n X_k$ diverge presque sûrement.

$\boxed{\textbf{b)}}$: montrer que si $x > 0$, $\mathbb{P}(\frac{1}{n}\max_{k\leq n} X_k \leq x) \to e^{-a/\pi x}$.

Sixième exercice.

"*Convergence p.s. et convergence en probabilité*".

Soient $(Z_k)_{k\geq 1}$ et Z des var. Soit $\varepsilon > 0$ et notons

$$A_{n,m}(\varepsilon) = \{|Z_k - Z| < \varepsilon : n \leq k \leq m\}.$$

a) : montrer que $Z_k \to Z$ $\mathbb{P}$-p.s. si et seulement si pour tout $\varepsilon > 0$,

$$\lim_n \lim_m \mathbb{P}(A_{n,m}(\varepsilon)) = 1.$$

b) : montrer que $Z_k \to Z$ en probabilité si et seulement si $\lim_n \mathbb{P}(A_{n,n}(\varepsilon)) = 1$ pour tout $\varepsilon > 0$.

c) : trouver $(Z_k)_{k\geq 1}$ et Z telles que $Z_k \to Z$ en probabilité mais pas presque sûrement.

Septième exercice.

"*Intégration des var*".

a) : soient X et Y deux var positives, telles que $\mathbb{E}(X) = +\infty$ et $\mathbb{E}(Y) = 0$. Sont-elles indépendantes ? A quelle condition de convention sur "$+\infty \times 0$" la formule $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ reste-elle valable ?

b) : soient X et Y deux var indépendantes. Soit $f : \mathbb{R}^2 \to \mathbb{R}^+$ borélienne. Notons $g(x) = \mathbb{E}(f(x, Y))$, $x \in \mathbb{R}$. Montrer que $\mathbb{E}(g(X)) = \mathbb{E}(f(X, Y))$. Montrer plus généralement que $\int_{X\in A} g(X)d\mathbb{P} = \int_{X\in A} f(X, Y)d\mathbb{P}$.

UNIVERSITÉ de Picardie Jules Verne

Mathématiques

MM3 Probabilités

15 Novembre 2001, 8h30–11h30

Salle H26

Polycopié de cours autorisé, autres documents interdits.

Calculettes interdites.

Les exercices sont indépendants les uns des autres.

La clarté, tant en matière de présentation que d'argumentation, est indispensable.

Premier exercice.

Soit (X, Y) un couple de variables aléatoires réelles. On note m_2 la mesure de Lebesgue sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, et m la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On note $\delta_n(B) = \mathbf{1}_B(n)$ pour $n \in \mathbb{N}^*$ et $B \in \mathcal{B}(\mathbb{R})$. On suppose que si $A \in \mathcal{B}(\mathbb{R}^2)$,

$$\mathbb{P}_{(X,Y)}(A) = C \sum_{n \geq 1} \frac{1}{2^n} \int \mathbf{1}_A(x, y) \mathbf{1}_{[0,1]}(y) y^n d(\delta_n \otimes m)(x, y),$$

où C est une constante positive.

1-1 : calculer C.

1-2 : déterminer la décomposition de Lebesgue de $\mathbb{P}_{(X,Y)}$ relativement à m_2.

1-3 : déterminer $\mathbb{P}_X$, $\mathbb{P}_Y$, et leurs décompositions de Lebesgue respectives relativement à m.

1-4 : X et Y sont elles indépendantes ?

1-5 : calculer, après en avoir justifié l'existence, la valeur de $\mathbb{E}(XY)$.

Second exercice.

Soient $(X_n)_{n \geq 1}$ et X des variables aléatoires.

2-1 : montrer que si $f : \mathbb{R} \to \mathbb{R}$ est continue et que si $X_n \to X$ $\mathbb{P}$-p.s., alors $f(X_n) \to f(X)$ $\mathbb{P}$-p.s..

2-2 : montrer que si $X_n, X \in \mathcal{L}^1(\mathbb{P})$, alors $\mathbb{E}(|X_n - X|) \to 0$ entraîne que $X_n \to X$ en probabilités.

2-3 : exhiber un contre-exemple à la réciproque naturelle de **2-2**.

Troisième exercice.

Soient X, Y deux variables aléatoires réelles, indépendantes, ayant chacune une densité relativement à m la mesure de Lebesgue sur $\mathbb{R}$: notons $f(x) = d\mathbb{P}_X \,/\, dm(x)$ et $g(y) = d\mathbb{P}_Y \,/\, dm(y)$ ces densités.

3-1 : montrer que $\mathbb{P}_{X+Y} \ll m$ et exprimer $\frac{d\mathbb{P}_{X+Y}}{dm}(t)$ en fonction de f et g.
3-2 : en déduire $\mathcal{L}(X+Y)$ si $\mathcal{L}(X) = \mathcal{N}(0,\sigma)$ et $\mathcal{L}(Y) = \mathcal{N}(0,\tau)$.

Quatrième exercice

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles. Montrer que la suite $(X_n)_{n\geq 1}$ est indépendante si et seulement si pour chaque n, $\sigma(X_1, \dots, X_{n-1})$ et $\sigma(X_n)$ le sont.

Cinquième exercice.

Soit X une variable aléatoire réelle, telle que $\mathbb{P}(X > 0) = 1$, et ayant une densité f relativement à la mesure de Lebesgue sur la droite.

Montrer que $\frac{1}{X}$ a une densité, et la calculer.

Sixième exercice.

Soit $(X_n)_{n\geq 1}$ une suite indépendante de variables aléatoires réelles, et soit Y une variable aléatoire réelle, qui soit $\sigma(X_n, X_{n+1}, \dots)$ mesurable, pour chaque $n \geq 1$.

Montrer qu'il existe une constante $a \in \mathbb{R}$ telle que $\mathbb{P}(Y = a) = 1$.

Septième exercice.

Supposons que X est une variable aléatoire réelle dont la fonction de répartition F_X est continue et strictement croissante.

Quelle est alors la loi de $F_X \circ X$?

Huitième exercice.

Soit $(\Omega, \mathcal{B})$ un espace mesurable, et soient μ et ν deux mesures de probabilités sur $(\Omega, \mathcal{B})$.

On dit que μ et ν sont équivalentes, noté $\mu \equiv \nu$, si à la fois $\nu \ll \mu$ et $\mu \ll \nu$. Montrer qu'alors

$$\frac{d\nu}{d\mu} = \mathbf{1}_{\frac{d\mu}{d\nu} > 0} \left(\frac{d\mu}{d\nu} \right)^{-1}.$$

Neuvième exercice.

Soit X une variable aléatoire réelle de loi $\mathbb{P}_X = \mathcal{N}(0,1)$. Notons, pour $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_{X|X\in[0,1]}(A) = \frac{\mathbb{P}_X(A \cap [0,1])}{\mathbb{P}_X([0,1])}.$$

9-1 : montrer que $\mathbb{P}_{X|X\in[0,1]}$ est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
9-2 : montrer que $\mathbb{P}_{X|X\in[0,1]} \ll \mathbb{P}_X$, et déterminer une version de $\frac{d\mathbb{P}_{X|X\in[0,1]}}{d\mathbb{P}_X}$.
9-3 : a-t-on $\mathbb{P}_{X|X\in[0,1]} \equiv \mathbb{P}_X$ au sens de l'exercice huit, ou pas ? Justifier.

<u>Tout document interdit, calculettes interdites.</u>

Les exercices 2,3,4,5,6 ne sont pas indépendants les uns des autres.
On pourra toutefois admettre les résultats de questions non traitées.

La clarté, tant en matière de présentation que d'argumentation, est indispensable.
Notamment, le détail concernant les "$\mathcal{O}$" est éxigé.

Premier exercice : augmenter une tribu.

Soit Ω un ensemble, et $\mathcal{B}$ une tribu de parties de Ω. Montrer que si $H \subset \Omega$,

$$\sigma(\mathcal{B} \cup \{H\}) = \{(A \cap H) \cup (B \cap H^c) : A, B \in \mathcal{B}\}.$$

Second exercice : éléments d'une algèbre.

Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$. On note $a(\mathcal{A})$ l'intersection des algèbres de parties de Ω qui contiennent $\mathcal{A}$.

2-1 : montrer que $a(\mathcal{A})$ est une algèbre qui contient $\mathcal{A}$, et que parmi celles qui la contiennent, c'est, au sens de l'inclusion, la plus petite.

2-2 : montrer que si $\mathcal{A} \neq \emptyset$, les éléments de $a(\mathcal{A})$ sont de la forme

$$\cup_{i=1}^{m} \left(\cap_{j=1}^{n_i} A_{i,j} \right),$$

où l'union est disjointe, et pour chaque i et j, $A_{i,j}$ ou $A_{i,j}^c$ est dans $\mathcal{A}$ (autrement dit que les éléments de $a(\mathcal{A})$ sont les unions d'atomes des partitions engendrées par des collections finies d'éléments de $\mathcal{A}$).

Troisième exercice : formule d'inclusion - exclusion.

Soit $\mathcal{A}$ une algèbre de parties de Ω, et soit $\mathbb{P} : \mathcal{A} \to [0,1]$ une fonction additive (non nécessairement σ-additive). Montrer par récurrence sur n que si $A_1, \ldots, A_n \in \mathcal{A}$, alors

$$\mathbb{P}(\cup_{i=1}^n A_i) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n+1} \mathbb{P}(\cap_{i=1}^n A_i).$$

Quatrième exercice : densités et formule d'Euler.

Désormais $\Omega = \mathbb{N}^* = \{1, 2, ...\}$, et $\mathcal{B} = \mathcal{P}(\Omega)$. Pour $n \geq 1$, on note $\mathbb{P}_n$ la probabilité définie sur $(\Omega, \mathcal{B})$ par

$$\mathbb{P}_n(A) = \frac{1}{n}\text{Card}(A \cap \{1, \dots, n\}).$$

Pour $A \in \mathcal{B}$, on note $D(A) = \lim_n \mathbb{P}_n(A)$, si la limite existe. On appelle la limite la densité de A, et on note $\mathcal{D}$ l'ensemble des A qui ont une densité.

4-1 : donner un exemple d'un A ayant une densité, et d'un autre n'en ayant pas.

4-2 : montrer que $\mathcal{D}$ est stable par unions finies disjointes, et que $D : \mathcal{D} \to [0, 1]$ est additive mais pas σ-additive.

4-3 : montrer les propriétés suivantes :

$$\begin{cases} \textbf{a)} : \emptyset, \Omega \in \mathcal{D}; \\ \textbf{b)} : A \in \mathcal{D} \Rightarrow A^c \in \mathcal{D}; \\ \textbf{c)} : A, B \in \mathcal{D},\ A \subset B \Rightarrow B \setminus A \in \mathcal{D}. \end{cases}$$

4-4 : montrer que $\mathcal{D}$ n'est pas stable par unions dénombrables disjointes.

4-5 : soit $a \geq 1$, et notons $M_a = \{ma : m \geq 1\}$. Soit $\mathcal{M} = \{M_a : a \geq 1\}$.

4-5-1 : montrer que $M_a \in \mathcal{D}$ et calculer $D(M_a)$.

4-5-2 : avec les notations du second exercice, montrer que $a(\mathcal{M}) \subset \mathcal{D}$, et que D sur $a(\mathcal{M})$ est entièrement déterminée par les valeurs $D(M_a)$ (*Indication* : on pourra montrer que si $a \vee b = ppcm(a, b)$, alors $M_a \cap M_b = M_{a \vee b}$. Puis en notant $\mathcal{N}_s$ la classe des ensembles de la forme $M_a \cap M_{b_1}^c \cap \ldots \cap M_{b_s}^c$, montrer que "$\mathcal{N}_{s+1} = \mathcal{N}_s \setminus \mathcal{N}_s$", puis conclure par induction sur s).

4-5-3 : on note $\phi(n)$ la fonction d'Euler : $\phi(n)$ désigne le nombre d'entiers inférieurs à l'entier positif n et premiers à n. Soient $p_1, \dots, p_r$ les diviseurs premiers de n. Montrer, à l'aide de la formule d'inclusion - exclusion du troisième exercice, que

$$\frac{\phi(n)}{n} = \prod_{i=1}^{r} \left(1 - \frac{1}{p_i}\right).$$

Cinquième exercice : puissance des facteurs premiers.

Pour un entier $m \geq 1$ et un nombre premier p, on note $\alpha_p(m)$ la puissance exacte à laquelle p apparaît dans la décomposition de m en facteurs premiers. On désignera toujours par p un nombre premier : on a donc

$$m = \prod_p p^{\alpha_p(m)}.$$

Notons encore $\delta_p(m) = 1$ si $p|m$, 0 sinon. Enfin, $[x]$ désigne la partie entière de x.

5-1 : soient $p_1, \dots, p_u$ des entiers premiers distincts. Montrer que

$$\begin{cases} \textbf{i)} : \mathbb{P}_n(\alpha_{p_i} \geq k_i : i \leq u) = \frac{1}{n}\left[\frac{n}{p_1^{k_1}\dots p_u^{k_u}}\right] \to \frac{1}{p_1^{k_1}\dots p_u^{k_u}}; \\ \textbf{ii)} : \mathbb{P}_n(\alpha_{p_i} = k_i : i \leq u) \to \prod_{i=1}^u \left(\frac{1}{p_i^{k_i}} - \frac{1}{p_i^{k_i+1}}\right); \\ \textbf{iii)} : \mathbb{P}_n(\delta_{p_i} = 1 : i \leq u) = \frac{1}{n}\left[\frac{n}{p_1\dots p_u}\right] \to \frac{1}{p_1\dots p_u}. \end{cases}$$

5-2 : montrer que si X est une v.a.r. prenant des valeurs entières positives ou nulles, alors intégrable ou pas,

$$\mathbb{E}(X) = \sum_{n\geq 1} \mathbb{P}(X \geq n). \tag{E}$$

5-3 : soit $f : \Omega \to \mathbb{R}$. On note, si $n \in \Omega$,

$$\mathbb{E}_n(f) = \frac{1}{n}\sum_{m=1}^n f(m) = \int_\Omega f(\omega)d\mathbb{P}_n(\omega).$$

Montrer avec (E) que

$$\mathbb{E}_n(\alpha_p) = \sum_{k\geq 1}\frac{1}{n}\left[\frac{n}{p^k}\right] \to \frac{1}{p-1}.$$

Sixième exercice : un petit théorème des nombres premiers.

Rappelons la formule de Stirling (polycopié p. 73)

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{S}$$

6-1 : utiliser (S) pour montrer que $\mathbb{E}_n(\log) = \log n + \mathcal{O}(1)$.

6-2 : montrer (cf. **5-2**) que $\mathbb{E}_n(\alpha_p) \leq \frac{2}{p}$. En utilisant la relation $\log m = \sum_p \alpha_p(m)\log p$, et **6-1**, en déduire que $\sum_p \frac{\log p}{p} = +\infty$, et donc qu'il existe une infinité de nombres premiers!

6-3 : notons $\log^* m = \sum_p \delta_p(m)\log p$. Montrer que $\mathbb{E}_n(\alpha_p - \delta_p) \leq \sum_{k\geq 2}\frac{1}{p^k}$ (on pourra utiliser **5-1** et (E)). En déduire que

$$\mathbb{E}_n(\log^*) = \sum_p \frac{1}{n}\left[\frac{n}{p}\right]\log p = \log n + \mathcal{O}(1), \tag{$F1$}$$

et donc que

$$\mathbb{E}_{2n}(\log^*) - \mathbb{E}_n(\log^*) = \mathcal{O}(1).$$

6-4 : montrer que $\left[\frac{2n}{p}\right] - 2\left[\frac{n}{p}\right] \geq 0$ et vaut 1 si $n < p \leq 2n$. En déduire, avec **6-3**, que

$$\sum_{n<p\leq 2n} \log p \leq 2n(\mathbb{E}_{2n}(\log^*) - \mathbb{E}_n(\log^*)) = \mathcal{O}(n).$$

En déduire, en découpant l'intervalle $[1, n]$ à l'aide de puissances entières de 2, que

$$\sum_{p\leq n} \log p = \mathcal{O}(n). \tag{$F2$}$$

6-5 : utiliser $(F1-2)$ pour montrer que

$$\sum_{p\leq x} \frac{\log p}{p} = \log x + \mathcal{O}(1). \tag{$F3$}$$

6-6 : on note $A(x) \approx B(x)$ pour $A, B > 0$ en x au voisinnage de $+\infty$, si il existe $0 < \alpha < \beta$ tels que $\alpha \leq \frac{A(x)}{B(x)} \leq \beta$ pour x assez grand.

Soit alors $K > 0$ tel que $|\mathcal{O}(1)| \leq K$, où $\mathcal{O}(1)$ apparaît dans $(F3)$. Montrer qu'alors, avec $(F3)$, pour un $\theta > 0$ bien choisi, et pour x assez grand,

$$\sum_{p\leq x} \log p \geq \theta x(\log \frac{1}{\theta} - 2K).$$

En déduire que

$$\sum_{p\leq x} \log p \approx x. \tag{$F4$}$$

6-7 : on note $\pi(x) = \mathrm{Card}\{p : p \leq x\}$. Remarquer que

$$\sum_{p\leq x} \frac{\log p}{\log x} \leq \pi(x) \leq \sum_{p\leq\sqrt{x}} 1 + \sum_{\sqrt{x}<p\leq x} \frac{\log p}{\log\sqrt{x}} \leq \sqrt{x} + \frac{2}{\log x}\sum_{p\leq x} \log p,$$

et en déduire que

$$\pi(x) \approx \frac{x}{\log x}. \tag{$F5$}$$

Remarque culturelle : le Théorème des Nombres Premiers affirme en fait que

$$\lim_{x\to+\infty} \frac{\pi(x)\log x}{x} = 1,$$

ce qui est plus fin que $(F5)$.

6-8 : on rappelle (intégrales de Bertrand pour ceux qui ne l'auraient jamais vu) que $\sum_{r\geq 1} \frac{1}{r\log r} = +\infty$. Noter p_r le $r^{\text{ième}}$ nombre premier, et calculer $\pi(p_r)$. Utiliser ensuite $(F5)$ pour montrer que

$$\sum_p \frac{1}{p} = +\infty.$$

UPJV, Maîtrise de Mathématiques **Année universitaire 2001/2002**
Session de Septembre **Module MM3 (Probabilités)**
Lundi 2 Septembre 2002 **14h–18h**

Tous documents autorisés, calculettes aussi.
E-mail, wap et communications téléphoniques interdits.

Le symbole "$\xrightarrow{\mathcal{L}}$" désigne la convergence en loi, et "$\mathbb{P}$" une probabilité.

$(\Omega, \mathcal{B}, \mathbb{P})$ *désignera l'espace probabilisé commun à toutes les variables aléatoires de ce sujet.*

Les exercices sont indépendants les uns des autres, sauf les deux premiers.

Le sujet étant court, les réponses devront être précises...!

Premier exercice : classements et records.

Soit $(X_i)_{i\geq 1}$ une suite de v.a.r. i.i.d. de loi commune ne présentant pas de charge ponctuelle (sans atome, i.e. continue). Soit B l'événement

$$B = \{\omega \in \Omega : \exists m \neq n : X_n(\omega) = X_m(\omega)\}.$$

I-1: montrer que $\mathbb{P}(B) = 0$.

On retire B de Ω, ce qui rend les égalités impossibles, tout en laissant inchangée la loi de la suite $(X_n)_{n\geq 1}$.

On note $T^{(n)}(\omega) = (T_1^{(n)}(\omega), \ldots, T_n^{(n)}(\omega))$ la permutation $(t_1, \ldots, t_n)$ de l'ensemble $\{1, \ldots n\}$ qui est telle que $X_{t_1}(\omega) < X_{t_2}(\omega) < \ldots < X_{t_n}(\omega)$. La variable $T^{(n)}$ enregistre donc le classement parmi $\{X_1, \ldots, X_n\}$.

On note Y_n le rang de X_n dans ce classement : autrement dit, $Y_n \in \{1, \ldots, n\}$ et $Y_n(\omega) = r$ si et seulement si $T_r^{(n)}(\omega) = n$.

I-2: montrer que $T^{(n)}$ est uniformément distribué sur l'ensemble des $n!$ permutations de $\{1, \ldots, n\}$.

I-3: montrer que $\mathbb{P}(Y_n = r) = \frac{1}{n}$, pour chaque $1 \leq r \leq n$.

I-4: montrer que Y_k est $\sigma(T^{(n)})$ mesurable pour chaque $k \leq n$.

I-5: montrer que la suite $(Y_j)_{j\geq 1}$ est indépendante.

Deuxième exercice : records.

Soit $(X_i)_{i\geq 1}$ une suite de v.a.r. i.i.d. de loi commune continue. On note A_n l'événement

$$A_n = \{\max_{1\leq k<n} X_k < X_n\}.$$

L'événement A_n modélise l'événement "*un record a eu lieu au temps n*".

II-1: montrer que la suite $(A_l)_{l\geq 1}$ est indépendante et que $\mathbb{P}(A_n) = \frac{1}{n}$.

II-2: montrer qu'un record ne dure jamais éternellement.

Troisième exercice : convergence faible.

III-1: on suppose que des v.a.r. X_n et X, $n \geq 1$, ont des lois de densités respectives f_n et f relativement à la mesure de Lebesgue sur la droite $\mathbb{R}$. Montrer que si $f_n(x) \to f(x)$ sauf pour x dans un ensemble Lebesgue négligeable, alors $X_n \xrightarrow{\mathcal{L}} X$.

III-2: on suppose qu'une suite de mesures de probabilités $(\mu_n)_{n\geq 1}$ sur la droite est telle qu'il existe $f : \mathbb{R} \to \mathbb{R}^+$ borélienne et telle que $\lim_{x\to\pm\infty} f(x) = +\infty$. Montrer que si $\sup_n \int f d\mu_n < +\infty$, alors la suite $(\mu_n)_{n\geq 1}$ est tendue.

Quatrième exercice : TCL pour une somme aléatoire.

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. centrées et de variance $\sigma^2 > 0$. On note $S_n = X_1 + \ldots + X_n$, $n \geq 1$. Pour chaque entier positif t, on suppose qu'il existe une variable aléatoire ν_t prenant des valeurs entières positives, non nécessairement indépendante des X_n. On suppose qu'il existe des constantes positives a_t et Θ telles que

$$\lim_{t\to+\infty} a_t = +\infty \qquad \text{et} \qquad \frac{\nu_t}{a_t} \underset{t\to+\infty}{\overset{\mathcal{L}}{\to}} \Theta.$$

Le but de ce problème est de montrer que

$$\frac{S_{\nu_t}}{\sigma\sqrt{\nu_t}} \underset{t\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}, \text{ et que } \frac{S_{\nu_t}}{\sigma\sqrt{\Theta a_t}} \underset{t\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}, \tag{$\star$}$$

où $\mathcal{N}$ désigne la loi normale centrée réduite.

IV-1 : montrer que l'on peut supposer sans perte de généralité que $\Theta = 1$ et $a_t \in \mathbb{N}$.

IV-2 : ceci étant fait, montrer que pour démontrer ($\star$), il suffit d'en montrer la seconde des deux convergences en loi.

IV-3 : montrer qu'il suffit de montrer que

$$\frac{S_{a_t} - S_{\nu_t}}{\sqrt{a_t}} \xrightarrow{\mathcal{L}} 0$$

("0" désigne la loi admettant une charge 1 en 0).

IV-4 : montrer que

$$\mathbb{P}(|S_{\nu_t} - S_{a_t}| \geq \varepsilon\sqrt{a_t}) \leq \mathbb{P}(|\nu_t - a_t| \geq \varepsilon^3 a_t) + \mathbb{P}(\max_{|k-a_t|\leq\varepsilon^3 a_t} |S_k - S_{a_t}| \geq \varepsilon\sqrt{a_t}).$$

Utiliser l'inégalité de Kolmogorov pour conclure.

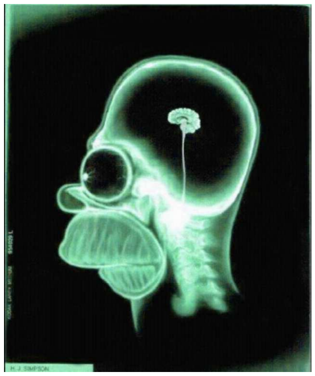
Cinqième exercice : martingales.

V-1 : on suppose que $(X_n)_{n\geq 1}$ est une martingale, et on suppose que $|X_n|$ et $|X_n - X_{n-1}|$ sont uniformément bornées par une constante indépendante de n. Montrer que si τ est un temps d'arrêt, alors $\mathbb{E}(X_\tau) = \mathbb{E}(X_1)$.

V-2 : soit $(Y_i)_{i\geq 1}$ une suite i.i.d. de v.a.r. centrées et de variance σ^2. On pose

$$X_n = (\sum_{k\leq n} Y_n)^2 - n\sigma^2.$$

Montrer que $(X_n)_{n\geq 1}$ est une martingale.



UPJV, Maîtrise de Mathématiques
Année universitaire 2002/2003
Partiel de MM3 (Probabilités)
Jeudi 21 Novembre 2002, 13h30–16h30

Tous documents autorisés, calculettes aussi.
E-mail, wap et communications téléphoniques interdits.

$(\Omega, \mathcal{B}, \mathbb{P})$ *désignera l'espace probabilisé commun à toutes les variables aléatoires de ce sujet.*

Les exercices sont indépendants les uns des autres.

On pourra admettre des résultats de questions antérieures, et poursuivre...

Il vaut mieux faire quelques questions proprement qu'en faire un tas sans rien jamais démontrer!

Premier exercice : une amélioration de l'inégalité de Biénaymé - Tchebyshev.

On se donne une suite $(X_n)_{n\geq 1}$ de v.a.r. i.i.d. de loi commune d'espérance m et de variance $\sigma^2 > 0$. On note $S_n = X_1 + \ldots + X_n$, et $\bar{X}_n = S_n \,/\, n$.

On suppose en outre qu'il existe une constante $A > 0$ telle que $|X_1| \leq A$ $\mathbb{P}$-p.s.. On suppose que $B \geq \sigma^2$. On se propose d'établir qu'alors si $0 \leq \beta \leq \frac{B}{A^2}$, et $n \geq 1$,

$$\mathbb{P}(|\bar{X}_n - m| \geq \beta A) \leq 2e^{-\frac{\beta^2 A^2 n}{4B}}$$

I-1: décrire en quoi une telle inégalité constitue une amélioration de l'inégalité classique de Biénaymé - Tchebyshev ?
I-2: montrer que si X est une v.a.r. positive, alors X est limite simple croissante d'une suite de v.a.r. positives simples et $\sigma(X)$-mesurables.
I-3: montrer que si $X \perp\!\!\!\perp Y$ sont deux v.a.r. indépendantes positives, alors intégrables ou pas

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \text{ avec la convention } 0 \times (+\infty) = 0.$$

I-4: montrer que si $\alpha, \beta \geq 0$ et $n \geq 1$, alors $e^{\alpha\beta A}\mathbb{P}(\bar{X}_n - m \geq \beta A) \leq \mathbb{E}(e^{\frac{\alpha}{n}(X_1 - m)})^n$.
I-5: montrer que si $|t| \leq 1$, alors $e^t \leq 1 + t + t^2$, et en déduire que si $\alpha \leq \frac{n}{2A}$, alors

$$\mathbb{E}(e^{\frac{\alpha}{n}(X_1 - m)}) \leq 1 + \frac{B\alpha^2}{n^2}.$$

I-6: utiliser $(1+\frac{t}{n})^n \leq e^t$ si $t \geq 0$ pour en déduire que si $\alpha \leq \frac{n}{2A}$, alors

$$\mathbb{P}(\bar{X}_n - m \geq \beta A) \leq e^{\frac{\alpha^2 B}{n} - \alpha\beta A}.$$

I-7: conclure en minimisant la fonction en α ci-dessus, et en changeant les signes.

Deuxième exercice.

Supposons que X est une var de fonction de répartition F continue strictement croissante. Montrer qu'alors
II-1: $F(X)$ suit une loi uniforme sur $[0,1]$.
II-2: $-\ln F(X)$ suit une loi exponentielle positive. De quel paramètre ?

Troisième exercice.

Soit (X,Y) un couple de v.a.r. dont la loi est donnée par

$$\mathbb{P}((X,Y) \in A) = C\left(\sum_{n\geq 0} \frac{1}{2^n} \int\int_A \mathbf{1}_{[n,n+1]}(x) d(m \otimes \delta_n)(x,y)\right),$$

où $C > 0$ est une constante, $A \in \mathcal{B}(\mathbb{R})$, δ_n désigne la masse de Dirac au point n, et m désigne la mesure de Lebesgue sur la droite $\mathbb{R}$.
III-1: calculer C, puis déterminer $\mathbb{P}_X$ et $\mathbb{P}_Y$.
III-2: les variables X et Y sont-elles indépendantes? Calculer $\mathbb{E}(XY)$ s'il y a lieu (justifier).

Quatrième exercice.

Soient f et g deux fonctions mesurables positives définies sur Ω, $(\Omega, \mathcal{B}, \mathbb{P})$ étant un espace probabilisé. Montrer que si $fg \geq 1$ $\mathbb{P}$-p.s., alors

$$\mathbb{E}(f)\mathbb{E}(g) \geq 1.$$

(*Indication* : on pourra s'intéresser à une inégalité de Hölder d'exposants $(\frac{1}{2}, \frac{1}{2})$).

Cinquième exercice.

Soit $h : \Omega \to]0, +\infty[$ mesurable, positive, $(\Omega, \mathcal{B}, \mathbb{P})$ étant un espace probabilisé. Noter $A = \mathbb{E}(h)$. Montrer que

$$\sqrt{1+A^2} \leq \mathbb{E}(\sqrt{1+h^2}) \leq 1 + A.$$

Si $\mathbb{P}$ désigne la mesure de Lebesgue sur $[0,1]$, et si h est continue, avec $h = f'$, donner une interprétation géométrique de cette double inégalité.

Sixième exercice.

Donner un exemple d'un triplet de trois v.a.r. (X,Y,Z) qui soit deux à deux indépedant sans être indépendant (on pourra penser à des évènements liés à trois lancés de dés indépendants, mais cela n'a rien d'obligatoire).