Cours de mathématiques

TOME IX

PROBABILITÉS

Mathématiques générales

France ~ 2025

Écrit et réalisé par Louis Lascaud

Table des matières

Chapitre 1

Théorie générale des probabilités

Résumé

Elle se fonde sur la théorie de la mesure de Borel-Lebesgue.

1.1 Axiomes des probabilités

Fait. ***(Généralité de la théorie des probabilités au sein de la mesure)***

Tout espace mesuré positif borné est un espace de probabilité à renormalisation près.

1.1.1 Espaces de probabilité classiques

Définition. ***(Espace probabilisé canonique)***

L'espace [0,1] muni de la tribu borélienne et de la mesure de Lebesgue restreintes à cet intervalle, est appelé l'*espace probabilisé canonique*.

Remarque. Le nom de cet espace est dû au fait que, comme on le verra, il est possible de construire, pour toute loi de probabilité donnée, une variable aléatoire sur cet espace, ayant pour loi, la loi en question.

1.2 Variables aléatoires

1.2.1 Définition

Définition. ***(Variable aléatoire)***

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $(F, \mathcal{F})$ un espace mesurable. On appelle *variable aléatoire* sur Ω (à valeurs dans F), toute application $X : \Omega \longrightarrow F$ qui soit $(\mathcal{A}, \mathcal{F})$ mesurable.

Remarques.

1. Notons que même le nom de l'objet variable aléatoire ne prévoit pas de mentionner la tribu d'arrivée, ce qui témoigne de son manque d'importance théorique. On reviendra sur ce point par la suite.
2. On devrait en toute rigueur dire que X est une $(\mathcal{A}, \mathcal{F})$-*variable aléatoire*, mais le contexte permet quasi toujours d'omettre cette précision. On commet cet abus dès maintenant et dans toute la suite.

Définitions. ***(Variable aléatoire finie)***

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, $(F, \mathcal{F})$ un espace mesurable et $X : \Omega \longrightarrow F$ une variable aléatoire. On dit que X est une variable aléatoire :

- *finie*, si $X(\Omega) \subseteq F$ est un ensemble de cardinal fini ;
- *discrète*, si $X(\Omega) \subseteq F$ est un ensemble de cardinal dénombrable ;
- *réelle*, si $X(\Omega) \subseteq F$ est inclus dans $\mathbb{R}$;
- *complexe*, si $X(\Omega) \subseteq F$ est inclus dans $\mathbb{C}$;
- *à valeurs vectorielles*, si $X(\Omega) \subseteq F$ est inclus dans un espace vectoriel normé E de dimension $2 \geqslant n < +\infty$.

$\longrightarrow$ *Convention.* Dans tous ces cas, on munit $X(\Omega)$ de la tribu grossière (cas discret) ou de la tribu trace de la tribu borélienne (cas continus).

Pour conclure cette section, on mentionne un résultat en pratique inutile mais d'ordre épistémologique important.

Théorème. ***(Lien entre tribu et fonction mesurable)***

La donnée d'une tribu est équivalente à celle d'une fonction mesurable.

$\longrightarrow$ *Notation.* On se permettra donc d'écrire qu'une fonction mesurable appartient à une tribu, pour dire qu'elle est mesurable par rapport à cette tribu (que la tribu qu'elle engendre y est contenue) :

$$Y \in \mathcal{T} \iff \sigma(Y) \subseteq \mathcal{T} \iff Y \text{ est } \mathcal{T}\text{-mesurable.}$$

Définition. ***(Mesurabilité par rapport à une variable)***

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, $(F, \mathcal{F})$ un espace mesurable et $X : \Omega \longrightarrow F$ une variable aléatoire. Soit $Y : \Omega \longrightarrow F$ une application. On dit que Y est X-*mesurable* si elle est $(\sigma(X), \mathcal{F})$-mesurable.

1.2.2 Loi d'une variable aléatoire

Définition. ***(Loi d'une variable aléatoire)***

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, $(F, \mathcal{F})$ un espace mesurable et $X : \Omega \longrightarrow F$ une variable aléatoire. On définit la *loi de* X comme l'application $\mathcal{L}_X : \mathcal{F} \longrightarrow [0,1]$ définie par :

$$\forall A \in \mathcal{F} \quad \mathcal{L}_X(A) = \mathbb{P}(X^{-1}(A)).$$

Théorème. ***(Loi de probabilité)***

La loi d'une variable aléatoire est une mesure de probabilité sur l'espace d'arrivée de cette variable.

▷ En reprenant les notations de la définition, la loi de X n'est autre que la mesure image de $\mathbb{P}$ par l'application mesurable X. C'est donc une mesure. De plus, $\mathcal{L}_X(F) = \mathbb{P}(X^{-1}(F))$, ce qui est licite car $F \in \mathcal{F}$. Or pour toute application définie partout $X^{-1}(F) = \Omega$, et par axiome $\mathbb{P}(\Omega) = 1$ ce qui permet de conclure que $\mathcal{L}_X$ est une mesure de probabilité. ■

Heuristique

On transporte sur $\mathcal{F}$, à l'aide de X, la mesure de probabilité $\mathbb{P}$ dont on dispose sur $\mathcal{A}$.

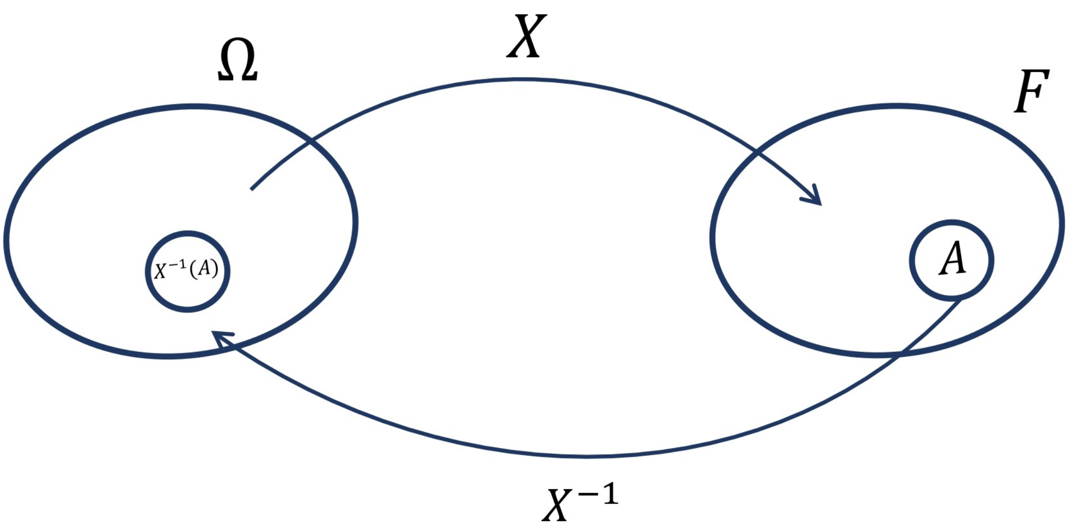


FIGURE 1.2.1 : *Transfert de la mesure de probabilité sur un ensemble quelconque.* —

On a les cas particuliers fondamentaux suivant pour les variables aléatoires discrètes et les variables aléatoires continues.

Théorème. ***(Loi discrète)***

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X un variable aléatoire **discrète** à valeurs dans $(F, \mathcal{F})$. Alors :

$$\mathcal{L}_X = \sum_{a \in \mathrm{Im}(X)} Pr(X = a)\delta_a.$$

Forme générale d'une loi de probabilité

Attention ! Il existe, même en pratique, des lois de probabilité qui ne sont ni discrètes, ni à densité (par rapport à la mesure de Lebesgue), mais des sommes des deux. Il existe également des mesures à densité qui ne le sont pas par rapport à la mesure de Lebesgue (par exemple, celle d'une variable aléatoire ayant pour fonction de répartition l'escalier de Cantor) ; on parle de *mesure singulière.*

Dans le cas général, on peut écrire :

$$\mathcal{L}_X = \sum_{a \in E \leqslant \mathbb{N}} \mathbb{P}(X = a)\delta_a + f\mathrm{d}x + g\mathrm{d}\mu$$

où $\mathrm{d}\mu$ est une mesure singulière. La démonstration de ce résultat, qui utilise le théorème de décomposition de Lebesgue, dépasse le cadre de ce chapitre.

1.2.3 Opérations sur les variables aléatoires

Attention ! Pour définir des opérations sur plusieurs variables aléatoires (somme, produit, etc.), il faut qu'elles soient définies **sur le même espace probabilisé !**

1.2.3.1 Somme et combinaison linéaire

1.2.3.2 Produit

1.2.3.3 Minimum, maximum de lois

1.2.3.4 Séries de variables aléatoires

Pour vérifier la bonne définition, le plus simple est souvent de vérifier la convergence normale.

1.3 Fonction de répartition

1.3.1 Fonction de répartition d'une mesure de probabilité

1.3.2 Fonction de répartition d'une variable aléatoire réelle

C'est la même chose en variant légèrement le formalisme. Nous reprenons tout à zéro néanmoins.

Définition. ***(Fonction de répartition d'une variable aléatoire)***

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire réelle. On définit sa *fonction de répartition* $F_X : \mathbb{R} \longrightarrow [0,1]$, par :

$$F_X(t) = \mathbb{P}(X \leqslant t).$$

Remarque. Par définition, on a également pour tout réel t : $\boxed{F_X(t) = \mathcal{L}_X(]-\infty,t])}$.

Propriété. ***(Inverse de la fonction de répartition)***

Pour tout variable aléatoire réelle X, pour tout réel t,

$$\mathbb{P}(X \geqslant t) = 1 - F_X(t).$$

VOC On immortalise ce constat en appelant la fonction, souvent rencontrée, $t \mapsto \mathbb{P}(X \geqslant t)$, 1 *moins la fonction de répartition.*

Théorème. ***(Théorème de la réciproque)***

Soit $F : \mathbb{R} \longrightarrow [0,1]$ une fonction croissante, continue à droite et de limites en $-\infty, +\infty$ respectivement 0 et 1. Alors F est la fonction de répartition d'une variable aléatoire. Plus précisément, F est la fonction de répartition d'une variable aléatoire sur l'espace probabilisé canonique.

▷ Soit une variable aléatoire U de loi uniforme sur $[0,1]$. Il est possible de construire une telle variable aléatoire : on prend pour espace $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), dx)$ et l'on prend $U = id_\Omega$, soit $U : \Omega \longrightarrow \Omega,\ \omega \longmapsto \omega$. C'est bien une application mesurable donc une variable aléatoire. Elle est bien de loi uniforme, car $\mathbb{P}(U \leqslant t) = \lambda([0,t]) = t$ pour $t \in [0,1]$.

Soit donc F une fonction vérifiant les hypothèses de l'énoncé. L'idée de la preuve est de poser $X = F^{-1}(U)$, ainsi, pour tout $t \in \mathbb{R}$, on a :

$$\mathbb{P}(X \leqslant t) = \mathbb{P}(F^{-1}(U) \leqslant t) = \mathbb{P}(U \leqslant F(t)) = F(t).$$

Le problème est que F^{-1} est mal définie dans le cas général. Si F était strictement croissante (sans défaut donné par les plateaux) et continue (sans défaut donné par les sauts), X serait bien défini et une variable aléatoire ; seulement a priori F^{-1} n'existe pas.

On va introduire la notion d'*inverse généralisée* que l'on notera toujours de la même manière et on va montrer que le raisonnement précédent tient bon. On pose :

$$\begin{aligned} F^{-1}: \quad [0,1] &\longrightarrow \overline{\mathbb{R}} \\ p &\longmapsto \inf\{x \in \mathbb{R},\ F(x) \geqslant p.\} \end{aligned}$$

Notons que $\boxed{\text{cette opération change les sauts de } F \text{ en plateaux de } F^{-1}}$. On montre maintenant que pour tout $p \in [0,1]$, pour tout $x \in \mathbb{R}$, $p \leqslant F(x) \iff F^{-1}(p) \geqslant x$. En effet, si $p \leqslant F(x)$, alors par définition de F^{-1}, on a $F^{-1}(p) \leqslant x$. Réciproquement, si $F^{-1}(p) \leqslant x$ alors la croissance de F et la définition de F^{-1} assurent que pour tout $\varepsilon > 0$, on a $F(x+\varepsilon) \geqslant p$ et ainsi par continuité à droite $F(x) \geqslant p$.

On calcule alors de même que dans notre laïus : $\mathbb{P}(X \leqslant t) = F(t)$. Donc X est une variable aléatoire de fonction de répartition F. ■

À la lumière du développement précédent, on peut introduire la notion suivante, qui est utile dans nombre de raisonnement sur les fonctions de répartition de variable aléatoire :

Définition. ***(Inverse généralisée, pseudo-inverse de Lévy))***

] Soit F une fonction de répartition. Alors l'*inverse généralisée* ou *pseudo-inverse* de F est donnée par :

$$\begin{aligned} F^{-1}: \quad [0,1] &\longrightarrow \overline{\mathbb{R}} \\ p &\longmapsto \inf\{x \in \mathbb{R},\ F(x) \geqslant p\} \end{aligned}.$$

Alors on a la propriété suivante qui tient : $p \leqslant F(x) \iff F^{-1}(p) \geqslant x$. L'inverse généralisée permet donc de généraliser la notion d'inverse pour des inégalités croissantes.

1.4 Moments d'une variable aléatoire

1.4.1 Espérance

1.4.1.1 Définition

En théorie de la mesure, on a défini l'intégrale d'une fonction mesurable par rapport à une mesure quelconque. En probabilité, on parle d'espérance (intégration d'une variable aléatoire par rapport à la mesure de probabilité sur l'espace de départ).

Définition. ***(Espérance d'une variable aléatoire réelle)***

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire réelle. On définit l'espérance de X par :

$$\mathbb{E}(X) = \int_\Omega X(\omega)\mathbb{P}(\mathrm{d}\omega).$$

$\longrightarrow$ *Notation.* La linéarité de l'opérateur espérance justifie la notation : $\mathbb{E}(X) = \mathbb{E}X$.

Le calcul de l'espérance d'une variable aléatoire fonction d'une autre, extrêmement courant, peut s'effectuer d'une manière très simple, grâce au théorème fondamental suivant :

Théorème. ***(Théorème de transfert)***

Soit X une variable aléatoire à valeurs dans l'espace mesurable $(F, \mathcal{F})$ et $h : F \longrightarrow \mathbb{R}$ une fonction mesurable, alors le transfert s'aplique.

1.4.1.2 Cas particuliers à masse

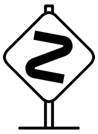
Proposition. ***(Espérance dans le cas discret)***

Soit $X \in L^1$ une v.a.r. discrète. Alors :

$$\mathbb{E}X = \sum_{a \in \mathrm{Im}(X)} a\mathbb{P}(X = a).$$

$\triangleright$ On écrit $\mathbb{E}X = \int_\Omega X(\omega)\mathrm{d}\mathbb{P}(\omega) = \int_\Omega \sum_a a\mathbb{1}_{\{X=a\}}(\omega)\mathrm{d}\mathbb{P}$ si $X(\omega) = \sum_a a\mathbb{1}_{\{X=a\}}(\omega)$. Par le théorème de Fubini L^1, on a : $\mathbb{E}X = \sum_{a \in \mathrm{Im}(X)} \int_\Omega a\mathbb{1}_{\{X=a\}}\mathrm{d}\mathbb{P}(\omega)$. $\blacksquare$

1.4.1.3 Formule de l'espérance par les queues



Toutes ses formules ne valent que pour des variables aléatoires positives ou presque sûrement positives.

Proposition. ***(Formule de l'espérance cumulée)***

Soit X une variable aléatoire à valeurs dans $\mathbb{Z}$. Alors $\mathbb{E}(X) = \sum_{k=0}^{+\infty} \mathbb{P}(X \geqslant k)$.

Proposition. ***(Formule de l'espérance cumulée)***

Soit X une variable aléatoire presque sûrement positive ou nulle. Alors :

$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X \geqslant x)\mathrm{d}x = \int_0^{+\infty} \mathbb{P}(X > x)\mathrm{d}x.$$

Proposition. ***(Transfert de l'espérance cumulée)***

Soit X une variable aléatoire réelle presque sûrement positive et φ une fonction positive, continûment dérivable, croissante sur $\mathbb{R}_+$ et telle que $\varphi(0) = 0$. Alors

$$\mathbb{E}(X) = \int_0^{+\infty} \varphi'(x)\mathbb{P}(X \geqslant x)\mathrm{d}x.$$

On peut là encore remplacer l'inégalité large par une inégalité stricte.

Proposition. ***(Formule des moments cumulés)***

Soit X une variable aléatoire presque sûrement positive ou nulle et $\alpha > 0$.

$$\mathbb{E}(X) = \int_0^{+\infty} \alpha x^{\alpha-1}\mathbb{P}(X \geqslant x)\mathrm{d}x.$$

Proposition. ***(Formule des moments cumulés, cas discret)***

Soit X une variable aléatoire à valeurs dans $\mathbb{Z}$ et $\alpha > 0$.

$$\mathbb{E}(X) = \sum_{k=0}^{+\infty} (k^\alpha - (k-1)^\alpha)\mathbb{P}(X \geqslant k).$$

1.4.1.4 Inégalités classiques sur l'espérance

Proposition. ***(Inégalité de la valeur absolue)***

Soit X un variable aléatoire réelle L^1. Alors :

$$|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|).$$

Proposition. ***(Inégalité de Jensen)***

Soit X un variable aléatoire réelle L^1. Soit φ une fonction convexe positive. Alors :

$$\varphi(\mathbb{E}X) \leqslant \mathbb{E}\varphi(X).$$

On introduit à présent, l'inégalité de Markov, qui malgré une grande simplicité de preuve, est hégémonique dans la théorie des probabilités.

Proposition. ***(Inégalité de Markov)***

Soit X un variable aléatoire réelle positive. Alors pour tout $t > 0$,

$$\mathbb{P}(X \geqslant t) \leqslant \frac{\mathbb{E}(X)}{t}.$$

Heuristique

L'inégalité de Markov est une première approche de la *majoration des queues de probabilités* grâce aux invariants « moments ».

1.4.2 Variance

Proposition. ***(Existence de la variance)***

Soit X une variable aléatoire réelle. Alors X admet une variance (finie) si et seulement si $\mathbb{E}[X^2]$ existe dans $\mathbb{R}$, c'est-à-dire $X \in L^2$.

▷ Si $X \in L^2$, $X \in L^1$ donc par König-Huygens, $\mathrm{Var}(X)$ existe dans $\mathbb{R}$. Réciproquement, par définition, si $\mathrm{Var}(X)$ admet une variance, alors $X \in L^1$ donc par König-Huygens, en faisant la différence de deux termes finis, $\mathbb{E}[X^2] = \mathrm{Var}(X) + \mathbb{E}[X]^2 < \infty$. ■

Proposition. ***(Caractère L^1 des variables à variance)***

Soit X une variable aléatoire réelle. Si X admet une variance, elle admet une espérance.

▷ C'est par définition ! ■

1.4.3 Moments

1.4.3.1 Définition générale des moments

Définition. ***(Moments d'une variable aléatoire)***

Soit X une variable aléatoire réelle sur un espace de probabilité passé sous silence. Soit $n \in \mathbb{N}$. Le *n-ième moment* de la variable X, ou *moment d'ordre n*, est défini par :

$$m_n = \mathbb{E}(X^n).$$

Proposition. ***(Calcul des moments par transfert)***

Soit X une variable aléatoire L^n pour n un entier naturel. Alors

$$m_n(X) = \int_{\mathbb{R}} x^n \mathrm{d}\mathbb{P}_X(x).$$

1.4.3.2 Existence des moments

Proposition. ***(Existence des moments a fortiori)***

Soit X une variable aléatoire réelle admettant un moment d'ordre $n \in \mathbb{N}$ fini. Alors pour tout $k \leqslant n$, X admet un moment fini d'ordre k.

▷ Vient de l'inégalité de Hölder avec un 1, qui montre également qu'en mesure fini, L^{p_2} est inclus dans L^{p_1} pour $p_2 \geqslant p_1$. ■

1.4.3.3 Inégalités classiques sur les moments

Proposition. ***(Inégalité des normes)***

Soient $p,q \in [1,+\infty]$. Soit X un variable aléatoire réelle quelconque. Alors :

$$\mathbb{E}(|X|^p)^{1/p} \leqslant \mathbb{E}(|X|^q)^{1/q}.$$

Proposition. ***(Inégalité de Cauchy-Schwartz)***

Soient X,Y deux variables aléatoires réelles quelconques. Alors :

$$\mathbb{E}(|XY|) \leqslant \mathbb{E}(|X|^2)^{1/2}\mathbb{E}(|Y|^2)^{1/2}.$$

Proposition. ***(Inégalité de Hölder)***

Soient $p,q \in [1,+\infty]$ deux exposants conjugués. Soient X,Y deux variables aléatoires réelles quelconques. Alors :

$$|\mathbb{E}(XY)| \leqslant \mathbb{E}(|XY|) \leqslant \mathbb{E}(|X|^p)^{1/p}\mathbb{E}(|Y|^q)^{1/q}.$$

Proposition. ***(Inégalité de Markov généralisée pour les moments)***

Soit X un variable aléatoire réelle positive. Soit $p \in [1,+\infty]$. Alors pour tout $t > 0$,

$$\mathbb{P}(X \geqslant t) \leqslant \frac{\mathbb{E}(X^p)}{t^p}.$$

1.4.4 Cumulants

1.5 Fonction caractéristique

Définition. ***(Fonction caractéristique)***

Soit X une variable aléatoire réelle. La *fonction caractéristique* de X est la fonction à valeurs complexes $\phi_X : \mathbb{R} \longrightarrow \mathbb{C}$ définie par $t \in \mathbb{R}$ et :

$$\phi_X(t) = \mathbb{E}(e^{itX}) = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)].$$

Il y a un léger abus dans le nom de cette fonction. En vérité, on devrait parler de *fonction caractéristique de la loi* de X, car deux variables aléatoires distinctes mais de même loi auront la même fonction caractéristique.

1.5.1 Intérêt de la fonction caractéristique

Lemme

Soit X une variable aléatoire de loi $\mathcal{N}(m, \sigma^2)$ pour des paramètres $m \in \mathbb{R}$ et $\sigma \geqslant 0$. Alors $\phi_X(t) = e^{imt - \frac{\sigma^2 t^2}{2}}$ pour tout réel t.

▷ Il suffit de prouver le résultat dans le cas centré, les propriétés opératoires de ϕ_X permettant l'extrapolation. Prenons donc $X \sim \mathcal{N}(0, \sigma^2)$. Par transfert, $\phi_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x = \dfrac{1}{\sigma\sqrt{2\pi}} \left(\int_{\mathbb{R}} \cos(tx) e^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x + \int_{\mathbb{R}} \sin(tx) e^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x \right)$ en ouvrant l'exponentielle imaginaire. Cette deuxième intégrale est nulle, car la fonction $x \longrightarrow sin(tx)e^{-\frac{x^2}{2\sigma^2}}$ est impaire donc son intégrale vaut 0. Ainsi, $\phi_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} g(t,x) \mathrm{d}x$ où $g(t,x) = \cos(tx) e^{-\frac{x^2}{2\sigma^2}}$. Par dérivation sous le signe intégral, on a une expression de ϕ'_X en fonction de ϕ_X, de classe $\mathbb{C}^1$. Or les propriétés de base donnent une valeur en 0. D'après la théorie des équations différentielles ordinaires, on obtient une expression de $\phi_X(t)$. ■

Remarque. La fonction caractéristique envoie à constante près une gaussienne sur une gaussienne d'écart-type inverse.

Théorème

La fonction caractéristique caractérise la loi.

▷ Le sens direct est simple. ■

1.5.2 Formule d'inversion des fonctions caractéristiques

La formule d'inversion de Fourier s'applique dans le cadre des probabilités, avec la fonction caractéristique, est énonce un résultat fort : l'intégrabilité de la fonction qui caractérise la loi, apporte une certaine régularité de la probabilité : d'abord, elle admet une densité, ensuite, on peut l'exprimer.

Théorème. ***(Formule d'inversion de Fourier)***

Soit X une variable aléatoire telle que $\phi_X \in L^1$. Alors X admet une densité continue et bornée, donnée par la formule suivante :

$$\forall x \in \mathbb{R} \qquad f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_X(t) \mathrm{d}t.$$

▷ Soit $\sigma > 0$. Soit $u \in \mathbb{R}$. Alors en prenant $Y \sim \mathcal{N}(0, \frac{1}{\sigma^2})$, on sait que $\phi_Y(u) = \mathbb{E}e^{iuY} = e^{\frac{-u^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\frac{1}{\sigma^2}}} \int_{\mathbb{R}}^{e} itu - \frac{t^2}{2\frac{1}{\sigma^2}} \mathrm{d}^{\mathrm{d}}u.$

On prend maintenant $t = X - y$ dans ce qui précède, et l'on passe à l'espérance.

Reste à vérifier que f est bien continue et bornée. La densité f est bornée, car $|f(x)| \leqslant \frac{1}{2\pi} \int \left|e^{-itx}\right| |\phi_X(t)| \,\mathrm{d}t = \frac{1}{2\pi} \|\phi_X\|_1$. On montre que la transformée de Fourier inverse d'une fonction L^1, est continue. C'est une conséquence du théorème de continuité sous le signe intégral. ■

1.6 Quelques généralités

Méthode. ***(Étudier une loi de probabilité)***

1. Fournir une description « expérimentale »/interprétation combinatoire
2. Alternativement, donner l'expression mathématique de la loi
3. Triturer les paramètres à leurs cas limites
4. Trouver le support
5. Donner la fonction de masse ou densité
6. Si possible aussi,la fonction de répartition
7. Calculer espérance, médiane, variance, dispersion, premiers moments
8. Écrire une condition de centrage, de réduction, et de centrage et réduction simultanés
9. Calculer mode, asymétrie
10. Calculer kurtosis, entropie de Shannon
11. Calculer fonctions génératrices, caractéristique

1.7 Notion d'indépendance

Avant de parler d'indépendance, nous préférons tout de suite rappeler que l'indépendance de variables aléatoires **n'est définie que pour des variables définies toutes sur le même espace.** Cette restriction technique n'est pas vraiment gênante, car on parlera surtout de somme, ou de produit, de variables aléatoires indépendantes, opérations qui, comme on l'a vu, ne peuvent elles aussi être définies que lorsque les variables mises en jeu sont définies sur un même espace probabilisé.

1.7.1 Indépendance d'événements

Dans toute cette section, on se donne un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

1.7.1.1 Indépendance de deux événements

Définition. ***(Indépendance de deux événements)***

Soient $A,B \in \mathcal{A}$. Les événements A,B sont dits *indépendants*, si $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cap \mathbb{P}(B)$.

Propriété. ***(Caractérisation de l'indépendance quant à un événement de probabilité non nulle)***

Soient $A,B \in \mathcal{A}$ où $\mathbb{P}(B) \neq 0$. Alors A et B sont indépendants si et seulement si $\mathbb{P}_B(A) = \mathbb{P}(A)$.

Corollaire

Soient A,B deux événements. Alors A et B sont indépendants si et seulement s'ils sont tous deux presque impossibles, ou si $\mathbb{P}_B = \mathbb{P}$, ou si $\mathbb{P}_A = \mathbb{P}$.

On remarque que cette caractérisation, quoique très intuitive (bien contrairement à la définition) n'est pas la bonne, car n'est pas aussi générale que la définition, et ne se généralise pas à une collection quelconque.

1.7.1.2 Indépendance d'une collection quelconque d'événements

Définition. ***(Indépendance d'un nombre fini d'événements)***

Soient n un entier naturel et $A_1,\ldots,A_n \in \mathcal{A}$. Les événements $A_1,\ldots,A_n$ sont dits *mutuellement indépendants*, ou plus simplement *indépendants*, si $\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \ldots \times \mathbb{P}(A_{i_k})$ pour toute partie $\{i_1, \ldots, i_k\} \subseteq \{1,\ldots,n\}$, $k \in [\![1,n]\!]$.

Remarque importante. Cette définition est la bonne, car on veut la stabilité par restriction.

Définition. ***(Indépendance d'un nombre quelconque d'événements)***

Soit $(A_i)_{i \in I}$ une collection quelconque d'événements, I un ensemble non nécessairement fini. On dit que les événements A_i sont *indépendants*, si toute sous-famille finie forme une famille indépendante d'événements. Autrement dit, si pour tout $J \subseteq I$ fini, $\prod\limits_{j \in J} \mathbb{P}(A_j) = \mathbb{P}(\bigcap\limits_{j \in J} A_j)$.

Dans quelle mesure peut-on généraliser cette définition à une collection infinie dénombrable ? On a la propriété suivante :

Propriété. ***(Produit infini d'événements indépendants)***

Soient $(A_n)_{n\in\mathbb{N}}$ une suite d'événements mutuellement indépendants. Alors $\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \prod_{n=1}^{+\infty} \mathbb{P}(A_n)$.

▷ On passe à la limite grâce au théorème de la limite monotone pour une intersection en mesure probabiliste. ■

Exercice 1

Donner un exemple d'une famille dénombrable d'événements vérifiant la propriété ci-dessous, dont les événements ne sont ni deux à deux indépendants, ni mutuellement.

▷ **Éléments de réponse.**

C'est simple !

Pour un famille d'événements mutuellement indépendants, le principe d'inclusion-exclusion se récrit un peu différemment.

Propriété. ***(Formule du crible indépendant)***

Soient $n \in \mathbb{N}$ et $A_1,\ldots,A_n$ des événements mutuellement indépendants. Alors :

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n}(-1)^{k-1} \sum_{1\leqslant i_1<i_2<\ldots<i_k\leqslant n} \prod_{j=1}^{k} \mathbb{P}(A_{i_j}).$$

On voudrait mieux prendre un vecteur de longueur n, ce qui permettrait de vérifier l'indépendance systématiquement. La propriété suivante réconcilie avec cette idée en prenant une tribu plus grande. La preuve est un enfer.

Corollaire

Si $(A_i)_{i\in I}$ est une famille d'événements indépendants, alors

$$(\overline{A_i})_{i\in I}$$

également.

1.7.2 Indépendance de tribus

Le théorème de l'enfer nous donne envie de définir une notion d'indépendance de tribus. On se donne un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition. ***(Indépendance de tribus)***

Soient $\mathcal{A}_1, ..., \mathcal{A}_n$, $n \in \mathbb{N}$, des sous-tribus de $\mathcal{A}$. Elles sont *indépendantes* si :

$$\forall A_1 \in \mathcal{A}_1, ..., \forall A_n \in \mathcal{A}_n, \quad \mathbb{P}(A_1 \cap ... \cap A_n) = \mathbb{P}(A_1)...\mathbb{P}(A_n).$$

$\longrightarrow$ *Notation.* Dans un calcul, on notera parfois : $\mathbb{P}(\mathcal{A}_1 \cap ... \cap \mathcal{A}_n) = \mathbb{P}(\mathcal{A}_1)...\mathbb{P}(\mathcal{A}_n)$.

De même que dans le cas des événements, l'indépendance mutuelle implique l'indépendance deux à deux (ah bon ?), mais la réciproque est fausse.

Fait. ***(Indépendance des événements par les tribus)***

Des événements sont indépendants, si et seulement si, les tribus qu'ils engendrent un à un sont indépendantes dans leur ensemble.

Pour contrôler cette définition, on préfère vérifier l'indépendance des tribus sur des ensembles plus petits.

Propriété. ***(Vérification pratique de l'indépendance des tribus)***

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soient n un naturel et $\mathcal{A}_1, ..., \mathcal{A}_n \subseteq A$ des tribus. Pour tout $i \in \{1,...,n\}$, soit $\mathcal{C}_i$ une collection d'ensembles stable par intersection finie, contenant Ω et telle que $\sigma(\mathcal{C}_i) = \mathcal{A}_i$. Si pour tous $A_1 \in \mathcal{C}_1, ..., A_n \in \mathcal{C}_n$, on a $\mathbb{P}(A_1 \cap ... \cap A_n) = \mathbb{P}(A_1)...\mathbb{P}(A_n)$, alors les tribus $\mathcal{A}_1, ..., \mathcal{A}_n$ sont indépendantes.

L'hypothèse contenant C est cruciale. Sinon, l'indépendance mutuelle pourrait se vérifier sur les vecteurs complets !

1.7.3 Indépendance de variables aléatoires

Définition. ***(Indépendance de variables aléatoires)***

Soient $X_1,...,X_n$, $n \in \mathbb{N}$ variables aléatoires définies sur le même espace $(\Omega,\mathcal{A}, \mathbb{P})$ à valeurs dans $(F_1, \mathcal{F}_1), ...,(F_n, \mathcal{F}_n)$ sont indépendantes si les σ-algèbres $\sigma(X_1),..., \sigma(X_n)$ le sont.

1.7.3.1 Lois jointes, lois produits

Heuristique

La loi jointe n'est pas la loi produit (sauf en cas d'indépendance) !
Soient X,Y deux variables aléatoires sur le même espace. On peut considérer sans problème $\mathcal{L}_X, \mathcal{L}_Y, \mathcal{L}_{(X,Y)}$, qui sont des mesures de probabilité.
D'autre part, on construit l'objet loi produit analytiquement, par une extension de Hahn-Kolmogorov, à partir de $\mathcal{L}_X$ et $\mathcal{L}_Y$ qui donne $\mathcal{L}_X \otimes \mathcal{L}_Y$; puisque $\mathcal{L}_X$ ne contient aucune information sur Y et réciproquement, $\mathcal{L}_X \otimes \mathcal{L}_Y$ ne peut pas contenir les informations sur le couplage de X et Y ! Pour s'en convaincre, il suffit de prendre $X = Y$...
En revanche, (X,Y) est une fonction à part entière donc sa loi contient les informations qui les *joignent* (les mots sont bien choisis, quand même). On comprend donc qu'on ne peut pas déduire $\mathcal{L}_{X,Y}$ à partir des seules lois marginales.

1.7.4 Lemmes de Borel-Cantelli

Exercice 2

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.r. sur cet espace.

1. Peut-on lier $\limsup\{X_n = 1\}$ et $\{\limsup X_n = 1\}$, ou leurs probabilités ?
2. Peut-on lier $\limsup\{X_n \leqslant 1\}$ et $\{\limsup X_n \leqslant 1\}$, ou leurs probabilités ?

▷ **Éléments de réponse.**

La méthode suivante doit éclaircir ces questions. On peut déjà remarquer, par exemple : $\limsup\{X_n = 1\} \subseteq \{\limsup X_n \geqslant 1\}$.

1.8 Probabilités conditionnelles

1.8.1 Arbres de probabilité

1.8.2 Probabilité sachant un événement

Remarquer que la mesure de probabilité conditionnelle ne donne de la masse qu'aux événements qui intersectent B, l'événement par rapport auquel on conditionne.

1.8.3 Formule des probabilités totales

Exercice 3 *(Ultra-classique)*

Un filtre à particules reçoit un flux de particules, dont le nombre par unité de temps suit une loi de Poisson de paramètre λ. Il filtre ce flux de telle sorte que les particules non toxiques sont rejetées dans l'air. Ces particules non toxiques sont présentes en proportion p dans le gaz originel. Quelle est la loi du nombre de particules rejetées dans l'air par unité de temps, supposée indépendante du nombre de particules arrivant dans le filtre ? Montrer que réciproquement, si cette loi est une loi de Poisson, alors l'hypothèse d'indépendance est automatique (!).

$\triangleright$ **Éléments de réponse.**

On modélise très naturellement le problème de la manière suivante. Soient X et Y des variables aléatoires réelles avec X une loi suivant une loi de Poisson de paramètre λ et la loi de Y conditionnée par $\{X = n\}$ est binomiale de paramètre (n,p). On cherche la loi de Y. C'est une simple application de la formule des probabilités totales : $\mathbb{P}(Y = k) = \sum_{n=0}^{\infty} \mathbb{P}_{X=n}(Y = k)$. Le calcul donne une loi de Poisson encore de paramètre λ.

En fait, on peut prendre l'exercice à l'envers, et remonter les formules pour montrer que, si X,Y suivent des lois de Poisson de même paramètre et le conditionnement de X à Y est binomial, alors pour tous $k,n \in \mathbb{N}$, $\mathbb{P}(Y = k, X = n) = \mathbb{P}(Y = k)\mathbb{P}(X = n)$, et le tour est joué.

Cet exercice très classique, et pas très compliqué, se décline sous toutes sortes d'enrobage : plaque d'or de l'expérience de Rutherford, ramasseur de champignons, etc, etc.

1.8.4 Formule des probabilités composées

1.9 Espérance conditionnelle

Le but de cette section est d'introduire la notion d'espérance conditionnelle, qui est un incontournable en probabilités. Cette notion généralise le conditionnement par rapport à un événement de probabilité nulle, mais sa définition précise est assez délicate. Notre présentation est progressive : nous repartons du conditionnement par rapport à un événement, et introduisons la notion d'espérance conditionnelle dans ce contexte, puis nous montons en degré de généralité. Remarquons dès maintenant que, dans notre théorie, le conditionnement d'une variable aléatoire ne se manifeste selon les définitions qu'à partir de son espérance.

1.9.1 Interprétation de la notion d'espérance conditionnelle

1.9.1.1 Lancer de dé : tirs pairs et tirs impaires

1.9.1.2 Un peu de météo

1.9.2 Conditionnement d'une variable aléatoire par rapport à un événement

On se place une fois pour toutes sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Notre objectif, dans un premier temps, est de définir l'espérance d'une variable aléatoire « conditionnée » à la réalisation d'un événement de probabilité positive. Pour cela, commençons par rappeler la définition de probabilité conditionnelle.

Définition-propriété. ***(Probabilité conditionnelle)***

Soit $B \in \mathcal{A}$ un événement tel que $\mathbb{P}(B) > 0$. On définit la probabilité conditionnelle sachant B, notée $\mathbb{P}(\,\cdot\,|B)$, relative à $\mathbb{P}$ bien sûr, par :

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

Il est facile de vérifier que cette expression définit bien une mesure de probabilité sur $(\Omega, \mathcal{A})$. En particulier, le facteur $\mathbb{P}(B)$ au dénominateur assure que la masse totale de cette mesure vaut bien 1.

Sans grande recherche, on peut définir l'espérance conditionnelle d'une variable aléatoire sachant un événement de probabilité non nulle, en remplaçant la probabilité sur Ω par la probabilité conditionnelle.

Définition. ***(Espérance conditionnelle d'une variable sachant un événement)***

Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $B \in \mathcal{A}$ un événement de probabilité non nulle. Si $X \geqslant 0$ ou $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, on définit l'*espérance conditionnelle de X sachant B* par :

$$\mathbb{E}(X|B) = \int_\Omega X \mathrm{d}\mathbb{P}(\,\cdot\,|B).$$

Si le lecteur est encore éveillé ici, il aura remarqué qu'il n'est pas immédiat que $\mathbb{E}(X|B)$ soit bien définie lorsque X est seulement dans L^1. En effet, dans le cas positif, l'intégrale est toujours bien définie dans $\overline{\mathbb{R}}$, mais dans le cas où la variable $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, il n'est pas évident que $X \in L^1(\Omega, \mathcal{A}, \mathbb{P}(\,\cdot\,|B))$. La proposition suivante permet de régler ce problème.

Proposition. ***(Identité de l'espérance conditionnelle)***

Soit X une variable aléatoire réelle sur Ω et B un événement non négligeable. Si $X \geqslant 0$, alors :

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)}.$$

Ainsi (s'en convaincre), la définition précédente fait sens pour $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et l'égalité ci-dessus reste vraie dans ce cas.

▷ Cette identité est vraie si X est une indicatrice. Par linéarité de l'espérance, elle est vraie pour tout variable étagée positive. Par approximation et théorème de convergence monotone, cette identité s'étend à toutes les variables mesurables positives. Dans le cas où X est L^1, en appliquant aux parties positive et négative, par linéarité de l'espérance, on a le même résultat. ■

Cette formule permet de justifier a posteriori l'intégrabilité contre $\mathbb{P}(\,\cdot\,|B)$.

Remarquer qu'on aurait pu poser les choses dans l'autre sens. Il n'y aurait eu aucun problème pour définir l'espérance conditionnelle immédiatement, mais heuristiquement, le fait suivant aurait été moins clair.

Principe. ***(Espérance conditionnelle)***

L'espérance conditionnelle est une espérance contre une certaine mesure de probabilité. Par conséquent, elle hérite de toutes les propriétés de l'intégrale et de l'espérance probabiliste.

On a de plus les propriétés suivantes :

Proposition. ***(Formule des probabilités totales pour l'espérance)***

Soit X une variable aléatoire réelle L^1 et B un événement ni négligeable ni presque sûr. Alors :

$$\mathbb{E}(X) = \mathbb{P}(B)\mathbb{E}(X|B) + \mathbb{P}(\overline{B})\mathbb{E}(X|\overline{B}).$$

Exercice 4

Cette formule se généralise-t-elle, comme la formule usuelle, pour une partition dénombrable de Ω ?

▷ **Éléments de réponse.**

La réponse est oui.

Corollaire. ***(Lien entre l'espérance et l'espérance conditionnelle)***

Soit X une variable aléatoire réelle L^1 et A un événement presque sûr. Alors $\mathbb{E}(X) = \mathbb{E}(X|A)$.

Corollaire

Soit X une variable aléatoire réelle L^1. Alors on a $\mathbb{E}(X) = \mathbb{E}(X|\Omega)$.

1.9.3 Conditionnement d'une variable aléatoire par rapport à une variable aléatoire discrète

Motivation. On souhaite cette fois définir l'espérance d'une variable aléatoire X sachant une autre variable aléatoire Y. Quel sens donne-t-on à cela ? De façon vague, l'espérance de X sachant Y pourrait être reliée à l'espérance de X sachant l'événement $\{Y = y\}$, pour tout y. On voit déjà que cet événement doit être de probabilité strictement positive pour que cela soit définissable ; on se restreint donc aux variables aléatoires Y discrètes.

Définition. ***(Espérance conditionnelle sachant une variable discrète)***

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, $X : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire réelle L^1 et $Y : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$ une variable aléatoire. On suppose de plus que E est dénombrable, que $\mathcal{E} = \mathcal{P}(E)$ et l'on note $E' = \{y \in E, \mathbb{P}(Y = y) > 0\}$. L'espérance conditionnelle de X sachant Y est définie pour tout $\omega \in \Omega$ par :

$$\mathbb{E}(X|Y)(\omega) = \sum_{y \in E} \mathbb{1}_{Y=y} \mathbb{E}[X|Y = y].$$

Remarque. Avec cette définition, l'espérance conditionnelle d'une variable sachant un événement pré-élémentaire (image réciproque d'un singleton par une variable aléatoire), n'est autre que l'espérance conditionnelle sachant une variable déterministe, en particulier discrète.

1.9.4 Construction générale de l'espérance conditionnelle : conditionnement d'une variable aléatoire par rapport à une tribu

1.9.4.1 Pré-cas L^2

Notons que L^2 est inclus dans L^1.

1.9.4.2 Cas général L^1

Résumons :

- ⋆ On a défini l'espérance conditionnelle sachant une variable discrète, donc à support dénombrable dans E', par $\mathbb{E}(X|Y) = \sum_{y\in E'} \mathbb{1}_{\{Y=y\}}(\omega)\mathbb{E}(X|\{Y=y\})$;
- ⋆ mais également l'espérance conditionnelle de $X \in L^2$ sachant $\mathcal{B}$;
- ⋆ on veut maintenant construire l'espérance conditionnelle L^1 sachant $\mathcal{B}$.

Comme dans le cas L^2, l'espérance conditionnelle d'une variable aléatoire L^1 sachant la tribu engendrée par une variable aléatoire discrète, coïncide avec l'espérance conditionnelle sachant une variable aléatoire discrète.

Proposition. ***(Consistance)***

Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et Y un variable aléatoire discrète. Alors $\mathbb{E}[X|\sigma(Y)] = \mathbb{E}[X|Y]$ presque sûrement.

▷ La variable Y est à valeurs dans un ensemble dénombrable. La preuve est quasi identique à la preuve L^2. ■

On définit maintenant :

Définition. ***(Espérance conditionnelle sachant une variable quelconque L^1)***

Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et Y une variable aléatoire. On définit l'espérance conditionnelle de X sachant Y, que l'on note $\mathbb{E}(X|Y)$, par l'espérance conditionnelle de X sachant $\sigma(Y)$.

Remarquons que la bonne notion est le conditionnement par rapport tribu et non une variable. On a défini par exemple le conditionnement seulement par rapport à une variable pour les discrètes, et l'on peut vérifier que ça coïncide bien. Mais on est bien au delà du conditionnement par rapport aux événements de mesure non nulle.

Remarque. (Espérance conditionnelle d'une variable positive) On étend toujours la notion d'espérance conditionnelle par des techniques d'approximation. La preuve de la construction de l'espérance conditionnelle L^1 montre que si $X \geqslant 0$, alors on peut définir l'espérance conditionnelle même si $\mathbb{E}[X] = \infty$. Sa caractérisation est alors : pour toute variable aléatoire Z qui soit $\mathcal{B}$-mesurable, bornée, positive, on a : $\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}]Z]$.

Cette remarque est utile car, bien que L^2 et L^1 ne soient pas disjoints de $\geqslant 0$, ce dernier n'est pas inclus dedans.

1.9.5 Propriétés de l'espérance conditionnelle

On donne maintenant les propriétés générales de l'espérance conditionnelle, dans le cadre L^1 qui est le plus sympathico-utile qu'on lui connaît.

Dans cette section, on fixe $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. On fixe également $Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ pour les propriétés à deux variables.

Proposition. ***(Linéarité de l'espérance conditionnelle)***

Pour tous réels a,b,

$$\mathbb{E}[aX + bY|\mathcal{B}] = a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B}) \text{ presque sûrement.}$$

On a certainement pas la linéarité de l'autre côté !

$\triangleright$ Par linéarité, $a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B}) \in L^1(\Omega, \mathcal{B}, \mathbb{P})$. Il nous suffit de montrer que pour toute variable Z $\mathcal{B}$-mesurable bornée,

$$\mathbb{E}[(a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B}))Z] = \mathbb{E}[(aX + bY)Z].$$

Or par linéarité de l'espérance, $\mathbb{E}[(a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B}))Z] = a\mathbb{E}[\mathbb{E}[X|\mathcal{B}]Z] + b\mathbb{E}[\mathbb{E}[Y|\mathcal{B}]Z] = a\mathbb{E}[XZ] + b\mathbb{E}[YZ]$ par définition de l'espérance conditionnelle de $\mathbb{E}[X|\mathcal{B}]$ et $\mathbb{E}[Y|\mathcal{B}]$. La linéarité de l'espérance simple permet de conclure. ■

Proposition. ***(Positivité de l'espérance conditionnelle)***

Si X est presque sûrement positive alors $\mathbb{E}[X|\mathcal{B}] \geqslant 0$ presque sûrement.

$\triangleright$ On pose $Z = \mathbb{1}_{\{\mathbb{E}[X|\mathcal{B}]<0\}}$. Alors Z est $\mathcal{B}$-mesurable et bornée et l'on a :

$$\mathbb{E}[X\mathbb{1}_{\{\mathbb{E}[X|\mathcal{B}]<0\}}] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}]\mathbb{1}_{\{\mathbb{E}[X|\mathcal{B}]<0\}}$$

d'où le calcul. (On peut ainsi faire la définition sans problème.) ■

Proposition. ***(Espérance conditionnelle d'une constante)***

Si a est une constante alors $\mathbb{E}(a|\mathcal{B}) = a$ p. s.

Proposition. ***(Espérance conditionnelle modulo la tribu grossière)***

Si $\mathcal{F} = \{\emptyset, \Omega\}$, alors $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$ presque sûrement, qui est donc déterministe.

Proposition. ***(Espérance conditionnelle modulo la tribu entière)***

On a $\mathbb{E}[X|\mathcal{A}] = X$ presque sûrement.

On énonce maintenant des propriétés plus avancées.

Proposition. ***(Espérance conditionnelle d'un produit par une variable non conditionnée)***

Soit Z une variable $\mathcal{B}$-mesurable telle que $XZ \in L^1$. Alors $\mathbb{E}[XZ|\mathcal{B}] = Z\mathbb{E}[X|\mathcal{B}]$. En particulier si X est $\mathcal{B}$-mesurable, $\mathbb{E}[X|\mathcal{B}] = X$.

▷ On se ramène au cas où X et Z sont positives en décomposant. Posons $Z_n = Z.\mathbb{1}_{\{Z \leqslant n\}}$. Pour toute v. a. Y $\mathcal{B}$-mesurable bornée et positive :

$$\mathbb{E}[Y(Z_n\mathbb{E}[X|\mathcal{B}])]$$

se calcule. ■

Proposition. ***(Espérance conditionnelle par rapport à une sous-sous-tribu)***

Soit $\mathcal{C} \subseteq \mathcal{B}$, alors $\mathbb{E}[\mathbb{E}[X|\mathcal{B}]|\mathcal{C}] = \mathbb{E}[X|\mathcal{C}]$.

On peut se demander quelle est l'espérance de l'espérance conditionnelle ; on obtient un résultat fondamental et pas si surprenant.

Proposition. ***(Formule de l'espérance totale)***

On a :

$$\mathbb{E}[\mathbb{E}[X|B]] = \mathbb{E}X.$$

Proposition. ***(Espérance conditionnelle triviale)***

Si $\sigma(X)$ et $\mathcal{B}$ sont indépendantes, alors $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$ presque sûrement.

▷ C'est une conséquence du théorème suivant en prenant $g(x,y) = x$. ■

Plus généralement :

Proposition. ***(Théorème de transfert)***

Pour toute variable aléatoire Y $\mathcal{B}$-mesurable, et pour toute fonctionnelle mesurable $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ telle que $g(X,Y)$ soit intégrable, on a :

$$\mathbb{E}(g(X,Y)|\mathcal{B}) = u(Y)$$

avec $u(y) = \displaystyle\int_{\mathbb{R}} g(x,y)\mathcal{L}_X(\mathrm{d}x)$ pour tout $y \in \mathbb{R}$.

▷ Par définition $\mathbb{E}[g(X,Y)|\mathcal{B}]$ est l'unique variable aléatoire dans $L^1(\Omega,\mathcal{B},\mathbb{P})$ tel que $\mathbb{E}[Z\mathbb{E}[g(X,Y)|\mathcal{B}]] = \mathbb{E}[Zg(X,Y)]$ pour toute variable Z $\mathcal{B}$-mesurable bornée. Or $\mathbb{E}[Zg(X,Y)] = \int zg(x,y)\mathcal{L}_{(X,Y,Z)}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}z)$ d'après le théorème de transfert classique. La variable X est indépendante de (X,Y), donc $\mathcal{L}_{(X,Y,Z)} = \mathcal{L}_X \bigotimes \mathcal{L}_{(Y,Z)}$. Ainsi, on a le résultat. ■

Proposition. ***(Inégalité de Jensen)***

Si $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ est convexe, alors

$$\phi\mathbb{E}[X|\mathcal{B}] \leqslant \mathbb{E}[\phi(X)|\mathcal{B}] \text{ presque sûrement.}$$

On peut étendre aussi des théorèmes de convergence classique. Pour cela, on doit considérer une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$. Remarquons qu'on doit utiliser, pour les propriétés de convergence monotone et de Fatou, il nous faut, même si la suite des X_n est dans L^1, la notion d'espérance conditionnelle positive ne serait-ce que pour la limite.

Proposition. ***(Convergence monotone pour l'espérance conditionnelle)***

Si $(X_n)_{n\geqslant 0}$ est croissante positive et $X = \lim\limits_{n\to+\infty} X_n$, alors presque sûrement, $\lim\limits_{n\to+\infty} \mathbb{E}[X_n|\mathcal{B}] = \mathbb{E}[X|\mathcal{B}]$.

Proposition. ***(Lemme de Fatou pour l'espérance conditionnelle)***

Si $(X_n)_{n\geqslant 0}$ est positive, alors $\mathbb{E}[\liminf\limits_{n\to+\infty} X_n|\mathcal{B}] \leqslant \liminf\limits_{n\to+\infty} \mathbb{E}[X_n|\mathcal{B}]$ presque sûrement.

Proposition. ***(Convergence dominée pour l'espérance conditionnelle)***

Si $\mathbb{X}_n \underset{n\to+\infty}{\longrightarrow} X$ p. s., et s'il existe Y intégrable positive telle que presque sûrement $|X_n| \leqslant Y$ pour tout $n \geqslant 1$, alors $\lim\limits_{n\to+\infty} \mathbb{E}[|\mathbb{E}[X_n|\mathcal{B}] - \mathbb{E}[X|\mathcal{B}]|] = 0$ et $\lim\limits_{n\to+\infty} \mathbb{E}[X_n|\mathcal{B}] = \mathbb{E}[X|\mathcal{B}]$ presque sûrement.

1.9.6 Méthodes de calcul d'espérance conditionnelle

Proposition. ***(Cas discret)***

Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et Y une variable aléatoire à valeurs dans un ensemble dénombrable E. Alors presque sûrement $\mathbb{E}[X|\sigma(Y)] = \phi(Y)$ où $\phi(y) := \begin{cases} \mathbb{E}[X|\{Y = y\}] \text{ si } \mathbb{P}(Y = y) > 0 \\ 0 \text{ sinon.} \end{cases}$

▷ Déjà montré précédemment. ■

Proposition. ***(Cas à densité)***

Soit X et Y deux variables aléatoires réelles tel que le couple (X,Y) admette une densité $f : \mathbb{R}^2 \longrightarrow R$. Soit $h : \mathbb{R} \longrightarrow \mathbb{R}$. Soit $h : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction mesurable telle que $h(X)$ est intégrable. Alors presque sûrement $\mathbb{E}[h(X)|\sigma(Y)] = \phi(Y)$ avec $\phi(y) :=$

$$\begin{cases} \frac{1}{\int f(x,y)\mathrm{d}x} \int h(x)f(x,y)\mathrm{d}x \text{ si } \int f(x,y)\mathrm{d}x > 0 \\ 0 \text{ sinon.} \end{cases}$$

▷ On pose $q(y) = \int_{\mathbb{R}} f(x,y)\mathrm{d}x$ pour tout $y \in \mathbb{R}$. On peut vérifier sans problème de que q est mesurable. On commence par montrer que la variable aléatoire Y admet pour densité q. En effet, pour tout $A \in \mathcal{B}(\mathbb{R})$,

$$\mathcal{L}_Y(A) = \int_y \int_x f(x,y)\mathrm{d}x \mathbb{1}_{A(y)}\mathrm{d}y = \int_y q(y)\mathbb{1}_{A(y)}\mathrm{d}y.$$

et je prétends que l'on peut l'écrire avec des probabilités en Y. Passons à la preuve de la formule. Supposons que l'on puisse prouver : $(\star)$ $\mathbb{E}[h(X)g(Y)] = \mathbb{E}[\phi(Y)g(Y)]$ pour toute fonction $g : \mathbb{R} \longrightarrow \mathbb{R}$ mesurable bornée. Comme toute variable aléatoire $\sigma(Y)$-mesurable est de la forme $g(Y)$ pour une certaine fonction mesurable $g : \mathbb{R} \longrightarrow \mathbb{R}$. On en déduira que $\mathbb{E}[h(X)Z] = \mathbb{E}[\phi(Y)Z]$ pour toute Z $\sigma(Y)$-mesurable bornée. Et alors $\mathbb{E}[h(X)|\sigma(Y)] = \varphi(Y)$ presque sûrement. Montrons $(\star)$. ■

Cette fonction curieuse ϕ de Y, est en fait d'une expression tout à fait raisonnable en fait.

Heuristique

Cette formule peut s'expliquer par le calcul informel suivant :

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

$$\frac{\int h(x)f(x,y)\mathrm{d}x}{\int f(x,y)\mathrm{d}x} = \phi(y)$$

On peut rendre le calcul presque rigoureux en prenant la probabilité que Y soit dans un voisinage de ε. Il s'agirait de passer à la limite sur son diamètre. En toute généralité on ne peut pas montrer ce résultat mais dès que l'on a un peu de régularité sur la densité on peut.

1.10 Convergence de variables aléatoires

On se limite formellement aux variables aléatoires à valeurs dans $\mathbb{R}$, mais tout s'applique en le changeant pour $\mathbb{R}^d$, où $d \in \mathbb{N}$.

1.10.1 Différentes notions de convergence

Dans toute la suite, on se donne un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ quelconque et $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoire, c'est-à-dire $X_n : \Omega \longrightarrow \mathbb{R}$ qui soient $(\mathcal{A},\mathcal{B}(\mathbb{R}))$-mesurables.

1.10.1.1 Problème de la notion de convergence simple

La convergence simple est déjà définie pour (X_n), à valeurs dans $\mathbb{R}$ et prenant leurs abscisses dans un ensemble quelconque ; c'est connu. Elle s'exprime alors : $\forall \omega \in \Omega \quad X_n(\omega) \longrightarrow X(\omega)$ lorsque n tend vers $+\infty$. On voit que cette notion de convergence n'est pas bonne pour les probabilités, car elle est trop forte.

Prenons l'exemple suivant : on tire à pile ou face une infinité de fois. Ainsi $X_n = +1$ si l'on tire pile et 0 si l'on tire face au n-ième lancer. Les (X_n) sont une suite de variables aléatoires indépendantes de loi $Ber(\frac{1}{2})$. Intuitivement, la moyenne $S_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ devrait converger vers $\frac{1}{2}$ à l'infini. On modélise avec $\Omega = \{0{,}1\}^{\mathbb{N}}$, $X_n(\omega) = \omega_n$ et S_n définie comme précédemment.

A-t-on $S_n(\omega)$ qui tend vers la probabilité moitié pour tout ω, c'est-à-dire, a-t-on convergence simple de S_n vers $\frac{1}{2}$?

Non ! Car pour tout ω contenant un nombre fini de 1, $S_n(\omega) \longrightarrow 0$ quand $n \to \infty$. Il faut donc considérer la convergence simple hors d'un ensemble de mesure nulle.

1.10.1.2 Convergence presque sûre

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé que l'on gardera dans toute la suite du chapitre.

Définition. ***(Convergence presque sûre)***

Soient (X_n), $n \geqslant 1$ et X des v. a. r. définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que X_n converge presque sûrement vers X si

$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \underset{n\to+\infty}{\longrightarrow} X(\omega))\} = 1.$$

On notera $X_n \xrightarrow[n\to+\infty]{\text{p. s.}} X$ ou $\lim\limits_{n\to+\infty} X_n = X$ p. s.

Les modes de convergence classiques sont définis pour des va sur le même espace. On ne se soucie plus de ça pour la convergence en loi.

Définition. ***(Convergence L^p)***

Soient (X_n), $n \geqslant 1$ et X des v. a. r. définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. On suppose que toutes ces v. a. appartiennent à $L^p(\Omega, \mathcal{A}, \mathbb{P})$ pour un certain $p \in [1, +\infty[$. On dit que X_n convergence vers X dans L^p si

$$\lim_{n\to+\infty} E[|X_n - X|^p] = 0.$$

On notera $X_n \xrightarrow[n\to+\infty]{L^p} X$ ou $\lim\limits_{n\to+\infty} X_n = X$ dans L^p.

Remarques.

1. Si X_n tend vers X dans L^1, alors $\mathbb{E}X_n \longrightarrow \mathbb{E}X$ et $\mathbb{E}|X_n|| \longrightarrow \mathbb{E}|X|$.
2. Si $1 \leqslant p < q < \infty$ alors si $X_n \xrightarrow[n\to+\infty]{L^q} X$ alors $X_n \xrightarrow[n\to+\infty]{L^p} X$ grâce à la croissance des normes L^p, propre aux mesures de proba.

Exemple fondamental

Soit X_n une loi uniforme sur $(0, \frac{1}{n})$ pour tout n supérieur à 1. Alors pour tout $p > 1$, X_n tend vers 0 en moyenne L^p.
En effet, il suffit de donner son espérance.

Définition. ***(Convergence en probabilité)***

Soient (X_n), $n \geqslant 1$ et X des v. a. r. définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que X_n converge vers X en probabilité (ou en mesure) si pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0.$$

On notera $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$.

De manière équivalente, $\boxed{\forall \varepsilon > 0, \forall \delta > 0, \exists N \geqslant 1, \forall n > N, Pr(|X_n - X| > \varepsilon) < \delta}$.

Attention ! La négation de la convergence en probabilité n'est pas $\forall \varepsilon$, ni il en existe une infinité telle que... Mais $\exists \varepsilon$.

Propriété. ***(Séparation des convergences probabilistes)***

Les limites en L^p, en probabilité et presque sûres sont presque sûrement uniques.

On verra dans la section suivante que, s'il y a plusieurs modes de convergence en même temps, les limites coïncident.

1.10.2 Comparaison entre les notions de convergence

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

Théorème. ***(Convergence $L^p \implies$ convergence en probabilité)***

Soient $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.
Si $X_n \xrightarrow[n\to+\infty]{L^p} X$, alors $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$.

▷ Soit $\varepsilon > 0$. Alors d'après l'inégalité de Markov,

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p > \varepsilon^p) \leqslant \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p}.$$

Or par hypothèse, cette quantité à droite tend vers zéro lorsque n tend vers l'infini. ■

Contre-exemple. *($\mathbb{P} \not\Longrightarrow$ p. s.)*

La convergence en probabilité n'implique pas la convergence presque sûre.

Sur $((0,1], \mathcal{B}((0,1]))$ muni de la mesure de Lebesgue. Pour tout $n \geqslant 1$ et pour tout $j \in \{1,\ldots,n\}$, on pose $Y_{n,j} = \mathbb{1}_{]\frac{j-1}{n},\frac{j}{n}]}(\omega)$. On pose le stroboscope infernal : $X_1 = Y_{1,1}$; $X_2 = Y_{2,1}$, $X_3 = Y_{2,2}$; $X_4 = Y_{3,1}$. Pour tout $m > 1$,

$$\mathbb{P}(|X_m - 0| > \varepsilon) \leqslant \mathbb{P}(X_m = 1) = \frac{1}{n}$$

où n est tel qu'il existe j tel que $X_m = Y_{n,j}$. Ainsi X_m tend vers 0 en $\mathbb{P}$, car $n(m)$ tend vers $+\infty$ pour $m \to \infty$.

Par contre, $\limsup_{m\to+\infty} X_m(\omega) = 1$ pour tout $\omega \in]0,1]$ et $\liminf_{m\to+\infty} X_m(\omega) = 0$ pour tout $\omega \in]0,1]$. Donc X_m ne converge pas presque sûrement, en fait, elle ne converge nulle part. □

Théorème. ***(Convergence presque sûre $\Longrightarrow$ convergence en probabilité)***

Soient $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.
Si $X_n \xrightarrow[n\to+\infty]{\text{p. s.}} X$, alors $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$.

▷ Soit $\varepsilon > 0$. On introduit $A_n = \{|X_n - X| > \varepsilon\}$. Alors $\limsup_{n\to+\infty} A_n = \{|X_n - X| > \varepsilon$ pour une infinité de $n\} \subseteq \{X_n$ ne converge pas vers $X\}$. Or ce dernier événement est de probabilité nulle. Ainsi $\mathbb{P}(\limsup_{n\to+\infty} A_n) = 0$. Ainsi $\mathbb{P}(\bigcup_{n\geqslant m} A_n) \xrightarrow[m\to+\infty]{} 0$ et ainsi $\mathbb{P}(A_n)$ tend vers 0 lorsque n tend vers l'infini. ■

Exemple. *($\mathbb{P} \not\Longrightarrow L^p$)*

La convergence en proba n'implique pas la convergence L^p.

Même exemple sauf que $Y_{n,j} = n^{\frac{1}{p}}\mathbb{1}_{]\frac{j-1}{n},\frac{j}{n}]}$. Alors $\mathbb{P}(|X_m - 0| > \varepsilon) \leqslant \mathbb{P}(X_m = n^{\frac{1}{p}})$ avec toujours n,j tels que $X_m = Y_{n,j}$. Mais à droite, ce terme est $1/n$ qui tend vers 0. Donc X_m tend vers 0 en probabilité. Si X_m converge dans L^p nécessairement sa limite est sa limite en probabilité d'après le théorème précédent. On calcule alors $\mathbb{E}[|X_m - 0|^p] = \mathbb{E}[X_m^p] = n.\frac{1}{n} = 1$ car $X_m = Y_{n,j}$. Donc X_m ne tend pas vers 0 dans L^p.

Exemple. *($Ł^p \not\Longrightarrow$ p. s.)*

La convergence L^p n'implique pas la convergence presque sûre.

On reprend le premier exemple. $Y_{n,j} = \mathbb{1}_{]\frac{j-1}{n},\frac{j}{n}]}$ où X_m tend vers 0 en proba et X_m ne tend pas vers 0 presque sûrement. Mais $\mathbb{E}[|X_m - 0|^p] = \mathbb{E}[X_m^p] = \frac{1}{n} \longrightarrow 0$ pour $m \to \infty$ donc X_m tend vers 0 dans L^p.

Exemple. *(p. s. $\not\Longrightarrow$ L^p)*

La convergence presque sûre n'implique pas la convergence L^p.

Sur $([0,1], \mathcal{B}([0,1]))$ muni de la mesure de Lebesgue. Prenons $X_n(\omega) = \mathbb{1}_{[0,\frac{1}{n}]}(\omega)$ pour tout $\omega \in [0,1]$. On a bien que X_n converge presque sûrement vers la variable nulle : en effet, $X_n(\omega) \longrightarrow 0$ pour tout $\omega \in]0,1]$ et la mesure Lebesgue du singleton 0 est nulle. Par contre $\mathbb{E}[|X_n - 0|^p] = n^{\frac{1}{p}}\frac{1}{n}$ ne tend pas vers zéro. Donc X_n ne tend pas vers 0 dans L^p.

Exercice 5

Dans l'exemple précédent, déterminer la loi de la limite.

1.10.3 Réciproques partielles aux théorèmes de comparaison

Théorème. *(Convergence en proba $\Longrightarrow$ convergence p. s. le long d'une sous-suite)*

Soient $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. Si $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$, alors il existe une extractrice φ telle que $X_{\varphi(n)} \xrightarrow[n\to+\infty]{\text{p. s.}} X$.

$\triangleright$ La convergence en probabilité implique la propriété suivante : il existe $n \leqslant 1$ tel que $\mathbb{P}(|X_n - X|) > 1) \leqslant 2^{-2}$ pour tout $n \leqslant n_1$. Récursivement, pour tout $k \geqslant 2$, il existe $n_k > n_{k-1}$ tel que $\mathbb{P}(|X_n - X| > \frac{1}{k}) \leqslant 2^{-k}$ pour tout $n \geqslant n_k$. Montrons alors que $(X_{n_k})_{k\geqslant 1}$ converge presque sûrement vers X. On pose $A_k = \{|X_{n_k} - X| > \frac{1}{k}\}$. Par le lemme de Borel-Cantelli. D'où $\mathbb{P}(\limsup_{k\to+\infty} A_k) = 0$. Par ailleurs tout $\omega \in \Omega$ tel que $\neg(X_{n_k}(\omega) \longrightarrow X(\omega))$ doit appartenir à $\limsup_{k\to+\infty} A_k$. En effet pour un tel ω, il existe $\varepsilon > 0$ tel que $|X_{n_k}(\omega) - X(\omega)| > \varepsilon$ pour une infinité de $k \geqslant 1$. Or $\frac{1}{k}$ passe en dessous de ε, donc $\omega \in \limsup_{k\to+\infty} A_k$. Ainsi, $X_{n_k} \xrightarrow[n\to+\infty]{\text{p. s.}} X$. ■

Comme l'a montré le ce précédent, la convergence en probabilité ne donne pas de contrôle quantitatif sur la limite : il est possible que les moments explosent. La domination permet de contrôler ceci d'où :

Théorème. *(Convergence en proba + Domination L^p $\Longrightarrow$ convergence L^p)*

Soient $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. Si $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$ et s'il existe $Y \in L^p$ telle que $|X_n| \leqslant Y$ presque sûrement pour tout $n \geqslant 1$, alors $X_{\varphi(n)} \xrightarrow[n\to+\infty]{\text{p. s.}} X$.

$\triangleright$ Comme $|X_n| \leqslant Y \in L^p$, $X_n \in L^p$. Par ailleurs, il existe $(n_k)_{k\geqslant 1}$ telle que X_{n_k} converge presque sûrement vers X. Par le lemme de Fatou :

$$\mathbb{E}[|X|^p] = \mathbb{E}[\liminf_{k\to-\infty} |X_{n_k}|^p] \leqslant \liminf_{k\to+\infty} \mathbb{E}[|X_{n_k}|^p] \leqslant \mathbb{E}[|Y|^p] < \infty.$$

Donc $X \in L^p$. Il reste à montrer que X_n tend vers X dans L^p. Par l'absurde, supposons que ce ne soit pas le cas. Il existe $\varepsilon > 0$ et $(n_k)_{k\geqslant 1}$ tel que $\mathbb{E}[|X_{n_k} - X|^p] > \varepsilon$. On sait que X_{n_k} tend vers X en probabilité. Par le théorème précédent, il existe $(n_{k_j})_{j\geqslant 1}$ tel que $X_{n_{k_j}}$ tende presque sûrement vers X. Par ailleurs, $|X_{n_{k_j}} - X|^p \leqslant 2^p|Y|^p$. Par le théorème de convergence dominée, $\mathbb{E}[|X_{n_{k_j}} - X|^p]$ tend vers 0 quand j tend vers l'infini. Contradiction. ∎

Remarque. Si l'on connaît déjà la limite X, il suffit d'établir un contrôle $X_n - X \leqslant Y$.

Rappel. La domination implique la bornitude dans L^p, mais la réciproque est fausse : dans tout L^p, on peut trouver une suite bornée qui n'est pas dominée.

Théorème. ***(Convergence en proba + Borne L^p $\implies$ convergence $L^{p-\varepsilon}$)***

Soient $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. Si $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$ et si $\sup\limits_{n\geqslant 1} \|X_n\|_{L^p} < \infty$ alors $X_{\varphi(n)} \xrightarrow[n\to+\infty]{L^q} X$ pour tout $q \in [1,p[$.

▷ Comme dans la preuve précédente, $X \in L^p$ d'après le lemme de Fatou. On fixe $q \in [1,p[$ et l'on pose $\delta = p - q$. Pour $\varepsilon > 0$, on calcule $\mathbb{E}[|X_n - X|^q] = \mathbb{E}[|X_n - X|^q \mathbb{1}_{\{|X_n - X| \leqslant \varepsilon\}}] + \mathbb{E}[|X_n - X|^q \mathbb{1}_{\{|X_n - X| < \varepsilon\}}] \leqslant \varepsilon^q + \mathbb{E}[|X_n - X|^q \mathbb{1}_{\{|X_n - X| < \varepsilon\}}]$. Pour contrôler le second terme, on applique l'inégalité de Hölder avec $a = \frac{p}{q}$ et $b = \frac{p}{p-q}$. Ainsi $\mathbb{E}[|X_n - X|^q \mathbb{1}_{\{|X_n - X| > \varepsilon\}}] \leqslant \mathbb{E}[(|X_n - X|^q)^a]^{\frac{1}{a}} \mathbb{E}[\mathbb{1}^b_{\{|X_n - X| > \varepsilon\}}]^{\frac{1}{b}} \leqslant \mathbb{E}[(|X_n - X|^p]^{\frac{q}{p}} \mathbb{P}\{|X_n - X| > \varepsilon\}^{\frac{p-q}{p}}$. Ainsi, $\mathbb{E}[|X_n - X|^q \mathbb{1}_{\{|X_n - X| < \varepsilon\}}] \leqslant C^q \mathbb{P}\{|X_n - X| > \varepsilon\}^{\frac{p-q}{p}} \longrightarrow 0$. Or $\|X_n - X\|_{L^p} \leqslant \|X_n\|_{L^p} + \|X\|_{L^p} \leqslant \sup\limits_{m\geqslant 1} \|X_m\|_{L^p} + \|X\|_{L^p} < C < \infty$ pour un certain $C > 0$. Ainsi $\limsup\limits_{n\to+\infty} \mathbb{E}[|X_n - X|^q] \leqslant \varepsilon^p$ pour tout $\varepsilon > 0$. Ainsi cette limite supérieur est nulle, en particulier, $\mathbb{E}[|X_n - X|^q]$ tend vers 0 lorsque $n \to \infty$. ∎

Théorème. ***(Composition de la limite presque sûre)***

Soit $f : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continue. Alors si $X_n \xrightarrow[n\to+\infty]{\text{p. s.}} X$, alors $f(X_n) \xrightarrow[n\to+\infty]{\text{p. s.}} f(X)$.

▷ Nul. Remarquer qu'on garde le même ensemble de convergence presque sûre. ∎

Théorème. ***(Composition de la limite en proba)***

Soit $f : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continue. Alors si $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$, alors $f(X_n) \xrightarrow[n\to+\infty]{\mathbb{P}} f(X)$.

▷ Par l'absurde, on suppose que $f(X_n)$ ne converge pas $f(X)$ en probabilité. Ainsi il existe $\varepsilon > 0$ tel que $\mathbb{P}(|f(X_n) - f(X)| > \varepsilon)$ ne tend pas vers zéro. Ainsi il existe $(n_k)_{k\geqslant 1}$, et il existe $\delta > 0$ tel que $\mathbb{P}(|f(X_n) - f(X)| > \varepsilon) > \delta$ pour tout $k \geqslant 1$. Par un théorème précédent, il existe $(n_{k_j})_j$ tel que $(X_{n_{k_j}})$ converge presque sûrement vers X. Par le résultat précédent, $f(X_{n_{k_j}})$ tend presque sûrement vers $f(X)$. Or la convergence presque sûrement implique la convergence en probabilité, absurde. ∎

Exercice 6

Montrer qu'il n'y a pas de composition de la convergence en moyenne L^p.

▷ **Éléments de réponse.**

Le contre-exemple est simple : il faut considérer la bonne puissance x^α. En fait, on peut montrer que c'est vrai si l'on impose à f une hypothèse d'uniforme continuité ou plutôt, de croissance linéaire à l'infini.

Propriété. ***(Lemme de Scheffé)***

Soient (X_n) une suite de var L^p définies sur le même espace. Si X_n converge presque sûrement vers X et $\mathbb{E}[|X_n|^p]$ tend vers $\mathbb{E}[|X|^p]$, alors X_n tend vers X dans L^p.

▷ Preuve identique au théorème de convergence dominée : on applique le lemme de Fatou à une double différence. ■

Théorème. ***(Théorème d'Egoroff)***

Soit (X_n) une suite de var définies sur le même espace. Si X_n converge presque sûrement vers X, alors pour tout $\varepsilon > 0$, il existe un événement N tel que $\mathbb{P}(N) < \varepsilon$ et X_n converge uniformément vers X sur $\Omega \setminus N$.

▷ C'est une manipulation de type Borel-Cantelli : pour tout $k \geqslant 1$ et $n \geqslant 0$, on introduit $A_n^{(k)} = \{\omega \in \Omega, |X_n(\omega) - X(\omega)| > \frac{1}{k}\}$. Par définition de la convergence simple, $\mathbb{P}(\limsup_{n\to+\infty} A_n^{(k)}) = 0$. De plus, par continuité décroissante, pour tout $k \geqslant 1$ et $\eta > 0$, il existe $n \geqslant 0$ tel que $\mathbb{P}\left(\bigcup_{j\geqslant n}\{\omega \in \Omega, |X_j(\omega) - X(\omega)| > \frac{1}{k}\right)\} \leqslant \eta$. Ainsi, pour tout $\varepsilon > 0$, pour tout $k \geqslant 1$, il existe n_k tel que ce terme soit $\leqslant \varepsilon 2^{-k}$. En notant A leur réunion, $\mathbb{P}(A) \leqslant \sum \varepsilon 2^{-k} \leqslant \varepsilon$. Ainsi $\mathbb{P}(A^c) \geqslant 1 - \varepsilon$ et X_n converge uniformément sur A^c. ■

1.10.4 Loi des grands nombres

On se donne un espace de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$.

Théorème. ***(Loi faible des grands nombres)***

Soit $(X_n)_{n\geqslant 1}$ une suite de v. a. r. indépendantes et de même loi. Alors :

(i) si $\mathbb{E}[|X_1|] < \infty$, alors $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[X_1]$.

(ii) si X_1 est de carré intégrable, alors $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow[n\to\infty]{L^2} \mathbb{E}[X_1]$.

▷ Montrons d'abord le deuxième point, dont le cadre est plus agréable. En écrivant la convergence L^2, puisque les X_j ont la même loi, on peut centrer les variables dans la somme. En développant le carré, on peut séparer les termes diagonaux et les termes extradiagonaux dont l'espérance est nulle par indépendance deux à deux. On obtient $\frac{\mathrm{Var}(X_1}{n}$ qui tend vers 0. Pour le second point, on n'a plus d'égalité. Il va falloir majorer la probabilité que l'écart soit grand. On veut majorer $\mathbb{P}(|\frac{1}{n}\sum_{j=1}^{n} X_j - \mathbb{E}[X_1]| > \varepsilon)$. On veut appliquer Tchebychev, mais on ne sait pas si le carré est intégrable. Astuce classique : on coupe la variable en deux. Pour $N > 0$ fixé, on pose $X_{j,N} := X_j \mathbb{1}_{\{|X_j| \leqslant N\}}$ et $X_{j,N^c} := X_j \mathbb{1}_{\{|X_j|>N\}}$, parties bornée et libre. On a $X_j = X_{j,N} + X_{j,N^c}$. Les $(X_{j,N})_{j\geqslant 1}$ sont indépendantes par coalition, de même loi et bornées, donc on peut appliquer le second point. Ainsi, $\mathbb{P}(|\frac{1}{n}\sum_{j=1}^{n} X_{j,N} - \mathbb{E}[X_{1,N}]| > \varepsilon) \leqslant \dfrac{\mathbb{E}[|\frac{1}{n}\sum_{j=1}^{n} X_{j,N} - \mathbb{E}[X_{1,N}]|^2]}{\varepsilon^2} \leqslant \dfrac{\mathrm{Var}(X_{1,N})}{n\varepsilon^2}$. Pour les X_{j,N^c}, on utilise un contrôle L^1. On a $\mathbb{P}(|\frac{1}{n}\sum_{j=1}^{n} X_{j,N^c} - \mathbb{E}[X_{1,N^c}]| > \varepsilon) \leqslant \mathbb{P}(\dfrac{1}{n}\sum_{j=1}^{n} |X_{j,N^c}| + |\mathbb{E}[X_{1,N^c}]| > \varepsilon) \leqslant \dfrac{1}{\varepsilon}\mathbb{E}[\dfrac{1}{n}\sum_{j=1}^{n} |X_{j,N^c}| + \mathbb{E}[|X_{1,N^c}|]] \leqslant \dfrac{2\mathbb{E}[X_{1,N^c}]}{\varepsilon}$ puisque les $X_{j,N}$ sont de même loi. Or, $\{|\frac{1}{n}\sum_{j=1}^{n} X_j - \mathbb{E}[X_1]| > \varepsilon\} \subseteq \{|\dfrac{1}{n}\sum_{j=1}^{n} X_{j,N} - \mathbb{E}[X_{1,N}]| > \varepsilon/2\} \cup \{|\dfrac{1}{n}\sum_{j=1}^{n} X_{j,N^c} - \mathbb{E}[X_{1,N^c}]| > \varepsilon/2\}$. Ainsi, $\mathbb{P}(|\frac{1}{n}\sum_{j=1}^{n} X_j - \mathbb{E}[X_1]| > \varepsilon) \leqslant \mathbb{P}(|\dfrac{1}{n}\sum_{j=1}^{n} X_{j,N} - \mathbb{E}[X_{1,N}]| > \varepsilon/2) \cup \mathbb{P}(|\dfrac{1}{n}\sum_{j=1}^{n} X_{j,N^c} - \mathbb{E}[X_{1,N^c}]| > \varepsilon/2) \leqslant \dfrac{4}{n\varepsilon^2}\mathrm{Var}(X_{1,N}) + \dfrac{4}{\varepsilon}\mathbb{E}[|X_{1,N^c}|]$. Or par convergence dominée, $\mathbb{E}[|X_{1,N^c}|]$ tend vers zéro pour $N \to +\infty$. Pour tout $\delta > 0$, il existe donc $N > 0$ tel que $\mathbb{E}[|X_{1,N^c}|] < \delta$. Pour ce choix de N, on obtient alors $\mathbb{P}(|\frac{1}{n}\sum_{j=1}^{n} X_j - \mathbb{E}[X_1]| > \varepsilon) \leqslant \dfrac{4}{n\varepsilon^2}\mathrm{Var}(X_{1,N}) + \dfrac{4\delta}{\varepsilon}$ d'où $\limsup_{n\to+\infty} \mathbb{P}(|\frac{1}{n}\sum_{j=1}^{n} X_j - \mathbb{E}[X_1]| > \varepsilon) \leqslant \frac{4\delta}{\varepsilon}$. Comme δ est arbitraire, on en déduit que $\limsup_{n\to+\infty} \mathbb{P}(|\frac{1}{n}\sum_{j=1}^{n} X_j - \mathbb{E}[X_1]| > \varepsilon) = 0$, d'où $\lim_{n\to+\infty} \mathbb{P}(|\frac{1}{n}\sum_{j=1}^{n} X_j - \mathbb{E}[X_1]| > \varepsilon)$, ce qu'il fallait montrer. ■

Les hypothèses de la loi faible des grands nombres sont largement atténuables. Citons que l'on peut ne supposer que :

1. les variables deux à deux indépendantes,
2. voire, deux à deux non corrélée,
3. tout simplement de carré intégrables (et l'une ou l'autre des trois hypothèses ci-dessus).

Théorème. ***(Loi forte des grands nombres)***

Soit $(X_n)_{n\geqslant 1}$ une suite de v. a. r. indépendantes et de même loi et telles que $\mathbb{E}[|X_1|] < \infty$. Alors $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow[n\to+\infty]{\text{p. s.}} \mathbb{E}[X_1]$.

▷ On pose $S_n = X_1 + ... + X_n$ pour $n \geqslant 1$. On veut montrer que $\frac{S_n}{n}$ converge presque sûrement vers $\mathbb{E}[X_1]$. Si l'on peut prouver que $\limsup\limits_{n\to+\infty} \frac{S_n}{n} \leqslant \mathbb{E}[X_1]$, alors par symétrie, presque sûrement, $\limsup\limits_{n\to+\infty} \frac{(-X_1-X_2-...-X_n}{n} \leqslant \mathbb{E}[-X_1]$ presque sûrement, c'est-à-dire p. s. $-\liminf\limits_{n\to+\infty} \frac{S_n}{n} \leqslant -\mathbb{E}[X_1]$, soit presque sûrement, $\liminf\limits_{n\to+\infty} \frac{S_n}{n} \geqslant \mathbb{E}[X_1]$. En combinant les deux, presque sûrement, $\frac{S_n}{n} \longrightarrow \mathbb{E}[X_1]$. Il suffit alors de montrer que pour tout $a \in \mathbb{R}$ tel que $a > \mathbb{E}[X_1]$, presque sûrement, $\limsup\limits_{n\to+\infty} \frac{S_n}{n} \leqslant a$. On introduit alors $M = \sup\limits_n 0(S_n - na)$. Il suffit de montrer que $M < \infty$ presque sûrement, car alors presque sûrement, $\frac{S_n}{n} \leqslant a + \frac{M}{n}$ d'où presque surement, $\limsup\limits_{n\to+\infty} \frac{S_n}{n} \leqslant a$. ■

Corollaire. ***(Projection de la loi des grands nombres)***

Soient (X_n) une suite de variables indépendantes et de même loi à valeurs dans $(E, \mathcal{E})$. Pour tout $f : E \longrightarrow \mathbb{R}$ mesurable telle que $\mathbb{E}([|f(X_1)|] < \infty$, alors $\frac{1}{n}\sum\limits_{i=1}^{n} f(X_i) \xrightarrow[n\to+\infty]{\text{p. s.}} \mathbb{E}[f(X_1)]$.

▷ La loi forte des grands nombres s'applique aux $f(X_i)$, indépendantes et de même loi. ■

Corollaire. ***(Projection double de la loi des grands nombres)***

Soient (X_n) une suite de variables indépendantes et de même loi à valeurs dans $(E, \mathcal{E})$. Pour tout $g : E \times E \longrightarrow \mathbb{R}$ mesurable telle que $\mathbb{E}([|g(X_1,X_2)|] < \infty$, alors $\frac{1}{n}\sum\limits_{i=1}^{n} g(X_i,X_{i+1}) \xrightarrow[n\to+\infty]{\text{p. s.}}$ $\mathbb{E}[g(X_1,X_2)]$.

▷ On applique la loi forte à $\frac{1}{n}\sum\limits_{i=1}^{n/2} g(X_{2i}, X_{2i+1})$ et séparément à $\frac{1}{n}\sum\limits_{i=1}^{n/2} g(X_{2i-1}, X_{2i})$. Pour chacun, j'obtiens une convergence presque sûre vers $\frac{1}{2}\mathbb{E}[g(X_1,X_2)]$. En sommant, on obtient le résultat. ■

La loi forte des grands nombres *ne rend pas caduque* pour le moins du monde la loi faible. Si la notion de convergence obtenue est plus forte, attention, la vitesse de convergence est amoindrie, ou en tout cas, n'est pas précisée ! (Rappelons qu'elle est quadratique dans le cas L^2 de la loi faible des grands nombres.)

1.10.5 Convergence en loi

On n'a plus besoin d'être sur le même espace ! Car les autres étaient des convergences fonctionnelles (en tant qu'applications définies sur un même espace). De même que précédemment, on se limite à R mais la dimension supérieur est sans problème.

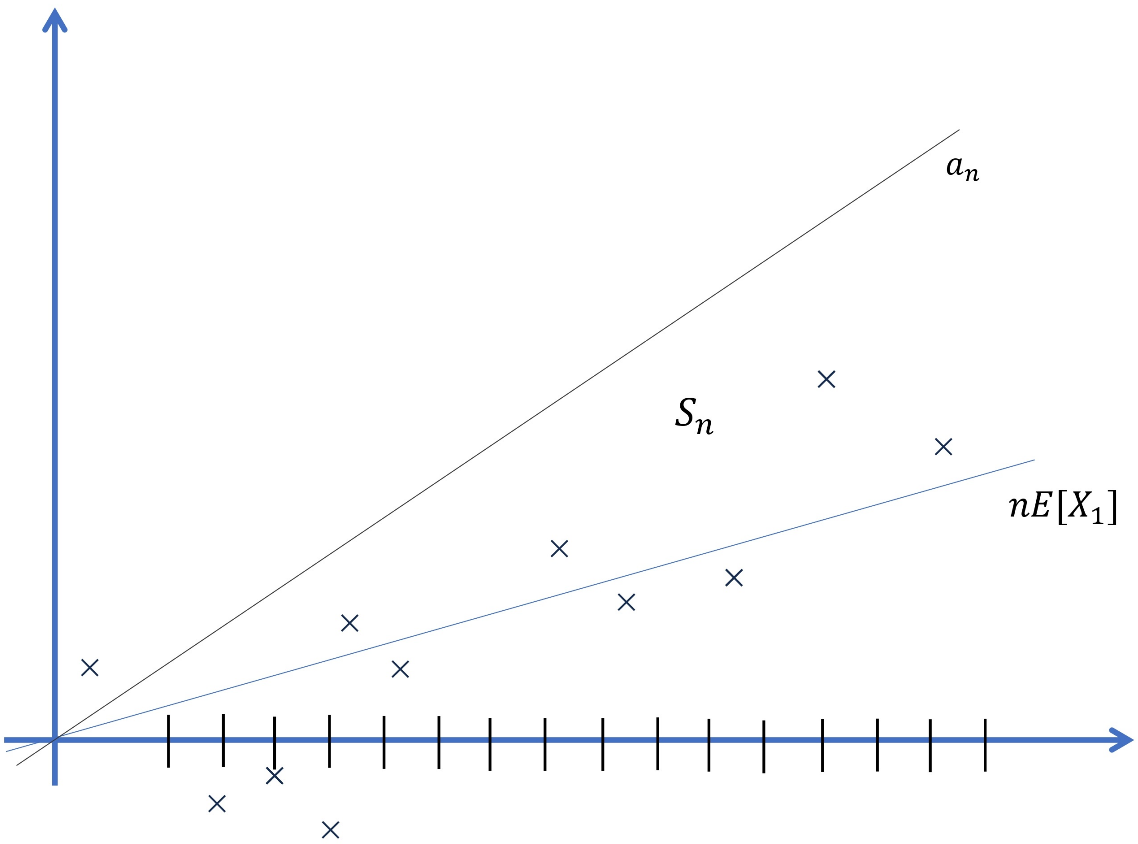


Figure 1.10.1 : *Explication statistique de la loi forte des grands nombres.* —

Définition. *(Convergence en loi)*

Soient X_n, $n \geqslant 1$ et X des va à valeurs dans $\mathbb{R}^d$, par forcément sur le même espace de probabilité. On dit que $(X_n)_{n\geqslant 1}$ converge en loi vers X, et l'on note $X_n \xrightarrow[n\to+\infty]{\text{loi}} X$ si la suite $(\mathcal{L}_X)_{n\geqslant 1}$ converge étroitement vers $\mathcal{L}_X$, c'est-à-dire, pour toute fonction continue bornée $h : \mathbb{R}^d \longrightarrow \mathbb{R}$, on a $\lim_{n\to+\infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)]$.

Ceci ne dépend que de la loi des X_n, et d'aucun loi jointe ; on peut changer les variables, une à une ou toute, pour des variables de même loi, ça ne change rien ; cependant, toutes les variables doivent prendre leurs valeurs dans le même ensemble qui est le départ de h.

La considération précédente permet l'abus qui n'est pas un abus d'écrire la convergence en loi pour des lois, et non plus des variables. On le rencontrera souvent pour des théorèmes et en exercice.

Remarques.

1. $E(h(X_n)) = \int_{\mathbb{R}^d} h(x) d\mathcal{L}_{X_n}(x)$.

2. Encore une fois, cette notion de convergence ne dépend que de la loi de chacune des X_n.
3. On dit parfois que $(X_n)_{n\geqslant 1}$ converge en distribution vers X. Ici le mot distribution est synonyme de loi et n'est pas relié à la théorie des distributions.
4. On peut généraliser $\mathbb{R}^d$ un espace topologique muni de sa tribu borélienne, et dans ce cas h doit être une fonction continue bornée de cet espace dans $\mathbb{R}$.

Attention ! Cela na pas de sens de dire que la différence de $h(X_n) - h(X)$ tend vers 0 en espérance, car sans autre précision elles ne sont pas définies sur le même espace. C'est pour ça que le lemme de Scheffé, la convergence L^p sont meilleures pour ces considérations.

On verra dans beaucoup d'exemples que c'est la bonne définition. On parlera de CV en loi si la loi converge vers une loi limite. Si les lois dépendent de n, on aura également un lien avec la convergence des fonctions de masse ou des densités.

On énonce le théorème suivant à $\mathbb{R}^d$ pour $d = 1$ mais ça se généralise à des détails près. Le lemme fonctionne sans modification pour $\mathbb{R}^d$ et $\mathbb{R}^k$.

Lemme

Soit $g : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continue. Si X_n converge en loi vers X, alors $g(X_n)$ converge en loi vers $g(X)$.

$\triangleright$ Soit $h : \mathbb{R} \longrightarrow \mathbb{R}$ continue bornée. Alors $\mathbb{E}[h(g(X_n))] = \mathbb{E}[(h\circ g)(X_n)]$. Or $h \circ g$ est continue bornée. Par convergence étroite, $\mathbb{E}[h \circ g(X_n)]$ tend vers $\mathbb{E}[h \circ g(X)]$ quand n tend vers $+\infty$, qui égale $\mathbb{E}[h(g(X))]$ et c'est terminé. ■

Exemples

1. Soit X_n de loi uniforme sur $\{\frac{1}{n}, \frac{2}{n}, ..., 1\}$. Nous montrons que X_n tend en loi vers X qui suit une loi uniforme sur $[0,1]$. Soit $h : \mathbb{R} \longrightarrow \mathbb{R}$ continue bornée. Alors $\mathbb{E}[h(X_n)] = \frac{1}{n}\sum_{k=1}^{n} h(\frac{k}{n})$. Cette somme de Riemann converge quand $n \to \infty$ vers $\int_0^1 h(x)\mathrm{d}x = \mathbb{E}[h(X)]$ par transfert.
2. Soit $X_n \sim \mathcal{N}(0, \sigma_n^2)$ avec σ une suite d'écarts typiques réels qui tendent vers zéro. Alors X_n tend en loi vers X où X est une variable aléatoire presque sûrement égale à 0.

On peut affaiblir les contraintes sur la fonction h :

Proposition. ***(Hypothèse faible de la convergence en loi)***

On a que X_n converge en loi vers X si et seulement si pour toute fonction continue à support compact $g : \mathbb{R} \longrightarrow \mathbb{R}$, on a :

$$\mathbb{E}[g(X_n)] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[g(X)].$$

▷ Bien évidemment, il suffit de montrer le sens réciproque, car la classe des fonctions continues bornées est plus grande que la classe des fonctions à supports compacts. L'idée graphique est toujours la même (c'est le principe des indicatrices régularisées) : à une fonction h continue, on fait correspondre une fonction g_L continue à support compact dans $[-L-1,L+1]$ en gardant h sur $[-L,L]$, et en la reliant à 0 sur les deux extrémités de longueur 1 du segment, ce qui peut se faire de manière C^∞ (d'où la propriété suivante). Dans ce cas, on aurait

$$|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| = |\mathbb{E}[h(X_n)] - \mathbb{E}[g_L(X_n)] + \mathbb{E}[g_L(X_n)] - \mathbb{E}[g_L(X)] + \mathbb{E}[g_L(X)] - \mathbb{E}[h(X)]|$$

ce qui nous donne trois termes à contrôler. Définissons rigoureusement g_L : on introduit $\psi_L : \mathbb{R} \longrightarrow \mathbb{R}$ qui satisfait $\psi_L(x) \in [0,1]$ pour tout $x \in \mathbb{R}$, $\psi_L(x) = 0$ si $x \in]-\infty, -L-1] \cup [L+1, +\infty[$, $\psi_L(x) = 1$ si $x \in [-L,L]$ et ψ_L est continue. On pose alors $g_L(x) = h(x)\psi_L(x)$ pour tout réel x. Les fonctions g_L et ψ_L sont continues à support compact, donc $\lim\limits_{n\to+\infty} \mathbb{E}[\psi_L(X_n)] = \mathbb{E}[\psi_L(x)]$. Ainsi le deuxième terme tend vers zéro quand n tend l'infini pour tout $L > 0$ fixé. Reste à contrôler les deux autres termes. On utilise un dépassement de Markov par TCD. Par ailleurs, $\mathbb{1}_{|X|\geqslant L}$ tend vers la fonction nulle presque sûrement pour L tendant vers l'infini, donc par théorème de convergence dominée, $\mathbb{P}(|X| \geqslant L)$ tend vers 0 quand L tend vers l'infini. Ainsi, pour tout $\varepsilon > 0$ fixé, il existe $L_0 > 0$ tel que pour tout $L \geqslant L_0$, $\mathbb{P}(|X| \geqslant L) < \varepsilon$ et donc $\mathbb{E}[1 - \psi_L(X)] \leqslant \mathbb{P}(|X| \geq L) < \varepsilon$ d'où $|\mathbb{E}[h(X)] - \mathbb{E}[g_L(X)]| = |\mathbb{E}[h(X) - g_L(X)]| = |\mathbb{E}[h(X)(1-\psi_L(X))]| \leqslant \|h\|_\infty \mathbb{E}[1-\psi_L(X)] \leqslant \|h\| \varepsilon$. Puisque $\mathbb{E}[\psi_L(X_n)]$ tend vers $\mathbb{E}[\psi_L(X)]$, on a $\mathbb{E}[1-\psi_L(X_n)]$ qui tend vers $\mathbb{E}[1-\psi_L(X)]$. Ainsi pour tout $L \geqslant L_0$ et pour tout n assez grand, $\mathbb{E}[1-\psi_L(X_n)] \leqslant \varepsilon + \mathbb{E}[1-\psi_L(X)] \leqslant 2\varepsilon$. Ainsi, le premier terme, qui vaut $|\mathbb{E}[h(X_n)] - \mathbb{E}[g_L(X_n)]| = |\mathbb{E}[h(X_n) - g_L(X_n)]| = |\mathbb{E}[h(X_n)(1-\psi_L(X_n))]| \leqslant \|h\|_\infty \mathbb{E}[1-\psi_L(X_n)] \leqslant \| h\|_\infty 2\varepsilon$. ■

Proposition. ***(Hypothèse (très !) faible de la convergence en loi)***

On a que X_n converge en loi vers X si et seulement si pour toute fonction C^∞ à support compact $g : \mathbb{R} \longrightarrow \mathbb{R}$, on a :

$$\mathbb{E}[g(X_n)] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[g(X)].$$

Propriété. ***(Lemme de Slutsky)***

Soient (X_n) et (Y_n) deux suites de variables aléatoires. On suppose que X_n converge en loi vers X et que Y_n converge en probabilité vers une constante c. Alors (X_n,Y_n) converge en loi vers (X,c).

1.10.5.1 Cas particulier des variables à fonctions de masse

Proposition. ***(Convergence en loi par les fonctions de masse)***

Soient X_n, $n \geqslant 1$ et X des variables aléatoires discrètes prenant leurs valeurs dans E. Si $\lim_{n \to +\infty} \mathbb{P}(X_n = a) = \mathbb{P}(X = a)$ pour tout $a \in E$, alors $X_n \xrightarrow[n\to+\infty]{\text{loi}} X$.

▷ Soit $h : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continue bornée et soit $\varepsilon > 0$. Il existe $k \geqslant 1$ un entier et $a_1,\ldots,a_k \in E$ tels que $\mathbb{P}(X \in \{a_1,\ldots,a_k\}) > 1 - \varepsilon$. Par ailleurs, l'hypothèse de l'énoncé assure que $\mathbb{P}(X_n \in \{a_1,\ldots,a_k\}) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}(X \in \{a_1,\ldots,a_k\})$. Ainsi, pour tout naturel n assez grand, $P(X_n \in \{a_1,\ldots,a_k\}) > 1 - 2\varepsilon$. On calcule alors $\mathbb{E}[h(X_n)] = \sum_{a\in E} h(a)\mathbb{P}(X_n = a) = \sum_{a\in\{a_1,\ldots,a_k\}} h(a)\mathbb{P}(X_n = a) + \sum_{a\in E\setminus\{a_1,\ldots,a_k\}} h(a)\mathbb{P}(X_n = a)$. Or $\sum_{a\in\{a_1,\ldots,a_k\}} h(a)\mathbb{P}(X_n = a) \underset{n\to+\infty}{\longrightarrow} \sum_{a\in\{a_1,\ldots,a_k\}} h(a)\mathbb{P}(X = a)$ comme somme finie de limites. Par ailleurs, $\left|\sum_{a\in E\setminus\{a_1,\ldots,a_k\}} h(a)\mathbb{P}(X_n = a)\right| \leqslant \|h\|_\infty \mathbb{P}(X_n \notin \{a_1,\ldots,a_k\}) \leqslant \|h\|_\infty 2\varepsilon$. Donc, pour conclure, $|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leqslant \left|\sum_{a\in\{a_1,\ldots,a_k\}} h(a)\mathbb{P}(X_n = a) - \sum_{a\in\{a_1,\ldots,a_k\}} h(a)\mathbb{P}(X = a)\right| + \left|\sum_{a\in E\setminus\{a_1,\ldots,a_k\}} h(a)\mathbb{P}(X_n = a) - \sum_{a\in E\setminus\{a_1,\ldots,a_k\}} h(a)\mathbb{P}(X = a)\right|$. Le premier terme tend vers zéro, le second terme est majoré par $3\varepsilon \|h\|_\infty$ en appliquant la même majoration sur le dernier terme. Ainsi, pour n assez grand, $|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leqslant 4\varepsilon \|h\|_\infty$. ■

Exemple

Si $X_n \sim Bin(n, \frac{\lambda}{n})$, alors X_n converge en loi vers $\mathcal{P}(\lambda)$. Typiquement, c'est une unique qui construit n pièces avec un grand paramètre et une petit probabilité que chacune, dont la production est indépendante des autres, soit défectueuse.

Il arrive que des variables discrètes tendent vers une variable continue (exemple ?).

Proposition. ***(Convergence en loi par les densités)***

Soient X_n, $n \geqslant 1$ une suite de variables aléatoires à valeurs dans $\mathbb{R}$, et à densités $(f_n)_{n\geqslant 1}$. On suppose qu'il existe $g : \mathbb{R} \longrightarrow \mathbb{R}$ intégrable contre la mesure de Lebesgue telle que $f_n \leqslant g$ Lebesgue-presque partout. Alors si $\lim_{n\to\infty} f$ Lebesgue presque partout, alors f est une densité de probabilité sur $\mathbb{R}$ et X_n tend en loi vers une variable aléatoire X à densité f.

▷ On applique le théorème de convergence dominée à f_n. Ceci nous donne la normalisation de f_n. Elle est positive par limite simple et mesurable également par limite simple ; c'est donc une densité de probabilité. Par ailleurs, soit $h : \mathbb{R} \longrightarrow \mathbb{R}$ continue bornée. Alors $\mathbb{E}[h(X_n)] = \int_{\mathbb{R}} h(x) f_n(x) \mathrm{d}x$. Or $|h(x)f_n(x)| \leqslant \|h\|_\infty g(x)$ intégrable et $h(x)f_n(x)$ tend vers $h(x)f(x)$ Leb-presque partout. Par convergence, dominée, la limite de cette expression est $\mathbb{E}[h(X)]$, par transfert. ■

1.10.5.2 Comparaison avec les autres notions de convergence

Proposition. ***(Convergence en proba $\implies$ convergence en lo)***

Soient X_n, $n \geqslant 1$ et X des variables aléatoires définies sur un même espace de proba. Si X_n tend vers X en probabilité, alors X_n tend vers X en loi.

▷ Soit $h : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continue bornée. Comme X_n tend vers X en mesure, et comme h est continue, par un résultat précédent, $h(X_n)$ tend vers $h(X)$ en probabilité. Par ailleurs, $|h(X_n)| \leqslant \|h\|_\infty$ qui est intégrable, donc $h(X_n)$ tend vers $h(X)$ dans L^1. En particulier, $\mathbb{E}[h(X_n)] \longrightarrow \mathbb{E}[h(X)]$. ■

Par conséquence, la convergence en moyenne ou presque sûre implique la convergence en loi. Il y a des contre-exemples pour toutes les réciproques (un suffit dans notre cas).

Proposition. ***(Convergence en loi + limite déterministe $\implies$ convergence en proba)***

Soit (X_n) une suite de variables aléatoires réelles sur $(\Omega, \mathcal{A}, \mathbb{P})$. Si X_n tend vers X en loi et si $X = a$ presque sûrement pour $a \in \mathbb{R}$. Alors X_n tend vers X en probabilité.

▷ Soit ε. Il suffit de montrer que $\mathbb{P}(|X_n - a| > \varepsilon)$ tend vers zéro pour n à l'infini. Soit $h : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction satisfaisant $h(x) = 0$ pour tout $x \in [a - \varepsilon/2, a + \varepsilon/2]$, $h(x) = 1$ pour tout $x \in]-\infty, a-\varepsilon] \cup [a+\varepsilon, \infty[$, et $h(x) \in [0{,}1]$ pour tout $x \in \mathbb{R}$. On a $h(x) \geqslant \mathbb{1}_{\{|x-a|>\varepsilon\}}$ et ainsi $\mathbb{P}(|X_n a| > \varepsilon) = \mathbb{E}[\mathbb{1}_{\{|x-a|>\varepsilon\}}] \leqslant \mathbb{E}[h(X_n)]$. Or $\mathbb{E}[h(X_n)] \longrightarrow \mathbb{E}[h(X)] = h(a) = 0$ donc $\mathbb{P}(|X_n - a| > \varepsilon) \longrightarrow 0$. ■

1.10.5.3 Récapitulatif des convergences des variables aléatoires

1.10.5.4 Théorème de la fonction de répartition

Théorème. ***(Convergence en loi $\implies$ convergence des fonctions de répartition)***

Soient $X_n, n \geqslant 1$ et X des variables aléatoires réelles. Alors X_n tend vers X en loi ssi F_{X_n} converge vers F_X en tout point de continuité de F_X.

▷ Soit $F_{X_n}(t) = \mathbb{P}(X_n \leqslant t) = \mathbb{E}[\mathbb{1}_{]-\infty,t]}(X_n)]$. On fixe $\varepsilon > 0$. Dans ce cas, on pose $H_{\varepsilon,t}(x) = \begin{cases} 1 \text{ si } x \leqslant t \\ 1 - \frac{x-t}{\varepsilon} \text{ si } t < x < t+\varepsilon \\ 0 \text{ sinon.} \end{cases}$ $H_{\varepsilon,t}$ est continue bornée et l'on a $\mathbb{1}_{]-\infty,t]}(x) \leqslant H_{\varepsilon,t}(x) \leqslant \mathbb{1}_{]-\infty,t+\varepsilon]}(x)$ pour tout $x \in \mathbb{R}$. Ainsi $\lim_{n\to+\infty} \mathbb{E}[H_{\varepsilon,t}(X_n)] = \mathbb{E}[H_{\varepsilon,t}(X)]$. On rappelle que limites inférieure et supérieure sont toujours définies pour une suite de fonctions, et qu'on ne peut parler de limite simple que si elles sont égales. On étudie donc les limites aux bornes. On a

$$\begin{aligned} \limsup_{n\to+\infty} F_{X_n}(t) &= \limsup_{n\to+\infty} \mathbb{E}[\mathbb{1}_{]-\infty,t]}(X_n)] \\ &\leqslant \limsup_{n\to+\infty} \mathbb{E}[H_{\varepsilon,t}(X_n)] = \lim_{n\to+\infty} \mathbb{E}[H_{\varepsilon,t}(X_n)] = \mathbb{E}[H_{\varepsilon,t}(X)] \\ &\leqslant \mathbb{E}[\mathbb{1}_{]-\infty,t+\varepsilon]}(X)] = F_X(t+\varepsilon). \end{aligned}$$

Par un raisonnement analogue, on montre que $F_X(t-\varepsilon) \leqslant \liminf_{n\to+\infty} F_{X_n}(t)$. Par conséquent, **si t est un point de continuité de F_X**, en faisant tendre ε vers 0, on obtient que $\liminf_{n\to+\infty} F_{X_n}(t) = \limsup_{n\to+\infty} F_{X_n}(t) = \lim_{n\to+\infty} F_{X_n}(t) = F_X(t)$.

Réciproquement, soit h une fonction C^∞ à support compact. On calcule

$$\begin{aligned} \mathbb{E}[h(X)] &= \int_{\mathbb{R}} h(x)\mathcal{L}_X(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \int_{[x,+\infty]} -h'(y)\mathrm{d}y\mathcal{L}_X(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} -h'(y)\mathbb{1}_{[x,+\infty]}(y)\mathrm{d}y\mathcal{L}_X(\mathrm{d}x) \\ &= \int_{\mathbb{R}} -h'(y) \int_{\mathbb{R}} \mathbb{1}_{[x,+\infty]}(y)\mathcal{L}_X(\mathrm{d}x)\mathrm{d}y \\ &= \int_{\mathbb{R}} -h'(y) \int_{\mathbb{R}} \mathbb{1}_{]-\infty,y]}(y)\mathcal{L}_X(\mathrm{d}x)\mathrm{d}y \\ &= \int_{\mathbb{R}} -h'(y)F_X(y)\mathrm{d}y. \end{aligned}$$

Par le même argument, $\mathbb{E}[h(X_n)] = \int_{\mathbb{R}} -h'(y)F_{X_n}(y)\mathrm{d}y$. Or l'ensemble des points de discontinuité d'une fonction de répartition est dénombrable, donc de mesure de Lebesgue nulle. Ainsi $-h'(y)F_{X_n}(y)$ tend quand $n \to \infty$ vers $-h'(y)F_X(y)$ Lebesgue-presque partout. Puisque $|-h'(y)F_{X_n}(y)| \leqslant |h'(y)|$ intégrable contre Lebesgue, le théorème de convergence dominée permet de conclure. ■

1.10.5.5 Tension

Soit I un ensemble quelconque, fini ou infini, dénombrable ou non.

Définition. ***(Variables aléatoires tendues)***

On dit qu'une collection de v. a. r. $(X_i)_{i\in I}$ est tendue si pour tout $\varepsilon > 0$, il existe un compact $K \subseteq \mathbb{R}$ tel que

$$\mathbb{P}(X_i \in K) \geqslant 1 - \varepsilon \qquad \forall i \in I.$$

Heuristique

Il s'agit de dire que le gros du support des X_i est au même endroit, ou plutôt, que la masse des variables est répartie topologiquement au même endroit.

Propriété. ***(Tension dans le cas seul)***

Si I est un singleton, alors la collection est tendue.

▷ Cela se voit sur la fonction de répartition. (Le compact pouvant être très grand !) ■

Propriété. ***(Tension d'une famille finie de variables)***

Toute collection finie de variables aléatoires réelles est tendue.

▷ On passe au minimum sur un ensemble fini. ■

On a les propriétés immédiates suivantes :

Propriété. ***(Réunion finie de collections tendues)***

Toute réunion de deux collections tendues est tendue.

Propriété. ***(Sous-suite d'une suite tendue)***

Toute sous-suite d'une suite tendue est tendue.

Contre-exemple. ***(Suite non tendue)***

Si X_n est constante égale à n, $(X_n)_{n\geqslant 1}$ n'est pas tendue. (En fait, par locale compacité de $\mathbb{R}$, c'est heuristiquement la seule manière de perdre la tension...) □

Contre-exemple. ***(Autre suite non tendue)***

Si $X_n \sim \mathcal{N}(n,1)$, (X_n) n'est pas tendue. □

On dispose de deux lemmes fondamentaux.

Lemme. ***(Bolzano-Weierstrass pour les suites-tendues)***

Soit $(X_n)_{n\geqslant 1}$ une suite tendue de variables aléatoires réelles. Alors il existe une sous-suite $(X_{n_k})_{k\geqslant 1}$ qui converge en loi vers une certaine v. a. X.

Lemme. ***(Convergence en loi $\implies$ tension)***

Soit $(X_n)_{n\geqslant 1}$ une suite de v. a. r. convergeant en loi. Alors la suite (X_n) est tendue.

▷ Soit X la limite en loi de $(X_n)_{n\geqslant 1}$. Soit $\varepsilon > 0$. Il existe $M > 1$ tel que $\mathbb{P}(|X| > M-1) < \varepsilon$. Soit h_M une fonction continue bornée vérifiant

$$\mathbb{1}_{]-\infty,-M]\cup[M,+\infty[}(x) \leqslant h_M(x) \leqslant \mathbb{1}_{]-\infty,-M+1]\cup[M-1,+\infty[}(x)$$

pour tout $x \in \mathbb{R}$. On a $\mathbb{E}[h_M(X_n)] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[h_M(X)]$. Or $\mathbb{E}[h_M(X_n)] \geqslant \mathbb{E}[\mathbb{1}_{]-\infty,-M]\cup[M,+\infty[}(x) \leqslant h_M(X_n)] = \mathbb{P}(|X_n| \geqslant M)$. De plus, $\mathbb{E}[h_M(X)] \leqslant \mathbb{E}[\mathbb{1}_{]-\infty,-M+1]\cup[M-1,+\infty[}(X)] = \mathbb{P}(|X| > M-1)$. DONC il existe $n_0 \geqslant 1$ tel que pour tout $n \geqslant n_0$, $\mathbb{P}(|X_n| \geqslant M) \leqslant 2\varepsilon$. On a montré que pour tout $n \geqslant n_0$, $\mathbb{P}(X_n \in [-M,M]) \geqslant 1-2\varepsilon$. ■

Heuristiquement, la tension est une notion asymptotique de deux propriétés mises ensemble en considérant les queues de distribution.

1.10.5.6 Théorème de Lévy

Propriété. ***(Convergence en loi $\implies$ CV des fonctions caractéristiques)***

Si $X_n \underset{n\to+\infty}{\overset{\text{loi}}{\longrightarrow}} X$, alors $X_{\phi(X_n)} \underset{n\to+\infty}{\overset{\text{CVS}}{\longrightarrow}} \phi(X)$ sur $\mathbb{R}$.

▷ L'exponentielle imaginaire est continue bornée, donc ses parties réelle et imaginaire le sont. Il suffit d'appliquer la convergence étroite à celles-ci pour conclure. ■

La réciproque est vraie, on le sait grâce à un théorème dû à Paul Lévy. Il donne même une propriété bien plus forte, modulo une condition supplémentaire (mais très raisonnable, car les fonctions caractéristiques sont uniformément continues !) de continuité en zéro. Deux conditions très faibles impliquent une convergence en loi, d'après le théorème de Lévy.

Théorème. ***(Théorème de Lévy)***

Soit $(X_n)_{n\geqslant 1}$ une suite de v. a. r. Si ϕ_{X_n} converge simplement vers ϕ sur $\mathbb{R}$, et ϕ est continue en zéro, alors il existe une variable aléatoire X de fonction caractéristique ϕ telle que $X_n \underset{n\to+\infty}{\overset{\text{loi}}{\longrightarrow}} X$.

Remarques.

1. Il est fréquent dans les exercices que la fonction ϕ que l'on obtient soit une fonction caractéristique connue.
2. Le résultat reste vrai en dimension d.
3. Le théorème de Lévy est une première étape dans la démonstration du théorème centrale limite ; on l'appliquera à des variables gaussiennes.

Pour montrer le théorème de Lévy, on aura besoin du lemme suivant, qui utilise la notion de tension.

Lemme

Sous les hypothèses précédentes, la suite $(X_n)_{n\geqslant 1}$ est tendue.

▷ La preuve est très astucieuse. On remarque que $\frac{1}{u}\int_{-u}^{u}(1-\phi_{X_n}(t))\mathrm{d}t = \mathbb{E}[\frac{1}{u}\int_{-u}^{u}(1-e^{itX_n})\mathrm{d}t]$ par théorème de Fubini, qui vaut après calcul $2\mathbb{E}[1-\mathrm{sinc}(uX_n)]$. Or on sait que $|\mathrm{sinc}|\leqslant 1$ et que si $|x|\geqslant 2$, $|\mathrm{sinc}(x)|\leqslant \frac{1}{2}$. Ainsi :

$$\begin{aligned}\mathbb{P}(|X|\geqslant \frac{2}{u}) &= \mathbb{E}[\mathbb{1}_{\{|X_n|\geqslant \frac{2}{u}\}}]\\ &= 2\mathbb{E}[(1-\frac{1}{2})\mathbb{1}_{\{|X_n|\geqslant \frac{2}{u}\}}]\\ &\leqslant 2\mathbb{E}[1-\mathrm{sinc}(uX_n)]\\ &\leqslant \frac{1}{u}\int_{-u}^{u}(1-\phi_{X_n}(t))\mathrm{d}t \leqslant \frac{1}{u}\int_{-u}^{u}|1-\phi_{X_n}(t)|\,\mathrm{d}t\end{aligned}$$

Par théorème de convergence dominée, ce dernier terme converge quand $n\to\infty$ vers $\frac{1}{u}\int_{-u}^{u}|1-\phi(t)|\mathrm{d}t$. On a donc montré que $\limsup_{n\to+\infty}\mathbb{P}(|X|\geqslant\frac{2}{u})\leqslant\int_{-u}^{u}|1-\phi_{X_n}(t)|\,\mathrm{d}t$, donc par continuité de ϕ en zéro et le fait que $\phi(0)=1$ en observant la limite simple, on a $\limsup_{u\downarrow 0}\limsup_{n\to+\infty}\mathbb{P}(|X|\geqslant\frac{2}{u})=0$. On déduit de cette convergence que

$$\forall\varepsilon>0\ \exists M>0\quad \limsup_{n\to+\infty}\mathbb{P}(|X_n|>M)<\varepsilon$$

donc

$$\exists n_0\geqslant 1\ \forall n\geqslant n_0\quad \mathbb{P}(|X_n|>M)<\varepsilon.$$

Quitte à augmenter M, on a $\forall n\geqslant 1\ \mathbb{P}(|X_n|>M)<\varepsilon$ puisque la tension est une notion asymptotique. ■

On peut reprendre la preuve du théorème.

Preuve.

▷ Par le lemme (X_n) est tendue. On peut donc extraire une sous-suite (X_{n_k}) qui converge en loi. Soit X sa limite. Alors $\phi_{X_{n_k}}$ converge simplement vers ϕ_X. Or par hypothèse ϕ_{X_n} converge simplement vers ϕ donc $\phi=\phi_X$. Donc ϕ est la fonction caractéristique d'une variable. On a donc montré que de toute sous-suite de (X_n) on peut extraire une sous-sous-suite qui converge en loi vers X. Ceci à conclure que (X_n) converge en loi vers X. En effet, supposons le contraire. Il existerait une fonction $h:\mathbb{R}\longrightarrow\mathbb{R}$, $\varepsilon>0$ et (X_{n_k}) tel que pour tout $k\geqslant 1$, $|\mathbb{E}[h(X_n)]-\mathbb{E}[h(X)]|>\varepsilon$. Par la propriété venant d'être prouvée, il existerait $(X_{n_{k_j}})$ une sous-sous-suite convergeant en loi vers X. En particulier, $\mathbb{E}[h(X_{n_{k_j}})]\longrightarrow\mathbb{E}[h(X)]$, contradiction. ■

1.10.5.7 Théorème centrale limite

On part du point de départ suivant : la loi forte des grands nombres donne que pour des $(X_n)_{n\geqslant 1}$ indépendantes, de même loi, et intégrables, $\frac{1}{n}\sum_1^n X_k$ converge presque sûrement vers

$\mathbb{E}[X_1]$. Le problème, c'est que l'on n'a pas de mesure de la marge d'erreur et seulement une propriété asymptotique.

Faisons un calcul simple dans le cas où les X_n sont dans L^2. On a $\mathbb{E}[(\sum_{k=1}^{n} X_k - n\mathbb{E}[X_1])^2] = \mathbb{E}[(\sum_{k=1}^{n}(X_k - \mathbb{E}[X_k]))^2] = \sum_{k=1}^{n} \mathrm{Var}(X_k) = n\mathrm{Var}(X_1)$ donc $\mathbb{E}\left[\left(\frac{\sum_{k=1}^{n} X_k - n\mathbb{E}[X_1]}{\sqrt{n}}\right)^2\right] \underset{n\to+\infty}{\longrightarrow} \mathrm{Var}(X_1)$.

Ainsi,

$$\frac{1}{n}\sum_{k=1}^{n} X_k \approx \mathbb{E}[X_1] + o(\frac{1}{\sqrt{n}}).$$

Théorème. ***(Théorème centrale limite)***

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles définies L^2, indépendantes et de même loi, forcément définies sur un même espace de probabilité. On note $m = \mathbb{E}[X_1]$ et $\sigma^2 = \mathrm{Var}(X_1)$. Alors

$$\frac{X_1 + \ldots + X_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\text{loi}} Y$$

où $Y \sim \mathcal{N}(0{,}1)$.

Remarque. On peut aussi énoncer le résultat de façon non normalisée en écrivant que

$$\frac{X_1 + \ldots + X_n - nm}{\sqrt{n}} \xrightarrow[n\to+\infty]{\text{loi}} Z$$

où $Z \sim \mathcal{N}(0, \sigma^2)$.

Attention ! Si on ne centre pas les variables, ça explose et ça n'a pas de sens. Si on veut centrer, il faudrait faire apparaître un biais. Ce n'est pas très naturel mais on peut l'énoncer ; c'est laissé en exercice.

On montre le théorème centrale-limite.

Preuve.

$\triangleright$ On va utiliser le théorème de Lévy. Con calcule donc la fonction caractéristique de $\frac{X_1+\ldots X_n - nm}{\sqrt{n}\sigma}$. Quitte à poser $\tilde{X}_k = X_k - m$, on peut supposer que les X_k sont centrées. On manipule donc $\frac{X_1+\ldots X_n}{\sqrt{n}\sigma}$ avec $E[X_k] = 0$ pour tout $k \geqslant 1$. On calcule pour tout $t \in \mathbb{R}$, grâce à l'indépendance et l'identique distribution :

$$\begin{aligned}
\mathbb{E}[e^{it\frac{X_1+\ldots X_n}{\sqrt{n}\sigma}}] &= \prod_{k=1}^{n} \mathbb{E}[e^{it\frac{X_k}{\sigma\sqrt{n}}}] \\
&= \mathbb{E}[e^{it\frac{X_1}{\sigma\sqrt{n}}}]^n \\
&= \phi_{X_1}(\frac{t}{\sigma\sqrt{n}})^n \\
&= (1+\frac{it}{\sigma\sqrt{n}}\mathbb{E}X_1 - \frac{t^2}{2\sigma^2 n}\mathbb{E}X_1^2 + o\left(\frac{1}{n}\right))^n \\
&= (1-\frac{t^2}{2\sigma^2 n}\mathrm{Var}(X_1) + o\left(\frac{1}{n}\right))^n \\
&= \exp(n\ln(1-\frac{t^2}{2\sigma^2 n}\mathrm{Var}(X_1) + o\left(\frac{1}{n}\right))) \\
&= \exp(-\frac{t^2}{2\sigma^2 n}\mathrm{Var}(X_1) + o(1)) \\
&= \exp(-\frac{t^2}{2} + o(1)) \underset{n\to+\infty}{\longrightarrow} \exp(-\frac{t^2}{2})
\end{aligned}$$

Or $t \longmapsto e^{-t^2/2}$ est la fonction caractéristique d'une v. a. $\mathcal{N}(0,1)$, donc par théorème de Lévy, $\frac{X_1+\ldots+X_n}{\sigma\sqrt{n}} \underset{n\to+\infty}{\overset{\text{loi}}{\longrightarrow}} \mathcal{N}(0,1)$. ■

1.11 Lois classiques

1.11.1 Lois finies

1.11.2 Lois discrètes

1.11.3 Lois à densité

Chapitre 2

Processus aléatoires

Résumé

Dans un premier temps, on étudie les processus à temps discret : martingales, chaînes de Markov. Dans un second temps, on étudie les processus modélisés à paramètre de temps continu : principalement, les mouvements browniens.

2.1 Processus à temps discret

2.1.1 Généralités sur les processus à temps discret

2.1.1.1 Définition

Définition. ***(Processus à temps discret)***

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On appelle *processus à temps discret*, ou *processus* dans le cadre de cette partie du cours, toute suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires, éventuellement de la forme $(X_n)_{n\geqslant n_0}$ pour un $n_0 \in \mathbb{N}$. De même que dans le cas des suites, on se ramène toujours au premier cas par simple changement de variables.

Au propos d'un processus, on peut donc définir tous les modes de convergence des variables aléatoires : convergences en moyenne, convergence presque certaine, convergence en mesure, et convergence en loi.

Heuristique

Un processus, en toute généralité, a pour objectif de modéliser un phénomène qui évolue selon le temps. Ce rôle est joué, dans le cas discret, par le paramètre n, de même que, dans le cas continu (plus proche de l'intuition réelle, car le temps nous paraît continu), c'est le paramètre $t \in \mathbb{R}$ qui signifie le temps.
Rappelons que le terme de *variable aléatoire*, quoique fustigé, représente très bien le sens donné à cet objet : bien que ce soit une fonction, qui dépende en fait des aléas $\omega \in \Omega$, c'est bien sa valeur, certes conditionnée par ω, qui nous intéresse[a].

Avec les processus, on a une infinité de phénomènes successifs à étudier. Notons qu'alors, pour une issue donnée, on a une suite de valeurs (ou fonction de la variable réelle dans le cas continu) donnée directement : $(X_n(\omega))_{n\in\mathbb{N}}$, voir le principe suivant. Par exemple, de façon plus parlante :

- tout jeu (pile ou face, dé, roulotte, etc.), joué un nombre arbitrairement grand de fois, dont on comptabilise le gain, cumulé ou non, est pris en compte par une suite de variables aléatoires souvent notée X_n ou S_n ;
- le cours de la bourse, en économie, est modélisé par un processus aléatoire ;
- la propagation d'une épidémie,
- l'évolution d'une opinion mesurée semaine après semaine par un sondage, est un processus aléatoire à temps discret.

[a] Un peu à la manière du chat de Schrödinger : avant d'ouvrir la boîte (avant d'évaluer en l'une des issues), on ne sait pas si le chat est mort ou vivant, mais on peut quand même étudier pour lui-même le phénomène X : « vie du chat »

Principe. ***(Réalisation d'un processus)***

À une issue de l'univers et un processus correspond une unique famille de réalisations déterminées (non plus aléatoires).

Un processus n'est autre qu'une suite, donc une famille de variables aléatoires. De même que dans le cas d'un couple, il faut distinguer loi jointe et loi produit.

Il est nécessaire de distinguer la loi d'un processus $(X_n)_{n\in\mathbb{N}}$, jointe, de la suite des lois des X_n, marginales. Pour s'en convaincre malgré l'infinité de la famille, prendre deux variables bien choisies pour $n = 0$ et $n = 1$ et considérer la suite constante, déterministe à partir du rang $n = 2$.
De même que dans le cas d'un couple, la première loi englobe l'information des corrélations entre les variables, mais pas le deuxième. Il faut donc s'intéresser à la loi du processus dans la théorie des processus.

2.1.1.2 Filtrations

Cette notion se définit sur un espace probabilisé en général, et même sur un espace probabilisable.

Définition. ***(Filtration, espace filtré)***

Soit $(\Omega, \mathcal{F})$ un espace probabilisable. On appelle *filtration* une suite croissante $(\mathcal{F}_n)_{n\in\mathbb{N}}$ de sous-tribus de $\mathcal{F}$, c'est-à-dire $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{F}$.
Si $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{F})$, on dit que le quadruplet $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}$ est un *espace*

de probabilité filtré.

Définition. ***(Processus adapté à une filtration)***

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}$ un espace probabilisé filtré et $(X_n)_{n\in\mathbb{N}}$ un processus sur $(\Omega, \mathcal{F})$, on dit que $(X_n)_{n\in\mathbb{N}}$ est *adapté* à $(\mathcal{F}_n)_{n\in\mathbb{N}}$ si pour tout $n \in \mathbb{N}$, X_n est $\mathcal{F}_n$-mesurable.

2.1.2 Quelques exemples de base

2.1.2.1 Marches aléatoires

2.1.2.2 Processus de Galton-Watson

2.1.2.3 Files d'attente

2.1.3 Notion de temps d'arrêt

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé muni d'une filtration $(\mathcal{F}_n)_n$.

Définition. ***(Temps d'arrêt)***

Une variable aléatoire $T : \Omega \longrightarrow \overline{\mathbb{N}}$ est un temps d'arrêt si $\{T \leqslant n\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}$, ou, ce qui est équivalent[a], si $\{T = n\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}$.

[a] Dans un sens, écrire une intersection de deux ensembles, dans l'autre, une union finie.

Remarques.

1. La valeur $+\infty$ est **autorisée**.
2. Typiquement, T est le rang à partir duquel on décide d'arrêter le jeu, mais la décision d'arrêter au temps n $\{T = n\}$ (*i. e.* juste après le même jeu) ne doit dépendre de l'histoire du processus que jusqu'à l'instant n : le joueur n'est pas devin (ni informé par un autre biais : tiers, météo...) !

Exemples

1. Soit $(A_n)_{n\in\mathbb{N}}$ un processus adapté à la filtration $(\mathcal{F}_n)$. Soit B un événement et $T = \inf\{n, A_n \in B\}$. Alors T est un temps d'arrêt, tel que $T = +\infty$ si (A_n) n'arrive jamais dans B.

   En effet, $\{T \leqslant n\} = \bigcup_{k=0}^{n} \{A_k \in B\} \in \mathcal{F}_n$, puisque $\{A_k \in B\} \in \mathcal{F}_k$, car (A_n) est adapté.
2. $T = \sup\{n \leqslant 4, A_n \in B\}$ est le dernier instant avant $n = 4$ où on atteint B. Ainsi $\{T \leqslant 3\} = \{A_4 \notin B\} \in \mathcal{F}_4$, pas dans $\mathcal{F}_{n-1}$ a priori.

   Un premier instant a de fortes chances d'être un temps d'arrêt, un dernier instant, non.

3. Un « temps d'arrêt » déterministe est un temps d'arrêt, mais pas forcément un dernier instant ! Dans l'exemple précédent, le premier instant à $n = 8$ n'est pas un temps d'arrêt.

2.1.4 Martingales

2.1.4.1 Définition et premières propriétés

Méthode. *(Montrer qu'un truc est une martingale)*

Les candidats pour être des martingales modélisent des phénomènes dépendant toujours du phénomène de rang précédent. Soit $(M_n)_n$ un processus.
On exprime $M_{n+1} - M_n$. Souvent, cette variable prend une certaine valeur selon les cas (souvent deux) avec une certaine probabilité. On calcule l'espérance naïve : $p_1V_1+(1-p_1)V_2$ et l'on montre algébriquement qu'elle est nulle. Ainsi $\mathbb{E}[M_{n+1} - M_n|\mathcal{F}_n] = 0$ avec $(\mathcal{F}_n)$ la filtration canonique.
Par suite, $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = \mathbb{E}[M_n|\mathcal{F}_n] = M_n$, donc (M_n) est une martingale.

2.1.4.2 Exemples de martingales

2.1.4.2.1 Urne de Polya

Cela marche encore avec m boules rajoutées à chaque tirage.

2.1.4.2.2 Processus de branchement

Soient $X_r^{(m)}$, $m,r \in \mathbb{N}$ iid de même loi que X et $\mathbb{E}X = \mu$. On pose :

$$\begin{cases} Z_0 = 1 \\ Z_{n+1} = X_1^{(n+1)} + ... + X_{Z_n}^{(n+1)}. \end{cases}$$

Dans ce cas, si $X_k^{(i)}$ modélise le nombre de descendants à la génération i du k-ième individu, Z_k représente la taille de la population à la génération k.

On a $\mathbb{E}[Z_{n+1}|Z_n) = \mu$. Dans ce cas, $M_n = \frac{Z_n}{\mu^n}$ est une martingale telle que $\mathbb{E}[M_{n+1}] = 1$.

On pose $f(\theta) = \mathbb{E}\theta^X$. Alors $\mathbb{E}[\theta^{Z_{n+1}}|Z_n = k] = f(\theta)^k$ et si $f_n(\theta) = \mathbb{E}[\theta^{Z_n}]$, $f_{n+1}(\theta) = f_n(f(\theta)) = f \circ f ... \circ f(\theta)$ $n+1$ fois. Alors $\Pi = \lim\limits_{n \to +\infty} \uparrow \mathbb{P}(Z_n = 0)$ est la probabilité d'extinction du processus (Z_n).

Π est le plus petit point fixe de $x = f(x)$ dans $[0,1]$. On peut montrer que $\Pi = 1$ si $\mu \leqslant 1$. Dans ce cas, avec probabilité 1, la probabilité s'éteint. On montrera plus tard (Théorèmes de convergence) que $\underset{M_n \to n}{\longrightarrow} +\infty M_\infty$ presque sûrement. $\mathbb{E}[M_n] = 1$, or $M_\infty = 0$ presque sûrement si $\mu < 1$.

2.1.4.3 Transformations de martingales

On reprend les notations précédentes.

Définition. ***(Processus prévisible)***

Un processus $C = (C_n)_{n\in\mathbb{N}}$ est dit prévisible si C_n est $\mathcal{F}_{n-1}$-mesurable pour tout $n \in \mathbb{N}$. Dans le cadre d'un jeu, C_n est la mise que l'on décide en fonction des parties réalisées auparavant. On gagne alors $C_n(X_n - X_{n-1})$ et $X_0 = 0$.

Exercice 1

Quels sont les processus prévisibles qui sont des martingales ?

Un joueur digne de ce nom varie sa stratégie : ici, sa stratégie est C.

Définition. ***(Gain d'un processus prévisible)***

Le gain total après n parties est

$$Y_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1}) = (C \cdot X)_n.$$

Définition. ***(Transformée d'une martingale)***

On dit que $(C \cdot X)$ est une *transformée* de X par le processus prévisible C.

Théorème

Soit un espace de probabilité avec les notations usuelles. Soit C un processus prévisible positif et bornée par k. Soit X une surmartingale (resp. une martingale). Alors $(C \cdot X)$ est une sur-martingale (resp. une martingale).

▷ En effet, puisque $C_n \geqslant 0$ et est F_{n-1}-mesurable, $\mathbb{E}[Y_n - Y_{n-1}|\mathcal{F}_{n-1}] = \mathbb{E}(C_n(X_n - X_{n-1})|\mathcal{F}_{n-1}) = C_n\mathbb{E}(X_n - X_{n-1}|\mathcal{F}_{n-1})$. ■

Remarque. Le théorème est encore vrai si $C \geqslant 0$ et $C,X \in L^2$.

VOC Un jeu est *juste* s'il vérifie $\mathbb{E}[X_{n+1} - X_n] = 0$. Un jeu est *truqué* s'il vérifie $\mathbb{E}[X_{n+1} - X_n] \leqslant 0$.

Propriété. ***(Composée convexe d'une martingale)***

Soit (M_n) une martingale. Soit $f : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction convexe telle que $\mathbb{E}|f(M_n)| < \infty$. Alors $Z_n = f(M_n)$ est une sous-martingale.

▷ Application de l'inégalité de Jensen. ■

2.1.4.4 Décomposition de Doob-Meyer

2.1.5 Théorèmes d'arrêt

2.1.5.1 Martingales et surmartingales arrêtées

Soient X une surmartingale et T un temps d'arrêt. On suppose que l'on mise 1 € à chaque jeu.

Définition. ***(Mise intégrant le processus d'arrêt)***

On définit le processus prévisible

$$C_n^{(T)} = \mathbb{1}_{\{T \geqslant n\}}$$

qui est la *mise intégrant le processus d'arrêt* (on arrête le jeu au temps T).

Avec les notations précédentes :

Définition. ***(Processus de gain)***

On définit le *processus de gain* $(C^{(T)} \cdot X) = X_{T \wedge n} - X_0$.

Définition. ***(Processus arrêté)***

On définit X^T le *processus arrêté au temps T* par

$$X_n^T = X_{T(\omega) \wedge n}(\omega).$$

On défini également le processus de gain du processus arrêté au temps T par $(C^{(T)} \cdot X) = X^T - X_0$.

Propriété. ***(Vérification des hypothèses de transformation par les processus d'arrêt)***

$C^{(T)}$ est prévisible, positif et borné par 1.

Théorème

Soit X une surmartingale et T un temps d'arrêt. Le processus arrêté X^T est une surmartingale et $\mathbb{E}X_{T\wedge n} \leqslant \mathbb{E}X_0$ pour tout n. Soit X une martingale et T un temps d'arrêt. Le processus arrêté X^T est une martingale $\mathbb{E}X_{T\wedge n} = \mathbb{E}X_0$ pour tout n.

▷ Le texte de la preuve. ■

Remarque. Soit X une marche aléatoire sur $\mathbb{Z}$. Alors $X_{n+1} - X_n$ iid. $\mathbb{E}[X_{n+1} - X_n) = 0$ de même loi que X, $\mathbb{P}(X = \pm 1) = \frac{1}{2}$ (attention à cette écriture...), $X_0 = 0$. T est le premier instant où $X_n = 1$. Alors $T < +\infty$ presque sûrement, sinon la marche aléatoire serait négative presque toujours, ce qui est aberrant. Alors T est un temps d'arrêt. On a $\mathbb{E}X_{T\wedge n} = \mathbb{E}X_0 = 0$ mais $\mathbb{E}[X_T] = 1$.

Dans la suite, on va donner des conditions pour que $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

2.1.5.2 Théorèmes d'arrêt de Doob

Théorème

Soient T un temps d'arrêt et $(X_n)_{n\geqslant 0}$ une surmartingale ou une martingale, alors X_T est intégrable et $\mathbb{E}X_T \leqslant \mathbb{E}X_0$ à une des conditions suivantes est vérifiée :

(i) T est borné,

(ii) X est bornée et $T < \infty$ presque sûrement,

(iii) $\mathbb{E}T < \infty$ et les accroissements consécutifs de X sont bornés.

▷ On sait que $\mathbb{E}X_{T\wedge n} \leqslant \mathbb{E}X_0$. Pour le premier point, on prend $n = [k] + 1$ et $Twedgen = T$. Pour le second point, on applique le théorème de convergence dominée et $X_{T\wedge n}$ converge presque sûrement vers X_T, $|X_{T\wedge n}| \leqslant k$ donc $\mathbb{E}[X_T] \leqslant \mathbb{E}X_0$. Pour le dernier point, on applique le TCD aux accroissements. On a $|X_{T\wedge n} - X_0| = \left|\sum_{k=1}^{T\wedge n} X_k - X_{k-1}\right| \leqslant kT\wedge n \leqslant kT$ et $\mathbb{E}[T] < +\infty$. Alors $\mathbb{E}[T] < +\infty$ et $T < +\infty$ presque sûrement, donc $X_{T\wedge n}$ tend presque sûrement vers X_T, d'où le résultat par le théorème de convergence dominée. ■

Corollaire

Soit M une martingale dont les accroissements sont bornés par k. Soit C un processus prévisible borné par k_2. Soit T un temps d'arrêt d'espérance finie. Alors $\mathbb{E}[C \cdot M)_T = 0$. On a une formule analogue pour une surmatingale positive et $T < \infty$ presque sûrement.

▷ $C \cdot M = X$ est une martingale et (iii) donne le résultat. $C \cdot M = X$ est une surmartingale positive et $X_{T\wedge k} \to X_T$. $\mathbb{E}[X_{T\wedge k} \leqslant \mathbb{E}[X_0]$ par Fatou. ■

Remarque. (Conditions pour avoir $\mathbb{E}T < \infty$.) Soit T un t.a. tel qu'il existe un entier positif N et $\varepsilon > 0$ tel que pour tout n, $\mathbb{P}(T \leqslant N + n|\mathcal{F}_n) > \varepsilon$ presque sûrement. Alors $\mathbb{E}T < +\infty$.

▷ $\mathbb{P}(T > kN) = \mathbb{P}(T > kN | T > (k-1)N)\mathbb{P}(T > (k-1)N) \leqslant (1-\varepsilon)\mathbb{P}(T > (k-1)N) \leqslant (1-\varepsilon)^k$, d'où $\mathbb{E}T/N < +\infty$. ■

Les théorèmes d'arrêt s'utilisent pour décrire un processus, notamment, ses trajectoires.

Exemple

Si X_i, $i \geqslant 1$ est tel que $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$, $S_0 = 0$ et $S_n = \sum_{i=1}^{n} X_i$ et $T = \inf\{n \geqslant 0, S_n = 1\}$.
Alors S_n est une martingale adaptée à $\mathcal{F}_n = \sigma(X,...,X_n)$ et T un ta.
On pose $\mathbb{E}e^{\theta X} = \frac{1}{2}(e^{\theta} + e^{-\theta}) = \cosh(\theta)$. Alors $M_n^\theta = \left(\frac{1}{\cosh(\theta)}\right)^n e^{\theta S_n}$ est une martingale.
$\mathbb{E}[M_{T\wedge n}^\theta] = 1$ pour tout n ; on prend $\theta > 0$. Alors $|\exp(\theta S_{T\wedge n})| \leqslant e^{\theta}$, car $S_{T\wedge n} \leqslant 1$. De plus, $T < \infty$ presque sûrement. Puisque $M_{T\wedge n}^\theta \to M_T^\theta$, en posant $M_T^0 = 0$ si $T = \infty$, ce qui arrive avec une probabilité nulle, on en déduit que $\mathbb{E}[M_T^\theta] = 1$. Donc $\mathbb{E}\left(\frac{1}{\cosh(\theta)}\right)^T = e^{-\theta}$.
Pour $\alpha = \frac{1}{\cosh(\theta)}$, on obtient $\mathbb{E}(\alpha^T) = e^{-\theta} = \alpha^{-1}(1 - \sqrt{1-\alpha^2})$ par un développement en série entière et par là la loi de T.

Exemple

Si X_i sont iid et $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$, alors $T_{a,b} = \inf\{n, S_n \notin (a,b)\}$. C'est le premier instant où la marche sort de $(-a,b)$. On sait que $T_{a,b} < +\infty$ presque sûrement par le théorème central limite. Ainsi $\mathbb{E}[S_{T_{-a,b}} \wedge n] = 0$.
Puisque $\left|S_{T_{a,b}\wedge n}\right| \leqslant b + a$ puisque $T_{-a,b} \wedge n \to T_{-a,b}$ par TCD et $\mathbb{E}[S_{T_{-a,b}\wedge n} \to \mathbb{E}S_T{-a,b}$.
On a $\mathbb{E}S_{T_{a,b}} = \mathbb{E}S_0$. En effet, $-a\mathbb{P}(S_{T_{a,b}} = -a) + b\mathbb{P}(S_{T_{a,b}} = b) = 0$ et $\mathbb{P}(S_{T_{-a,b}} = -a) + \mathbb{P}(S_{T_{-a,b}} = b) = 1$, d'où $\mathbb{P}(S_{T_{-a,b}} = b) = \frac{a}{a+b}$ et $T_{-a,b} < +\infty$ ps.

2.1.6 Théorèmes de convergence

Soit $X_n - X_{n-1}$ le gain au jeu n si on mise un euro. La trajectoire de $(X_n)_{n\geqslant 0}$ est le graphe discret de X_n dans le plan (dessin ?). Soient maintenant deux réels a,b avec $a < b$. On va définir une stratégie prévisible C. On attend que X passe sous le niveau a. On mise un euro, jusqu'à ce qu'on dépasse b, et alors on attend de repasser sous le niveau a. Dans ce cas $C_1 = \mathbb{1}_{X_0 < a}$. $C_n = \mathbb{1}_{C_{n-1}=1}\mathbb{1}_{X_{n-1}\leqslant b} + \mathbb{1}_{C_{n-1}=0}\mathbb{1}_{X_{n-1}<a}$. C_n est bien prévisible, positif et borné.

2.1.6.1 Crochet

On utilise un théorème très utile, que je n'ai pas eu le temps de noter.

Définition-propriété. ***(Crochet d'une martingale)***

Soit M une martingale nulle en 0 et dans L^2. Alors M^2 est une sous martingale par Jensen conditionnel.
Donc on sait qu'elle admet une décomposition de Doob $M^2 = N + A$ avec N une martingale

nulle en 0 et A un processus prévisible nul en 0. On pose $A_\infty = \lim_{n\to+\infty} \uparrow A_n$ presque sûrement.
On appelle A le *crochet* de M.

Immédiatement :

Remarques.

1. On sait que $\mathbb{E}A_n = \mathbb{E}M_n^2$ pour tout n.
2. M est bornée dans $L^2 \iff \sup_n \mathbb{E}M_n^2 < +\infty \iff \mathbb{E}A_\infty < \infty$ par le théorème de convergence monotone.
3. $A_n - A_{n-1} = \mathbb{E}((M_n - M_{n-1})^2|\mathcal{F}_{n-1}) = \mathbb{E}(M_n^2 - M_{n-1}^2|\mathcal{F}_{n-1})$.

Définition-propriété. ***(Crochet de deux martingales)***

Soient M,N deux martingales bornées dans L^2. On définit :

$$\langle M, N\rangle = \frac{1}{2}(\langle M+N\rangle - \langle M\rangle - \langle N\rangle .$$

Alors $MN - \langle M,N\rangle$ est une martingale.

▷ Utiliser l'identité : $xy = \frac{1}{2}[(x+y)^2 - x^2 - y^2]$. ■

Théorème

Soit M une martingale L^2.

(i) $\lim_{n\to+\infty} M_n$ existe pour presque tout ω pour lequel $A_\infty(\omega) < \infty$.

(ii) Si les incréments de M_n sont uniformément bornés par K, soit $|M_n - M_{n-1}| \leqslant k \quad \forall n\ \forall\omega$. Alors $A_\infty(\omega) < \infty$ pour tout ω tel que $\lim_{n\to+\infty} M_n(\omega)$ existe.

▷ Vue A est prévisible, on définit un temps d'arrêt par $S_k = \inf\{n \in \mathbb{N}, A_{n+1} \geqslant k\}$. Dans ce cas A^{S_k} est également prévisible. De plus,

$$A_{S_k\wedge n} \in B = \bigcup_{r=0}^{n-1} \{S_k = r \wedge A_r \in B\} \cup (\{S_k \geqslant n-1\} \cap A_n \in B).$$

Ainsi $(M^{S_k})^2 - A_{S_k} = (M^2 - A)^{S_k}$ est une martingale telle que $\left\langle M^{S_k}\right\rangle = A^{S_k}$. Le processus A^{S_k} est borné par k et donc $\mathbb{E}A_\infty < \infty$. Par un théorème connu, M^{S_k} est donc bornée dans L^2 et donc $\lim_{n\to+\infty} M_{n\wedge S_k}$ existe presque sûrement ; en effet, une martingale L^2 est presque sûrement finie. Or $\{A_\infty < \infty\} = \bigcup_k \{S_k = \infty\}$, donc on peut faire tendre $n \to \infty$.

Pour le second point, si pour un $c > 0$ quand on pose $T_c = \inf\{n, |M_n| > c\}$, on a $\mathbb{P}(T_c = \infty, A_\infty = \infty) \quad (\star)$, alors $\mathbb{E}(M_{T\wedge n}^2 - A_{T\wedge n}) = 0$ et $|M^{T_c}| \leqslant k + c$ par borne sur les incréments, et l'on en déduit $\mathbb{E}A_{T_c\wedge n} \leqslant (k+c)^2 \quad \forall n$. Par convergence dominée, on en déduit $\mathbb{E}A_\infty \mathbb{1}_{T_c=\infty} > +\infty$. Avec $(\star)$, c'est absurde. D'où la preuve. ■

2.1.7 Uniforme intégrabilité

On cherche des conditions suffisantes aux théorèmes de convergence.

2.1.7.1 Définitions

On commence par un lemme de théorie de la mesure.

Lemme

Soit X une variable aléatoire $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Alors étant donné $\varepsilon > 0$, il existe $\delta > 0$ tel que si $F \in \mathcal{F}$ avec $\mathbb{P}(F) < \delta$, $\mathbb{E}|X|\mathbb{1}_F < \varepsilon$.

$\triangleright$ Sinon, il existe $\varepsilon > 0$ tel qu'il existe F_n tel que $\mathbb{P}(F_n) < 2^{-n}$ et $\mathbb{E}|X|\mathbb{1}_{F_n} > \varepsilon$. On a $F = \overline{\lim} F_n = \mathbb{P}(F) = 0$ par Borel Cantelli, d'où $\mathbb{E}|X|\mathbb{1}_F \geqslant \varepsilon$, absurde. ■

Corollaire

Si $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $\varepsilon > 0$, il existe k tel que $\mathbb{E}|X|\mathbb{1}_{|X|>k} < \varepsilon$.

Définition. ***(Uniforme intégrabilité)***

Une collection $\mathcal{C}$ de variables aléatoires est dite uniformément intégrable si étant donné $\varepsilon > 0$, il existe $k > 0$ tel que $\mathbb{E}|X|\mathbb{1}_{|X|\geqslant k} < \varepsilon \qquad \forall X \in \mathcal{C}$.

Remarques.

1. Une famille $\mathcal{C}$ uniformément intégrable est bornée dans L^1 : il suffit de prendre $\varepsilon = 1$ dans l'exemple précédent, et le k correspondant vérifie $\mathbb{E}|X| = \mathbb{E}|X|(\mathbb{1}_{|X|\leqslant k} + \mathbb{1}_{|X|>k}) \leqslant k + 1$.
2. Une famille *finie* de variables aléatoires dans L^1 est automatiquement uniformément intégrable.
3. Une famille $\mathcal{C}$ bornée dans L^1 **n'est pas forcément** uniformément intégrable. Par exemple, sur $([0,1], \mathcal{B}([0,1]), \lambda)$, $X_n = n\mathbb{1}_{[0,1/n]} \geqslant 0$ et $\mathbb{E}X_n = 1$. Si $k > 0$, $\mathbb{E}|X_n|\mathbb{1}_{|X_n|>k} = n\mathbb{P}(X_n)$ si $n \geqslant k$. Alors $(X_n)_{n\geqslant 1}$ n'est pas uniformément intégrable, alors que X_n tend vers 0 en probabilité et $\mathbb{E}X_n = 1$.
4. Toute famille uniformément intégrable est bornée dans $\mathbb{R}$.

2.1.7.2 Conditions suffisantes pour l'uniforme intégrabilité

Propriété

Si $\mathcal{C}$ est une collection de v. a. bornée dans $L^p, p > 1$, c'est-à-dire s'il existe M tel que $\mathbb{E}|X|^p \leqslant M$ pour tout $X \in \mathcal{C}$, alors $\mathcal{C}$ est uniformément intégrable.

$\triangleright$ En effet, soit $v \geqslant k > 0$, alors $v \leqslant v^p k^{1-p}$ donc $\frac{v}{k} \leqslant (\frac{v}{k})^p$. Ainsi pour $|X| = v$, $\mathbb{E}|X|\mathbb{1}_{|X|>k} \leqslant \mathbb{E}|X|^p k^{1-p} \mathbb{1}_{|X|>k} \leqslant k^{1-p}M$, d'où le résultat. ■

Propriété

Soit $\mathcal{C}$ une collection de v. a. telle qu'il existe $V \geqslant 0$, $Y \in L^1$, avec $|X| \leqslant Y$ pour tout $X \in \mathcal{C}$. Alors $\mathcal{C}$ est uniformément intégrable.

$\triangleright$ En effet, $\mathbb{E}(|X|\mathbb{1}_{|X|>k} \leqslant \mathbb{E}Y\mathbb{1}_{Y>k} < \varepsilon$ pour k assez grand. ■

Théorème

Soit X une variable aléatoire dans $L^1(\Omega, \mathcal{F}, \mathbb{P})$. La collection $\{\mathbb{E}(X|\mathcal{G}), \mathcal{G}$ sous-tribu de $\mathcal{F}\}$, est uniformément intégrable.

$\triangleright$ En exercice. ■

Théorème

Soit $(X_n)_n$ une suite de variables aléatoires dans $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Soit $X \in L^1$. Alors X_n tend vers X dans L^1 si et seulement si :

(i) X_n tend vers X en probabilité,

(ii) $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable.

$\triangleright$ Montrons le sens réciproque. On définit pour $x > 0$, la fonction $\varphi_k : x \mapsto x$ si $|x| \leqslant k$, k si $x > k$ et $-k$ si $x < -k$ une fonction seuil. Alors $\mathbb{E}|\varphi_k(X_n) - X_n| \leqslant \mathbb{E}|X_n|\mathbb{1}_{|X_n|>k} < \varepsilon$ si k assez grand, par uniforme intégrabilité. Par $\mathbb{E}|\varphi_k(X) - X| < \varepsilon$ aussi. Or $|\varphi_k(x) - \varphi_k(y)| \leqslant |x - y|$. Donc $\varphi_k(X_n)$ tend en probabilité vers $\varphi_k(X)$. Par théorème de convergence dominée, $\mathbb{E}|\varphi_k(X_n) - \varphi_k(X)| < \varepsilon$ pour n assez grand. En sommant, on montre que $\mathbb{E}|X_n - X| < 3\varepsilon$ pour n assez grand. Donc X_n tend vers X dans L^1.

D'autre part, par l'inégalité de Markov, si X_n tend vers X dans L^1, alors X_n tend vers X en probabilités. On note $M = \sup_{n\geqslant 0} \mathbb{E}|X_n|$. On fixe un entier N assez grand. Alors $\mathbb{E}|X_n - X| < \varepsilon$ pour $n \geqslant N$. De plus $\{(X_n)_{n\in[\![1,N]\!]}, X\}$ est uniformément intégrable. Soit donc $\delta > 0$ tel que si $\mathbb{P}(F) < \delta$, $\mathbb{E}|X_n|\mathbb{1}_F < \varepsilon$ pour $n \geqslant N$. Ainsi $\mathbb{E}|X|\mathbb{1}_F < \varepsilon$. On prend K tel que $M/K < \delta$. Si $n \geqslant N$, $\mathbb{E}|X_n|\mathbb{1}_{|X_n|>k} \leqslant \mathbb{E}(|X| + |X - X_n|)\mathbb{1}_{|X_n|>k} \leqslant 2\varepsilon$. Alors $\mathbb{P}(|X_n| > k) \leqslant M/k < \delta$. ■

2.1.7.3 Martingales uniformément intégrables

Définition. ***(Martingale intégrable uniformément)***

Une martingale uniformément intégrable est une martingale qui est uniformément intégrable.

Corollaire

Une martingale uniformément intégrable est bornée dans L^1, donc on sait que $M_\infty = \lim\limits_{n\to+\infty} M_n$ existe presque sûrement. On a aussi par uniforme intégrabilité, $\mathbb{E}|M_n - M_\infty| \to 0$. On peut montrer dans ce cas $M_n = \mathbb{E}(M_\infty|\mathcal{F}_n)$ presque sûrement.

▷ Si $r \geqslant n$ et $F \in \mathcal{F}_n$, $M_r\mathbb{1}_F = \mathbb{E}M_n\mathbb{1}_F$ $(\star)$. Or $\mathbb{E}|M_r\mathbb{1}_F - M_\infty\mathbb{1}_F| \leqslant \mathbb{E}|M_r - M_\infty| \underset{r\to+\infty}{\longrightarrow} 0$, donc on peut faire tendre $r \to \infty$ dans $(\star)$. ■

Théorème. ***(Lévy)***

Soit $\zeta \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, et $M_n = \mathbb{E}(\zeta|\mathcal{F}_n)$ presque sûrement.
Alors $(M_n)_{n\geqslant 1}$ est une martingale uniformément intégrable et $M_n \underset{r\to+\infty}{\longrightarrow} \mathbb{E}(\zeta|\mathcal{F}_\infty) = \eta$ presque sûrement et dans L^1.

▷ On sait que (M_n) est une martingale uniformément intégrable, donc on sait qu'il existe M_∞ limite presque sûre et L^1 de M_n. Le seul point à montrer est que $M_\infty = \eta$ presque sûrement. Quitte à couper $\zeta = \zeta_+ + \zeta_-$, on suppose ζ positive. On considère les mesures Q_1,Q_2 sur $(\Omega, \mathcal{F}_\infty)$. Pour $F \in \mathcal{F}_\infty$, on pose $Q_1(F) = \mathbb{E}(\eta inf F)$ et $Q_2(F) = \mathbb{E}(M_\infty\mathbb{1}_F)$. Alors Q_1,$Q_2 \geqslant 0$ et $\zeta \in L^1$, donc Q_1,Q_2 sont des mesures. Or si $F \in F_n$, $Q_1(F) = \mathbb{E}(\zeta\mathbb{1}_F)$ par le propriété de tour et $Q_2(F) = \mathbb{E}(M_\infty\mathbb{1}_F) = \mathbb{E}(M_n\mathbb{1}_F)$ ce qui provient de la preuve précédente. Ainsi Q_1,Q_2 coïncident sur F_n pour tout n, donc sur F_∞. ■

Théorème. ***(Loi du $0-1$ de Kolmogorov avec les martingales)***

Soit (X_i) une suite de v. a. indépendantes et $\mathcal{T}_n = \sigma(X_n, X_{n+1}, \ldots)$. On pose $\mathcal{T} = \bigcap\limits_{n\geqslant 0} T_n$.
Alors pour tout $F \in \mathcal{T}$, $\mathbb{P}(F) = 0$ ou 1.

▷ Soit $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Soit $F \in \mathcal{T} = \mathcal{F}_\infty$. On pose $\eta = \mathbb{1}_F$, $\mathcal{T}$-mesurable. Alors $\eta = \mathbb{E}(\eta, \mathcal{F}_\infty) = \lim\limits_{n\to+\infty} \mathbb{E}(\eta|\mathcal{F}_n)$ presque sûrement. Or η est indépendante de $\mathcal{F}_n$ donc $\mathbb{E}[\eta|\mathcal{F}_n] = \mathbb{E}\eta = \mathbb{P}(F)$ presque sûrement. Donc η est une variable aléatoire qui prend les valeurs 0 ou 1. ■

On termine avec un théorème jamais utilisé.

Théorème. *(Un second théorème de Lévy)*

Soit $(\mathcal{G}_{-n})_n$ une suite de tribus avec $\mathcal{G}_{-\infty} \subseteq ... \subseteq \mathcal{G}_{-1}$. Soit $\gamma \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $M_{-n} = \mathbb{E}(\gamma|\mathcal{G}_{-n})$. Alors $M_{-\infty} = \lim\limits_{n\to+\infty} M_{-n}$ existe presque sûrement dans L^1 et $M_{-\infty} = \mathbb{E}(\gamma|\mathcal{G}_{-\infty})$ presque sûrement.

2.1.8 Inégalités de Doob

Théorème. *(Première inégalité de Doob)*

Soit $(Z_n)_{n\geqslant 0}$ une sous-martingale positive et $c > 0$. Alors

$$c\mathbb{P}(\sup_{0\leqslant k\leqslant n} Z_k \geqslant c) \leqslant \mathbb{E}(Z_n \mathbb{1}_{\sup\limits_{0\leqslant k\leqslant n} Z_k \geqslant c}) \leqslant \mathbb{E}Z_n.$$

Où $\mathbb{E}Z_n$ est la *valeur terminale.*

Remarque. La majoration est par la valeur terminale des $\mathbb{E}Z_k$, $k \leqslant n$.

$\triangleright$ Pour $F = \{\sup\limits_{0\leqslant k\leqslant n} Z_k \geqslant c\}$, $F = F_0 \cup F_1 \cup ... \cup F_n$ où les F_i sont deux à deux disjoints. Ainsi $F_k = \{Z_0 < c, Z_{k-1} < c, Z_k \geqslant c\} = \{T = k\}$ où $T = \inf\limits_{m\leqslant n} \{Z_m \geqslant c\}$. Ainsi $\mathbb{Z}_n \mathbb{1}_{F_k} \geqslant \mathbb{E}F_k \mathbb{1}_{F_k} \geqslant c\mathbb{P}(F_k)$. Et maintenant $\mathbb{P}(F) = \sum \mathbb{P}(F_k)$. ■

Application. *(Inégalités de Doob)*

1. Si (M_n) est une martingale, $c : \mathbb{R} \longrightarrow \mathbb{R}_+$ convexe telle que $c(M_n) \in L^1$ pour tout n, alors $C(M)$ est une sous-martingale positive. On peut en effet appliquer l'inégalité précédente à $c(M)$.
2. Si $(X_i)_{i\geqslant 1}$ est une suite de v. a. i. centrées, et $\mathrm{Var}(X_i) = \sigma_i^2$, on pose $S_n = \sum\limits_{i=1}^{n} X_i$. Alors $c^2\mathbb{P}(\sup\limits_{0\leqslant k\leqslant N} |S_k| \geqslant i) \leqslant V_n = \sum\limits_{i=1}^{n} \sigma_i^2$. Il suffit en effet d'appliquer l'inégalité de Doob à $Z_n = S_n^2$.
3. *(Loi du log itéré)* Soient (X_n) iid $N(0,1)$. Alors $\overline{\lim} \frac{S_n}{(2n \ln\ln n)^{1/2}} = 1$ presque sûrement, où $S_n = \sum\limits_{i=1}^{n} X_i$. On va poser $h(n)$ le dénominateur et $x \mapsto e^{\theta x}$ fonction convexe à valeurs positives.

2.1.9 Chaînes de Markov

2.1.9.1 Chaînes de Markov et martingales

Exemples

1. En fait, on peut le montrer plus vite avec des martingales.
On a $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = m\frac{X_n}{m} = X_n$. Maintenant, $(X_n)_n$ est une martingale bornée par m. Elle converge presque sûrement dans L^2 vers une variable aléatoire X_∞. Comme $\tau < +\infty$ presque sûrement, X_∞ est nécessairement à valeurs dans $\{0,m\}$. Si la chaîne est issue de x, $x = \mathbb{E}X_0 = \mathbb{E}X_n \longrightarrow \mathbb{E}X_\infty = m\mathbb{P}(X_\infty = m)$.
2. *(Urne de Polya par les chaînes de Markov)* Une urne contient des boules de m couleurs possibles ; à chaque instant entier, on tire au hasard une boule qu'on replace ensuite dans l'urne avec une boule de la même couleur. Cette modélisation intéresse également les instituts de sondage pour lesquels une personne ayant rencontré une autre personne dans la vraie vie, va prendre la même opinion. Au temps t, on note N_t^i le nombre de boules de couleur i, pour $i = 1,\ldots,m$. On part de $n_0 = N_0^1 + \ldots + N_0^m$ boules dans l'urne au temps 0 et donc au temps t, il y en a $n_0 + t$. La proportion $X_t^i = \frac{N_t^i}{n_0+t}$ est la proportion de boules de couleur i. (N_t^i) est une chaîne de Markov transiente (car croissante). On a $\mathbb{E}(N_{t+1}^i - N_t^i|\mathcal{F}_t) = X_t^i$. Ainsi $N_{t+1}^i = N_t^i$ avec une probabilité $1 - X_t^i$ et $X_t^i + 1$ avec la probabilité X_t^i. Ainsi $\mathbb{E}(X_{t+1}^i - X_t^i|\mathcal{F}_t) = \mathbb{E}\left(\frac{(n_0+t)N_{t+1}^i-(n_0+t+1)N_t^i}{(n_0+t)(n_0+t+1)}|\mathcal{F}_t\right) = 0$. Donc (X_t^i) est un martingale bornée, elle converge presque sûrement dans L^2. Le vecteur $(X_t^1\ldots X_t^m)$ converge également presque sûrement et dans L^2. Or $\mathbb{E}X_\infty^i = \mathbb{E}X_0^i = \frac{N_0^i}{n_0}$. Si $m = 2$, $N_0^1 = N_0^2 = 1$, alors $\mathbb{E}X_\infty^1 = 1/2$. On va chercher une autre martingale sous forme quadratique en N_t^i, sous la forme $a_tN_t(N_t+1)$. On a : $\mathbb{E}(a_{t+1}N_{t+1}^i(N_{t+1}^i(N_{t+1}^i + 1)|\mathcal{F}_t) = a_{t+1}[N_t^i(N_t^i+1)\frac{n_0+t-N_t^i}{n_0+1} + (N_t^i+1)(N_t^i+2)\frac{N_t^i}{n_0+t}) = a_{t+1}N_t(N_t+1)\frac{n_0+t+2}{n_0+t}$. Pour que ce soit une martingale, il suffit que $a_{t+1}(n_0+t+2) = a_t(n_0+t)$ soit $a_t(n_0+t)(n_0+t+1) =$ constante. Donc $Z_t^1 = \frac{N_t^1(N_t^1+1)}{(n_0+t)(n_0+t+1}$ est une martingale (c'est un produit de proportions). Alors $Z_t^{(k)} = \frac{N_t^1(N_t^1+1)\ldots(N_t^1+k-1}{(n_0+t)(n_0+t+1)\ldots(n_0+t+k-1)}$ est aussi une martingale. Elles sont bornées. Or $\lim_{t\to\infty} Z_t^{(k)} = \lim_{t\to\infty}\left(\frac{N_t^i}{t}\right)^k = (X_\infty^1)^k$. Donc $\mathbb{E}(X_\infty^1)^k = \frac{1}{k+1}$. On peut alors montrer que $X_\infty^1 \sim$ une loi uniforme.

2.1.9.2 Théorie du potentiel

Définition. ***(Matrice potentielle)***

Soit Q une matrice stochastique irréductible sur E. On va définir la *matrice potentielle* $G = (G(x,y))_{x,y\in E}$, par $G = I + Q + Q^2 + \ldots$ avec $G(x,y) = \sum_{n\geqslant 0} Q^n(x,y) = \mathbb{E}_X N_g \in \overline{\mathbb{N}}$.

En fait on a que $G(x,y) = \mathbb{P}_X(T_y < \infty)G(y,y) = +\infty$ dans le cas récurrent, $< \infty$ dans le cas transient, où $T_y = \inf\{n \geqslant 0, X_n = y\}$.

Définition. ***(Fonction harmonique)***

Soit $f : E \longrightarrow E$ une fonction telle que

$$\mathbb{E}_x|f(X_1)| < \infty \qquad \forall x \in E$$

et telle qu'il existe une partie A de E avec $Qf(x) = f(x)$ pour tout $x \in A$ (resp. $Qf(x) \leqslant f(x)$, $Qf(x) \geqslant f(x)$). On dit que f est *harmonique* (resp. *sur-harmonique*, *sous-harmonique*) sur A.
Si $T_{A^c} = \inf\{n \geqslant 0, X_n \in A^c\}$ désigne le temps de sortie de A, alors la suite $M_n = f(X_{n \wedge T_{A^c}})$ est une martingale (resp. une surmartingale, une sous-martingale).

▷ En exercice : $\mathbb{E}(M_{n+1}|\mathcal{F}_n) = Qf(M_n)$. ■

Définition. ***(Potentiel)***

Soit maintenant $D \subseteq E$ et $T = \inf\{n \geqslant 0, X_n \in D^c\}$. On se donne des fonctions $c : D \longrightarrow \mathbb{R}_+$ et $f : D^c \longrightarrow \mathbb{R}^+$ qu'on suppose bornées.
On définit un *potentiel* $\phi(x) = \mathbb{E}_x(\sum_{n=0}^{T-1} c(X_n) + f(X_T) indT < \infty)$.
On peut voir cette fonction ϕ comme le prix à payer avant d'atteindre D^c (T peut valoir $+\infty$ et $D = E$ est permis).

Proposition

1. Le potentiel vérifie $(\star)$: $\begin{cases} \phi = Q_\phi + c \text{ sur } D \\ \phi = f \text{ sur } D^c. \end{cases}$
2. Si $\mathbb{P}_x(T < \infty) = 1$ pour tout x, le problème $(\star)$ admet une unique solution bornée ϕ.

▷ Sur D^c, c'est clair que $\phi = f$. Si $x \in D$,

$$\phi(x) = c(x) + \mathbb{E}_x(\sum_{n=1}^{T-1} c(X_n) + f(X_T)\mathbb{1}_{T<\infty}) = c(x) + \mathbb{E}_x\mathbb{E}(\sum_{n=1}^{T-1} c(X_n) + f(X_T)\mathbb{1}_{T<\infty}|\mathcal{F}_1) = c(x) + \mathbb{E}_x\phi(X_1)$$

par la propriété de Markov.

Maintenant, soit $\psi : E \longrightarrow \mathbb{R}_+$, tel que $\psi \geqslant Q\psi + c$ sur D et $\psi \geqslant f$ sur D^c. Soit $x \in D$. Alors $\psi(x) \geqslant c(x) + \sum_{yinD^c} Q(x,y)f(y) + \sum_{y\in D} Q(x,y)\psi(y) = \mathbb{E}_x(c(X_0)\mathbb{1}_{T>0} + f(X_1)\mathbb{1}_{i=1} + \psi(X_1)\mathbb{1}_{T>1})$. On recommence en appliquant le même procédé au dernier terme de cette dernière somme. On arrive à montrer comme cela que $\psi(x) \geqslant \mathbb{E}_x(\sum_{i\leqslant n-1} c(X_i)\mathbb{1}_{T>i} + f(X_T)\mathbb{1}_{T\leqslant n} + \psi(X_n)\mathbb{1}_{T>N}))$. On a l'égalité ci-dessus si on a l'égalité dans la première inégalité du problème.

Si ψ est une solution bornée par M, alors $\mathbb{E}_x|\psi(X_n)|\mathbb{1}_{T>n} \leqslant M\mathbb{P}_x(T < n) \underset{n\to+\infty}{\longrightarrow} 0$ parle deuxième point. Dans ce cas, $\psi = \lim_{n\to+\infty} \mathbb{E}_x \sum_{i=1}^{N-1} c(X_i)\mathbb{1}_{T>1} + f(X_T)\mathbb{1}_{T\leqslant n}$, ie $\psi = \phi$ par le théorème de convergence monotone. ■

Exemples

1. Le $\phi(x) = \mathbb{P}_X(T < \infty)$ est une fonction harmonique sur D qui vaut 1 sur D^c. Donc $c = 0$ et $f = 1$. On a que $h \cong 1$ est toujours une solution et si elle est unique alors $\mathbb{P}_x(T < \infty) = 1$ pour tout x. Réciproquement, si $\mathbb{P}_x(T < \infty) = 1$ $\forall x$, alors on a unicité de la solution.

Remarque. Si $D = E$, $G = \sum_{n \geqslant 0} Q^n$, alors $\phi(x) = \mathbb{E}_x \sum_{n=0}^{\infty} c(X_n) = G_c(x)$.

2.1.9.3 Retournement du temps

La propriété de Markov dit que conditionnellement au présent, le passé et le futur sont indépendants, ce qui donne une impression de symétrie par rapport au présent.

On va voir ce qui se passe réellement lorsque l'on inverse la courbe du temps.

Définition. ***(Retournement du temps)***

Soit $(X_n)_{n \geqslant 0}$ une chaîne de Markov de noyau Q. La suite retournée dans le temps $Y = (Y^{(N)})^{n=0 \dots N}$ est définie par :

$$Y_n^{(N)} = X_{N-n}.$$

C'est une chaîne de Markov qui n'est pas homogène en temps en général.

▷ On a $\mathbb{P}(Y_n^{(N)} = x_n | Y_0^{(N)} = x_0, \dots, Y_{n-1}^{(N)} x_{n-1}) = \mathbb{P}(X_{N-n} = x_n | X_N = x_0 \dots X_{N-n+1} = x_{n-1}) = \frac{\mu_0 Q_1^{N-n}(x_n) Q(x_n,x_{n-1}) Q(x_{n-1},x_{n-2}) \dots Q(x_1,x_0)}{\mu_0 Q^{N-n+1}(x_{n-1}) Q(x_{n-1},x_{n-2}) \dots Q(x_1,x_0)}$ par la formule de Bayes, qui égale $\frac{\mu_0 Q^{N-n}(x_n)}{\mu_0 Q^{N-n+1}(x_{n-1}}$. C'est bien une chaîne de Markov. ■

Proposition

Si Π est une mesure invariante par Q avec $\Pi(x) > 0$ pour tout $x \in E$, alors $\tilde{Q}(x,y) = \frac{\Pi(y)}{\Pi(x)} Q(y,x)$ est une noyau = matrice de transition qui admet Π pour mesure invariante. De plus, la loi de $Y^{(N)}$ sachant les points de départ Y_0 et Y_N est la loi de la chaîne de Markov $\tilde{X}_n$ de noyau $\tilde{Q}$ conditionnée en ses valeurs extrêmes X_0 et $\tilde{X}_N$.

▷ On a $\sum_{y \in E} \tilde{Q}(x,y) = \frac{1}{\Pi(x)} \sum_{y \in E} \Pi(y) Q(y,x)$, cette somme égalant $\Pi(x)$, car Π est invariante. On veut montrer $\Pi \tilde{Q} = \Pi$. On a $\sum_y \Pi(y) \tilde{Q}(y,x) = \sum_y \Pi(Y) \frac{\Pi(x)}{\Pi(y)} Q(x,y) = \Pi(x)$. On va poser $x = (x_0,\dots,x_N)$ où $x_r = (X_N, \dots, x_0)$. Alors $\mathbb{P}(Y^N = x | Y_0 = x_0, Y_N = x_n) = \mathbb{P}(X = x_r | X_N = x_0, X_0 =$

$$x_n) = \frac{\prod_{i=1}^{N} Q(x_{N-i+1}, x_{N-i})}{\sum_{y_1,\dots,y_{N-1}} \prod_{i=1}^{N} Q(y_{N-i+1}, y_{N-i})} = \frac{\prod \tilde{Q}(x_{,N-i}, x_{N-i+1})}{\sum \prod \tilde{Q}(y_{N-i}, y_{N-i+1})} = \mathbb{P}(\tilde{X}_n = x | \tilde{X}_0 = x_0, \tilde{X}_s = x_z$$ en posant les notations évidentes. ■

Définition. ***(Chaîne stationnaire)***

Si Π est une probabilité invariante de noyau Q, la chaîne stationnaire est la suite $(X_n, n \in \mathbb{Z})$ telle que $X_0 \sim \Pi$ et $\mathbb{P}(X_{n+1} = y | \mathcal{F}_n^{\times}) = Q(X_n, y)$ pour tout n.

Remarque. Une telle chaîne existe. Soit $Y^{(m)}$ la chaîne qui démarre au temps $-m$ suivant la loi Π est de noyau de transition Q. La suite des lois des $Y^{(m)}$, $m \in \mathbb{N}$ est une famille compatible de probabilités. Par le théorème de Kolmogorov, $(X_n, n \in \mathbb{Z})$ est bien définie.

Conséquences

1. Si Π est une mesure réversible sur Q avec $\Pi(x) > 0$ pour tout $x \in E$, alors $\tilde{Q} = Q$ et donc $\mathbb{P}(X = x | X_0 = x_0, X_N = x_n) = \mathbb{P}(X = x_r | X_0 = x_n, X_N = x_0)$.
2. Si Π est une probabilité réversible et si X_n est une chaîne stationnaire, on pose $Y_n = X_{-n}$. Alors $X \overset{d}{=} Y$.

▷ On montre le deuxième point. On a $\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \Pi(x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n) = \Pi(x_1)Q(x_1, x_0)Q(x_1 x_2)\dots Q(x_{n-1}, x_n) = Q(x_1, x_0)Q(x_2, x_1)\Pi(x_2)\dots = \Pi(x_n)Q(x_n, x_{n-1})\dots Q(x_1, x_0)$. Donc $(x_0, \dots x_n)^d = (x_n \dots x_0)$. Ainsi $(X_N \dots X_0) \overset{d}{=} (X_0, \dots, X_{-N}) = (Y_0, \dots, Y_N)$. On peut le faire pour tout entier N d'où le résultat. ■

2.1.9.4 Graphes et conductances

On fait une analogie avec l'électricité.

Définition. ***(Graphes)***

Un *graphe orienté* est un couple $(E, \vec{\mathcal{E}})$ avec $\vec{\mathcal{E}} \subseteq E \times E$ un ensemble d'*arêtes orientées*. Le graphe est dit *connexe* si $\forall x,y \in E$, il existe un chemin de x à y le long des arêtes. Le graphe est *non orienté* si $\forall x, y \in E$, $(x,y) \in \vec{\mathcal{E}} \iff (y,x) \in \vec{\mathcal{E}}$.
Un graphe pondéré $(E, \mathcal{E}, c_{(xy)}, (x,y) \in \mathcal{E})$ est un graphe pour lequel chaque arête est muni d'une poids $c_{x,y} \geqslant 0$ qu'on interprète comme une conductance.
Si $c_e = 0$, alors l'arête e est *impraticable*. Si $c_e = \infty$, et $e = (x,y)$ alors $x = y$ et on agrège les deux états. Enfin, si $r = 1/c$, r est une *résistance*.

Proposition. ***(Graphes pondérés et chaînes de Markov)***

Soit Q un noyau de transition réversible pour une mesure Π. On lui associe le graphe simple pondéré $c_{xy} = \Pi(x)Q(x,y)$. Inversement, à tout graphe pondéré $(E,\mathcal{E},c_{xy)})$, on associe une chaîne de Markov réversible par $Q(x,y) = \frac{c_{x,y}}{c_x}$ avec $c_x = \sum_y c_{xy}$ (théorème de Milman). Une mesure réversible est alors $\mu(x) = c_x$ pour tout $x \in E$. On peut la normaliser si $c = \sum_{x\in E} c_x < \infty$. Alors $\Pi(x) = c_x/c$ définit une probabilité invariante.

$\triangleright$ $c(x,y) = c(y,x)$: le graphe est non orienté. Q est bien un noyau de transition réversible pour μ. ■

On introduit maintenant les réseaux de conductance.

Définition. ***(Flots)***

Un *flot* sur un graphe $(E,\vec{\mathcal{E}})$ est une fonction $f : \vec{\mathcal{E}} \longrightarrow \mathbb{R}$ tel que $f_{(x,y)} = -f_{(y,x)}$.
Le *flot algébrique* sortant de x est $f_x = \sum_{y\in E} f_{x,y}$. Alors on a que $\sum_{x\in E} f_x = 0$.
Un *flot unité* de B vers A est un flot f tel que $\sum_{x\in B} f_x = 1$ et $f_y = 0$ si $f \notin A \cup B$.

Remarque. En particulier, $f_{(x,x)} = 0$ pour tout $x \in E$. Heuristiquement, f peut être vue comme une intensité.

Le flot sortant de B vaut 1 et c'est un flot entre A et B.

Appendice