Notes de cours de
l'École Normale Supérieure de Lyon

13 janvier 2016

Probabilités

Florian Lavigne

Cours de préparation à l'agrégation de 2015-2016 de M. Xiaolin Zeng

Table des matières

1 Notions d'intégration

Définition. Un π-système est une famille stable par intersection finie.

Définition. Un λ-système est une famille f telle que :
- $\Omega \in f$.
- Si $A, B \in f$, avec $A \subset B$, alors $B \setminus A \in f$.
- Soit $A_n \in f$, avec A_n suite croissante. Alors $\cup A_n \in f$

Théorème 1. *Un λ - système engendré par un π-système est une tribu.*

Corollaire 2. *Les tribus engendrées par des π-systèmes indépendants sont indépendantes.*

Exemple. *L'ensemble des $]-\infty, a]$ avec a un réel est un π-système qui engendre la tribu borélienne.*

2 Espaces probabilisés

2.1 Définitions

Rappel. *Soit Ω un espace. Une tribu F sur Ω est un ensemble de parties de Ω vérifiant :*
- $\emptyset \in F$.
- $A \in F \Rightarrow A^c \in F$.
- $(A_n)_{n\in\mathbb{N}} \in F \Rightarrow \cup_n A_n \in F$.

Définition. Un *espace de probabilité* est un espace mesuré $(\Omega, F, \mathbb{P})$, où $\Omega \neq \emptyset$, F est une tribu et $\mathbb{P}$ une mesure sur l'espace mesurable telle que $\mathbb{P}(\Omega) = 1$.

Définition. Une telle mesure $\mathbb{P}$ est une *mesure de probabilité.*

Définition. Un *événement* est un éléments de F.

Définition. Un événement E de probabilité 1 *id est* $\mathbb{P}(E) = 1$ est dit *presque sûr.*

2.2 Probabilités conditionnelles

Définition. Soit A et B deux événements tels que $\mathbb{P}(B) \neq 0$. On appelle *probabilité conditionnelle* de A sachant B le réel :

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

Définition. Deux événements A et B sont *indépendants* si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. On note $A \amalg B$.

Définition. Une suite d'événements $(A_i)_{i\in I}$ est *indépendante* si quel que soit $J \subset I$ fini :

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

Définition. Deux sous-tribus G et H de F sont *indépendantes* si tout élément de G et tout élément de H sont indépendants.

2.3 Lemmes de Borel-Cantelli

Lemme 1. *Soit (A_n) une suite d'événements de l'espace de probabilités $(\Omega, F, \mathbb{P})$. On suppose que :*

$$\sum_{n\geq 0} \mathbb{P}(A_n) < \infty.$$

Alors une infinité de A_i se produisent simultanément $\mathbb{P}$-presque sûrement.

Démonstration. On pose A l'événement "une infinité de A_i se produisent simultanément".

Remarquons d'abord que :

$$A = \bigcap_{N\geq 0} \bigcup_{n\geq N} A_n.$$

On a alors :

$$\mathbb{P}(A) = \liminf_N \mathbb{P}\left(\bigcup_{n\geq N} A_n\right) \leq \lim_N \sum_{n\geq N} \mathbb{P}(A_n) = 0.$$

□

Lemme 2. *Si la suite d'événements (A_n) est indépendante, alors si $\sum \mathbb{P}(A_n) = \infty$ et si A est l'événement "une infinité de A_i se produisent simultanément". Alors $\mathbb{P}(A) = 1$.*

Démonstration. Soit $C_n = \cap_{p\geq n} A_p^c$. Alors $\mathbb{P}(\cup \uparrow C_n) = \limsup \mathbb{P}(C_n)$.

Or si $x \geq 0$, $1 - x \leq e^{-x}$, et donc :

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

Donc pour tout n, $\mathbb{P}(C_n) = 0$ et $\mathbb{P}(C_n^c) = 1$. Or $A = \cap C_n^c$. Donc $\mathbb{P}(A) = 1$. □

3 Variables aléatoires

3.1 Définitions

Définition. On appelle *variable aléatoire* à valeurs dans l'espace mesurable (Γ, E) toute application mesurable X de (Ω, F) dans (Γ, E).

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

Définition. La *loi* de la variable aléatoire X à valeurs dans l'espace mesurable (Γ, E), la mesure de probabilités sur (Γ, E) image de $\mathbb{P}$ par X. On la notera μ_X.

La mesure μ_X est donc définie par $\mu_X(A) = \mathbb{P}(X \in A)$ pour $A \in \Gamma$, avec :

$$(X \in A) \Leftrightarrow \{\omega, X(\omega) \in A\}.$$

3.2 Lois discrètes

On considère ici le cas où la variable aléatoire serait à valeur dans $\mathbb{N}$. On rappelle ici la *mesure de Dirac* au point n, δ_n, qui vaut 0 sauf en n où elle vaut 1. Alors la loi pourra se mettre sous la forme :

$$\textstyle\sum_{n\geq 0} p_n \delta_n \text{ avec } (p_n) \text{ une suite positive telle que } \sum p_n = 1.$$

Loi de Bernoulli

Soit un paramètre $0 \leq p \leq 1$. La *loi de Bernoulli* de paramètre p est définie par :

$$(1-p)\delta_0 + p\delta_1.$$

Loi binomiale

Soit $n \in \mathbb{N}^*$ et $0 \leq p \leq 1$. La *loi binomiale* de paramètre (n,p) est définie par :

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k.$$

Loi géométrique

Soit $0 < a < 1$. La *loi géométrique* de paramètre a est donnée par :

$$(1-a)\sum_{k \geq 0} a^k \delta_k.$$

Loi de Poisson

Soit $\lambda > 0$. La *loi de Poisson* est définie par :

$$e^{-\lambda} \sum_{k \geq 0} \frac{\lambda^k}{k!}.$$

3.3 Lois absolument continues

Un autre cas important est celui où la loi est définie à travers la donnée d'une fonction mesurable positive f, d'intégrale 1 par rapport à la mesure de Lebesgue *id est* $\int f(x)dx = 1$. On dit alors que f est la *densité* de la loi.

Loi uniforme

Soit $a < b$. La *loi uniforme* sur $[a,b]$ a pour densité :

$$\frac{1}{b-a}\mathbf{1}_{[a,b]}(x).$$

Loi normale ou Loi de Gauss

Soit $m \in \mathbb{R}$, et $\sigma > 0$. La *loi normale* de paramètre (m, σ^2) est définie par :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

On la notera $\mathcal{N}(m, \sigma^2)$.

Loi exponentielle

Soit $\lambda > 0$. La *loi exponentielle* de paramètre λ a pour densité :

$$\lambda e^{-\lambda x}\mathbf{1}_{x \geq 0}.$$

Loi Gamma

Soit $a > 0$. Alors la *Loi Gamma* de paramètre a est de densité :

$$\frac{1}{\Gamma(a)}\mathbf{1}_{t\geq 0}t^{a-1}e^{-t}dt.$$

Loi Beta

Soit a, b. Alors la *loi Beta* de paramètre (a, b) est définie par :

$$\frac{t^{a-1}(1-t)^{b-1}}{\beta(a,b)}\mathbf{1}_{0\leq t\leq 1}dt.$$

Remarque. (Simulation d'une variable aléatoire)
Si F est la fonction de répartition d'une loi μ et U est une variable aléatoire de loi uniforme $[0, 1]$. Alors $F^{-1}(U)$ est de loi μ. En effet :

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

Ainsi on peut simuler via ordinateur des variables aléatoires qui ont une loi différente de la loi uniforme.

3.4 Exemple de calcul de loi

Soit X de loi $\mathcal{N}(m, \sigma^2)$. On cherche la loi du couple $(X, Y) = (X, X^2)$.
Alors il suffit de trouver $z(x, y)$ tel que pour tout f, $\mathbb{E}(f(X, Y)) = \int\int f(x, y)z(x, y)dxdy$.
Alors

$$\begin{aligned}
\mathbb{E}[f(X, X^2)] &= \int f(x, x^2)e^{-\frac{(x-m)^2}{2\sigma^2}}\frac{1}{\sqrt{2\pi\sigma^2}}dx \\
&= \int\int f(x, y)e^{-\frac{(x-m)^2}{2\sigma^2}}\frac{1}{\sqrt{2\pi\sigma^2}}\delta_{x^2}(y)dxdy
\end{aligned}$$

3.5 Fonction de répartition

Définition. Soit X une *v.a.r.*. Sa *fonction de répartition* est la fonction :

$$\begin{aligned}
F_X : \mathbb{R} &\to [0, 1] \\
t &\mapsto \mathbb{P}(X \leq t) = \mathbb{P}_X(]-\infty, t]).
\end{aligned}$$

Proposition 3. *Deux lois qui ont la même fonction de répartition sont égales.*

Proposition 4. *F_X a les propriétés suivantes :*
(i) elle est croissante.
(ii) elle est continue à droite.
(iii) $\lim_{-\infty} F_X = 0$.
(iv) $\lim_{+\infty} F_X = 1$.

Remarque. Une fonction qui vérifie les points de la propriété 4 est alors fonction de répartition d'une variable aléatoire.

Proposition 5. *Si $\mathbb{P}_X$ a une densité f continue. Alors F_X est une primitive de f et réciproquement si $F_X \in C^1$ alors $\mathbb{P}_X$ a la densité F'_X.*

3.6 Espérance

On considère ici que des variables aléatoires réelles (*v.a.r.*).

Définition. Soit X une variable aléatoire. L'*espérance* de X est la quantité définie par :

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

Définition. La loi d'une *v.a.r.* X est dite *centrée* si $\mathbb{E}(X) = 0$.

Théorème 6. *Théorème de transfert*

Soit (Ω', F') un autre espace probabilisable et soit $f : \Omega \to \Omega'$ une fonction (F, F')-mesurable. Alors pour toute application borélienne $g : \Omega' \to \mathbb{R}^d$, on a :

$$g \in L^1(\mu_f) \Leftrightarrow g \circ f \in L^1(\mathbb{P}).$$

Dans ce cas, on a alors : $\mathbb{E}_\mathbb{P}(g \circ f) = \mathbb{E}_{\mu_f}(g)$.

Démonstration. L'intégrabilité de g équivaut à celle de toutes ses composantes. On peut donc se restreindre au cas où $d = 1$.

Par définition de μ_f, si $A \in F'$, et $g = \mathbf{1}_A$, alors :

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

Donc l'énoncé est vrai pour g une variable aléatoire.

Si g est positive, alors elle est limite d'une suite croissante de variables aléatoires positives, ce qui permet d'étendre l'égalité à ce cas-là.

De plus :

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

Dans le cas général, on décompose : $g = g^+ - g^-$. On remarque que $(g \circ f)^+ = g^+ \circ f$. On écrit alors :

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

L'égalité se déduit du cas précédent et de la linéarité de l'espérance. □

Définition. On appellera le *moment d'ordre* p de la *v.a.r.* X la quantité $\mathbb{E}(X^p)$, quand celle-ci est définie.

Définition. On note L^p l'ensemble des variables aléatoires réelles admettant un moment d'ordre p.

Proposition 7. *Si $1 \leq p < q$, alors $L^q \subset L^p$.*

Démonstration. Soit $X \in L^q$. Par inégalité de Hölder,$\|XY\|_1 \leq \|X\|_a \|Y\|_{a'}$, pour $a \geq 1$. En considérant $Y = 1$, on a :

$$\mathbb{E}[X] \leq \mathbb{E}[X^a]^{1/a}.$$

Alors $\mathbb{E}[|X|^p] \leq \mathbb{E}[|X^p|^{q/p}]^{p/q} < \infty$. On obtient donc l'inclusion. □

Remarque. Soit X une v.a.r. admettant un moment d'ordre 2. Posons $\phi : a \mapsto \mathbb{E}[(X-a)^2]$. On sait que :

$$\phi'(a) = -2\mathbb{E}[X] + 2a.$$

Alors ϕ est minimisée en $\mathbb{E}[X]$. Cette information justifie la prochaine notion.

Définition. La *variance* d'une *v.a.r.* X qui admet un moment d'ordre 2 est :

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

Proposition 8. *Si $Var(X) = 0$, alors $X = \mathbb{E}[X]$ p.p.*

Démonstration. Par définition de la variance, on sait que $\mathbb{E}[(X-\mathbb{E}[X])^2] = 0$, qui est l'espérance d'une variable positive. Donc $(X - \mathbb{E}[X])^2 = 0$ p.p. ou encore $X = \mathbb{E}[X]$ p.p. □

Définition. La loi X est *réduite* si sa variance est 1.

Proposition 9. *Soit X une v.a.r. et c une constante. Alors :*

$$Var(X + c) = Var(X).$$

Démonstration. On a :

$$\begin{aligned} Var(X+c) &= \mathbb{E}(X^2 + 2cX + c^2) - (\mathbb{E}(X)^2 + 2\mathbb{E}(c)\mathbb{E}(X) + \mathbb{E}(c)^2) \\ &= Var(X), \end{aligned}$$

car $\mathbb{E}(c) = c$. □

Définition. Soit X une *v.a.r.* qui admet un moment d'ordre 2. L'*écart-type* de X est :

$$\sigma(X) = \sqrt{Var(X)}.$$

Théorème 10. *Inégalité de Markov*

Soit X une v.a.r. positive. Alors pour tout $a > 0$ on a :

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}.$$

Démonstration. On a $a\mathbb{1}_{X>a} \leq X$. Ainsi en passant par l'espérance on obtient :

$$a\mathbb{P}(X > a) \leq \mathbb{E}(X).$$

□

Corollaire 11. *Inégalité de Bienaymé-Tchebychev*

Soit X une v.a.r. admettant un moment d'ordre 2. Alors pour tout $\varepsilon > 0$:

$$\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) \leq \frac{Var(X)}{\epsilon^2}.$$

Démonstration. La preuve applique l'inégalité de Markov à la variable positive $|X - \mathbb{E}(X)|^2$ et au réel ε^2, et utilise le fait que :

$$\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) = \mathbb{P}(|X - \mathbb{E}(X)|^2 > \varepsilon^2).$$

□

3.7 Indépendance

Définition. Soit (X_n) une suite de variables aléatoires réelles. Elles sont *indépendantes* si pour tout $n \geq 1$ pour tout choix d'entiers $1 \leq i_1 < ...i_k \leq n$ et de fonctions mesurables $f : \mathbb{R}^k \to \mathbb{R}$ et $g : \mathbb{R}^{n-k} \to \mathbb{R}$ on a :

$$\mathbb{E}(f(X_{i_1}, ..., X_{i_k})g(X_{j_1}, ..., X_{j_{n-k}})) = \mathbb{E}(f(X_{i_1}, ..., X_{i_k}))\mathbb{E}(g(X_{j_1}, ..., X_{j_{n-k}})),$$

avec $j_1, ..., j_{n-k}$ les entiers inférieurs à n différents des i_r.

Définition. On appelle des variables aléatoires *indépendantes et identiquement distribuées*, abrégé en *i.i.d.*, des variables aléatoires indépendantes de même loi.

Exemple. *Soit $X_1, \cdots, X_{n-1}$ variables aléatoires indépendantes de loi uniforme sur $[-1, 1]$.*
Soit $X_n = sign\,(\prod X_k)$.
Alors $(X_1, ..., X_n)$ ne sont pas indépendants, car :

$$\mathbb{P}(X_1, ..., X_{n-1} \geq 0, X_n = -1) = 0 \neq \mathbb{P}(X_n = -1)\prod \mathbb{P}(X_i \geq 0).$$

On va montrer que si on prend $n-1$ de ces variables aléatoires, alors elles sont indépendantes. En effet :

$$\begin{aligned}\mathbb{P}(\forall i \leq n-2, a_i < X_i < b_i, X_n = 1) &= \mathbb{P}(X_n = 1|\forall i \leq n-2, a_i < X_i < b_i)\mathbb{P}(\forall i \leq n-2, a_i < X_i < b_i) \\ &= \mathbb{P}(X_n = 1)\mathbb{P}(a_1 < X_1 < b_1, ..., a_{n-2} < X_{n-2} < b_{n-2}) \\ &= \mathbb{P}(X_n = 1)\mathbb{P}(a_1 < X_1 < b_1)...\mathbb{P}(a_{n-2} < X_{n-2} < b_{n-2})\end{aligned}$$

Proposition 12. *La suite $(X_i)_{i\in I}$ est indépendante ssi pour tout $J \subset I$ fini, en posant $\chi = (X_j)_{j\in J}$ on a $\mu_\chi = \bigoplus_{j\in J} \mu_{X_j}$.*

Démonstration. Soit $(A_j) \in B(\mathbb{R})^J$. La loi μ_χ est entièrement déterminée par ses valeurs du type $\mu_\chi(\prod A_j)$. Il en est de même pour l'autre loi. Mais par indépendance :

$$\begin{aligned}\mu_\chi(\prod_j A_j) &= \mathbb{P}(\cap_j X_j^{-1}(A_j)) \\ &= \prod_j \mathbb{P}(X_j^{-1}(A_j)) \\ &= \prod_j \mu_{X_j}(A_j) \\ &= \bigoplus_j \mu_{X_j}(\prod A_j).\end{aligned}$$

Réciproquement, l'indépendance se déduit facilement de l'égalité. □

Corollaire 13. *Si $X \sqcup Y$, alors quels que soient $g, h : \mathbb{R} \to \mathbb{R}$ boréliennes telles que $g(X)$ et $h(Y)$ soient dans $L^2(\mathbb{P})$, on a :*

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

Démonstration. Il suffit de le vérifier d'abord pour les fonctions en escaliers, puis pour les limites. □

Proposition 14. *La suite (X_i) est indépendante ssi pour tout $J \subset I$ fini, on a en posant $\chi = (X_j)_{j\in J}$:*

$$F_\chi((t_j)_{j\in J}) = \prod_j F_{X_j}(t_j).$$

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

$$F_\chi(t_j) = \mathbb{P}(\cap X_j^{-1}(]-\infty, t])) = \prod_j \mathbb{P}(X_j^{-1}(]-\infty, t])) = \prod_j F_{X_j}(t).$$

Supposons maintenant l'égalité des fonctions.

L'ensemble des pavés $\prod_j]-\infty, t]$ forme un système générateur de $B(\mathbb{R}^J)$. Comme on a l'égalité précédente, $\mathbb{P}_\chi$ et $\oplus \mathbb{P}_{X_j}$ coïncident sur ce système, et donc sont égales. On a donc l'indépendance. □

Remarque. Les résultats suivants N'implique PAS l'indépendance.
- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.
- $\phi_{X+Y} = \phi_X \phi_Y$.

Théorème 15. *Loi du 0-1*

Soit (X_n) une suite de variables aléatoires indépendantes. On pose $G = \cap_n \sigma(X_k, k \geq n)$ la tribu de queue. Alors pour tout $A \in G$, $\mathbb{P}(A) = 0$ ou 1.

Démonstration. Soit $F_\infty = \sigma(X_0, ..., X_n, ...)$. Remarquons que $G \subset F_\infty$. Soit $A \in G$ et :

$$C = \{B \in F_\infty, \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}.$$

On vérifie que C est une classe monotone. De plus, C contient l'algèbre de Boole :

$$B = \bigcup_{n \geq 0} \sigma(X_0, ..., X_n).$$

Par le théorème de classe monotone, il contient donc la tribu engendrée par B, c'est à dire F_∞. En particulier, $A \in C$ et donc A est indépendant de lui même. Donc $\mathbb{P}(A) = 0$ ou 1, ce qui donne le résultat souhaité. □

Remarque. (Associativité)

Soit $X_1, ..., X_n$ indépendantes. Alors pour tout $I \cap J = \emptyset$ et $I, J \subset \mathbb{N}^\star$, on a que $(X_i)_{i \in I}$ et $(X_j)_{j \in J}$ sont indépendantes.

Proposition 16. *Si X et Y sont indépendantes, alors la loi de (X, Y) s'écrit $f(x)g(y)dxdy$.*

4 Transformations exponentielles

4.1 Fonction caractéristique

Définition. Soit X une *v.a.* dans $\mathbb{R}^n$.

Sa *fonction caractéristique* est : $\phi_X(u) = \mathbb{E}\left(e^{i\langle u, X\rangle}\right)$.

Remarque. Il s'agit en fait de la transformée de Fourier de μ_X, ce qui par inversion donne que deux *v.a.r.* qui ont la même fonction caractéristique ont la même loi (nous détaillerons la preuve un peu plus loin dans le polycopié).

Proposition 17. *Une fonction caractéristique est uniformément continue.*

Démonstration. On ne traite que le cas $n = 1$.

On a :

$$|\phi_X(t+h) - \phi_X(t)| \leq \mathbb{E}(|e^{itX}(e^{ihX} - 1)|) \leq \mathbb{E}(|e^{ihX} - 1|).$$

Or la limite quand h tend vers 0 de $e^{ihX} - 1$ est 0, et que cette fonction est majorée en norme par 2, on a par convergence dominée que $\mathbb{E}(|e^{ihX} - 1|) \to 0$. □

Proposition 18. *Si X admet des moments jusqu'à l'ordre m alors ϕ_X est m-fois continûment dérivable et :*

$$\frac{\partial^m}{\partial u_{j_1}...\partial u_{j_m}}\phi_X(u) = i^m \mathbb{E}(X_{j_1}...X_{j_m}e^{i\langle u,X\rangle}).$$

Corollaire 19. *Si X réelle admet un moment à l'ordre n, alors :*

$$\phi_X^{(n)}(0) = i^n \mathbb{E}(X^n).$$

Proposition 20. *Soit X v.a. à valeurs dans $\mathbb{R}^n$, $a \in \mathbb{R}^m$ et $A \in M_{m,n}(\mathbb{R})$ Alors :*

$$\phi_{a+AX}(u) = e^{i\langle u,a\rangle}\phi_X(A^T u).$$

Proposition 21. *X et Y sont indépendantes ssi $\mathbb{E}(e^{i(\lambda X+\mu Y)}) = \mathbb{E}(e^{i\lambda X})\mathbb{E}(e^{i\mu Y})$*

Théorème 22. *Soit μ une mesure de probabilité et $\phi(t) = \int e^{itx}\mu(dx)$. Soit $a < b$.*

Alors la limite suivante existe et :

$$\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\phi(t)dt = \mu(]a,b[) + \frac{1}{2}\mu(\{a,b\}).$$

Démonstration. Soit I_T l'intégrale dont on souhaite déterminer la limite. Alors :

$$\begin{aligned} I_T &= \int_{-T}^{T}\int \frac{e^{-ita} - e^{-itb}}{it}e^{itx}\mu(dx)dt \\ &= \int \left[\int_{-T}^{T} \frac{\sin(t(X-a))}{t}dt - \int_{-T}^{T} \frac{\sin(t(X-b))}{t}\right] dt, \end{aligned}$$

car $\int_{-T}^{T} \frac{\cos(ct)-\cos(dt)}{t}dt = 0$.

Posons $S(t) = \int_0^T sinc(t)dt$. Alors :

$$\int_{-T}^{T} \frac{\sin(t\theta)}{t}dt = 2.sign(\theta).S(T|\theta|) = R(\theta, T).$$

Donc $I_T = \int R(X-a,T) - R(X-b,T)\mu(dx)$, avec $S(t) \to \pi/2$.

Donc, sous réserve de convergence de dominée (assurée par la finitude de μ) :

$$I_T \to \pi\mu(\{a,b\}) + 2\pi\int_a^b \mu(dx).$$

La preuve est donc terminée. □

Théorème 23. $\mu(\{a\}) = \lim \frac{1}{2\pi}\int_{-T}^{T} e^{ita}\phi(t)dt.$

Remarque. Si ϕ est à valeurs réelles, alors X et $-X$ ont même loi.

Théorème 24. *Si $\phi \in L^1$ alors μ est à densité continue et bornée. De plus, $f(y) = \frac{1}{2\pi}\int e^{-ity}\phi(t)dt$.*

Démonstration. Par le théorème précédent, on a :

$$\mu(]a,b[) + \frac{1}{2}\mu(\{a,b\}) \leq \frac{|b-a|}{2\pi}\int |\phi(t)|dt < \infty.$$

On prend $a = b$. Donc la mesure n'a pas d'atomes. Ainsi :

$$\begin{aligned} \mu([x,x+h]) &= \frac{1}{2\pi}\int \frac{e^{-itx} - e^{-it(x+h)}}{it}\phi(t)dt \\ &= \frac{1}{2\pi}\int\int_x^{x+h} e^{-ity}y\phi(t)dt \\ &= \int_x^{x+h}\left[\frac{1}{2\pi}\int e^{-ity}\phi(t)dt\right] dy, \end{aligned}$$

ce qui nous donne la fonction de densité. □

4.2 Transformée de Laplace

Définition. Soit X v.a.r. positive. La *transformée de Laplace* de X est :

$$\psi_X(\lambda) = \mathbb{E}(e^{-\lambda X}).$$

Remarque. En général, pour les v.a. quelconque, il se peut que la transformée de Laplace ne soit pas définie.

Remarque. La transformée de Laplace n'est utile que quand elle est définie sur un voisinage de $0 \in \mathbb{R}^n$.

Théorème 25. *(Théorème de Kac)*
Soit X et Y deux variables aléatoires positives. Alors elles sont indépendantes ssi

$$\forall \lambda, \mu \in \mathbb{R}^+, \mathbb{E}(e^{-\lambda X - \mu Y}) = \mathbb{E}(e^{-\lambda X})\mathbb{E}(e^{-\mu Y}).$$

Définition. Soit X une v.a.r. La *transformée de Laplace* de X est :

$$\begin{aligned} L_X : I &\to \mathbb{R} \\ t &\mapsto \mathbb{E}[e^{-tX}] \end{aligned}$$

avec I un intervalle contenant 0.

Remarque. Si X est positive, on a donc $\mathbb{R}^+ \subset I$.

Lemme 3. *Soit μ une mesure sur $\mathbb{R}^+$ et $\int_0^\infty e^{-\alpha x} d\mu(x) < \infty$. Alors :*

$$\forall t > \alpha, \int_0^\infty e^{-tx} d\mu(x) < \infty.$$

Démonstration. On a :

$$\int_0^\infty e^{-tx} d\mu(x) \leq \int_0^\infty e^{-\alpha x} d\mu(x) < \infty.$$

□

Remarque. On a (s'il n'y a pas d'atome) :

$$\mathbb{E}[e^{-tX}] = \mathbb{E}[e^{tX^+}] + \mathbb{E}[e^{tX^-}].$$

Or $L_X(0) = 1$. Donc I est bien un intervalle contenant 0.

Théorème 26. *(admis)*
Si X et Y sont deux variables aléatoires réelles telles que leur transformée de Laplace coïncident en une infinité de points équi-espacé alors X et Y sont de même lois.

Corollaire 27. *Si X, Y sont positives de même transformée de Laplace alors X et Y ont même loi.*

Corollaire 28. *Si X, Y sont bornés inférieurement de même transformée de Laplace alors X et Y ont même loi.*

5 Convergences de suites de variables aléatoires

5.1 Définitions

Soit (X_n) une suite de *v.a.r.* et X une *v.a.r.*

Définition. La suite (X_n) *converge presque sûrement* (p.s.) vers X si $\mathbb{P}(\{\omega, X_n(\omega) \to X(\omega)\}) = 1$.

Définition. La suite (X_n) *converge dans L^p* vers X si $\mathbb{E}(|X - X_n|^p) \to 0$.

Définition. La suite (X_n) *converge en probabilités* vers X si :

$$\forall \varepsilon > 0, \mathbb{P}(|X_n - X| > \varepsilon) \to 0.$$

Proposition 29. *La convergence dans L^p entraîne la convergence en probabilités.*

Démonstration. Notons X sa limite. Soit $\varepsilon > 0$. On a par l'inégalité de Markov :

$$\mathbb{P}(|X - X - n|^p > \varepsilon^p) < \frac{1}{\varepsilon^p}\mathbb{E}(|X - X_n|^p).$$

Le membre de droite tend vers 0. La convergence en probabilité s'en déduit. □

Proposition 30. *La convergence presque sûre entraînent la convergence en probabilités.*

Démonstration. Notons encore X la limite. On a alors $\mathbf{1}_{|X-X_n|>\varepsilon}$ qui converge p.s. vers 0 et qui est dominée par 1. Par convergence dominée, $\mathbb{P}(|X - X_n| > \varepsilon) \to 0$. □

Proposition 31. *Si (X_n) converge en probabilités vers X, il existe une sous-suite (X_{n_i}) qui converge p.s. vers X.*

Démonstration. Soit $\varepsilon > 0$. Pour $k \geq 1$, on peut trouver n_k tel que $\mathbb{P}(|X_{n_k} - X| > 1/k) \leq \frac{1}{k^2}$..

Ainsi $\sum \mathbb{P}(|X_{n_k} - X| > 1/k) < \infty$. Par le lemme de Borel-Cantelli, presque sûrement à partir d'un rang m, $|X_{n_k} - X| < 1/k$, ce qui donne la convergence voulue. □

Proposition 32. *Si $X_n \to X$ p.s. et si pour tout n, $|X_n| \leq Y \in L^q$, pour $1 < q < \infty$. Alors :*

$$X_n \to X \text{ dans } L^p, \text{ avec } 1 \leq p \leq q.$$

Démonstration. Par le lemme de Fatou, on a :

$$\mathbb{E}[|X|^q] = \mathbb{E}[\liminf |X_n|^q] \leq \liminf \mathbb{E}[|X_n|^q] \leq \mathbb{E}[Y^q] < \infty.$$

Donc $X \in L^q$. Soit $1 \leq p \leq q$. Alors :

$$\begin{aligned}
\mathbb{E}[|X_n - X|^q] &\leq \mathbb{E}[|X_n - X|^q \mathbf{1}_{|X_n - X|>\varepsilon}] + \varepsilon^p \\
&\leq \mathbb{E}[(|X_n - X|^q)^{q/p}]^{p/q} + \varepsilon^p, \text{ par l'inégalité de Hölder} \\
&\leq 2^p \mathbb{E}[Y^q]^{p/q} \mathbb{P}(|X_n - X| > \varepsilon) + \varepsilon^p \leq 2\varepsilon^p,
\end{aligned}$$

pour n suffisamment grand. □

Définition. La suite (X_n) converge en loi vers X si pour toute fonction $f : \mathbb{R} \to \mathbb{R}$ continue et bornée :

$$\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X)).$$

On notera alors $X_n \xrightarrow{e} X$.

Théorème 33. *(Slutsky)*
On a les propriétés suivantes :
(i) Si $X_n \xrightarrow{e} X$ *et* $Y_n \xrightarrow{e} 0$*, alors* $X_n + Y_n \xrightarrow{e} X$*.*
(ii) Si $X_n \xrightarrow{e} X$ *et* $Y_n \xrightarrow{e} a$*, alors* $X_n . Y_n \xrightarrow{e} aX$*.*

Démonstration. (i) Remarquons que $Y_n \xrightarrow{\mathbb{P}} 0$. Soit $x \in \mathbb{R}$ et $z' < x < z''$ avec :

$$\mathbb{P}(X = z'') = \mathbb{P}(X = z') = 0.$$

Si $z' < x - \varepsilon < x < x + \varepsilon < z''$, alors :

$$\mathbb{P}(X_n \leq z') - \mathbb{P}(|Y_n| \geq \varepsilon) \leq \mathbb{P}(X_n + Y_n \leq x) \leq \mathbb{P}(X_n \leq z'') + \mathbb{P}(|Y_n| \geq \varepsilon).$$

En passant à la limite, on a :

$$\mathbb{P}(X \leq z') \leq \liminf \mathbb{P}(X_n + Y_n \leq x) \leq \limsup \mathbb{P}(X_n + Y_n \leq x) \leq \mathbb{P}(X \leq z'').$$

Comme la fonction de répartition de X est continue, on obtient le résultat cherché.

(ii) Notons que $X_n Y_n = X_n a + X_n(Y_n - a)$. On sait que $Y_n - a \xrightarrow{e} 0$ et $aX_n \xrightarrow{e} aX$. Alors grâce au (i) on s'est ramené au cas $a = 0$.
Soit $\varepsilon > 0$ et $K > 0$ tel que $\mathbb{P}(|X| > K) \leq \varepsilon$. Soit $\lambda > 0$ tel que $\mathbb{P}(|X| = \frac{\varepsilon}{\lambda}) = 0$ et $\frac{\varepsilon}{\lambda} > K$. On a donc :

$$\mathbb{P}(|X_n| > \varepsilon/\lambda) \to \mathbb{P}(|X| > \varepsilon/\lambda) \leq \varepsilon.$$

Ainsi :

$$\limsup \mathbb{P}(|X_n Y_n| > \varepsilon) \leq \limsup \mathbb{P}(|Y_n| > \lambda) + \limsup \mathbb{P}(|X_n| > \varepsilon/\lambda) \leq \varepsilon.$$

$\square$

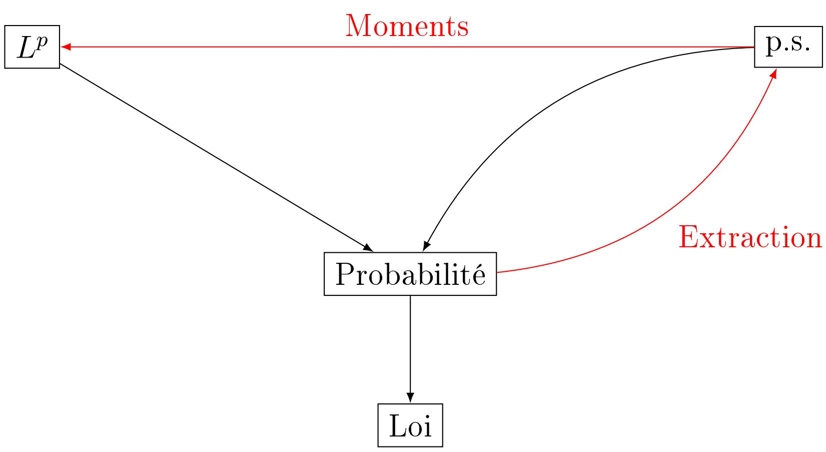
Proposition 34. *La convergence en probabilités entraîne la convergence en loi.*

Démonstration. Soit X sa limite. Soit f une fonction continue à support compact sur $\mathbb{R}$, qui est alors uniformément continue sur $\mathbb{R}$.

Pour $\varepsilon > 0$, il existe $a > 0$ tel que $|x - y| < a \Rightarrow |f(x) - f(y)| < \varepsilon$. On a alors :

$$\begin{aligned} |\mathbb{E}(f(X_n) - f(X))| &\leq \mathbb{E}(|f(X_n) - f(X)|) \\ &\leq \mathbb{E}(|f(X_n) - f(X)|\mathbf{1}_{|X - X_n| < a}) + \mathbb{E}(|f(X_n) - f(X)|\mathbf{1}_{|X - X_n| \geq a}) \\ &\leq \varepsilon + 2\|f\|_\infty \mathbb{P}(|X - X_n| \geq a) \leq 2\varepsilon, \end{aligned}$$

pour n assez grand. On conclut par densité le résultat voulu. $\square$



5.2 Contre-exemples

Proposition 35. *Soit X_n une suite de variables de Bernoulli de paramètre (p_n) alors :*

- $X_n \xrightarrow{\mathbb{P}} 0$ *ssi* $\lim(p_n) = 0$.
- $X_n \xrightarrow{p.s.} 0$ *ssi* $\sum_n p_n < \infty$.

Démonstration. Le premier point est évident. Montrons le second. Alors :

$$X_n(w) \to 0 \text{ ssi } X_n(w) \text{ stationnaire.}$$

Or $\{w, X_n(w) = 0 \text{ à partir d'un certain rang}\} = \liminf\{X_n = 0\}$. On conclut la dernière équivalence via les lemmes de Borel-Cantelli. □

Exemple. *Soit X de loi de Bernoulli de paramètre $1/2$. Soit $X_n = 1 - X$. Alors $X_n \xrightarrow{e} X$.*
Cependant, pour tout $\varepsilon < 1, \mathbb{P}(|X_n - X| > \varepsilon) = 1$.

Exemple. *Soit X_n avec $\mathbb{P}(X_n = n) = 1/n$ et $\mathbb{P}(X_n = 0) = \frac{n-1}{n}$. Alors :*

$$\mathbb{P}(|X_n - 0| > \varepsilon) = 1/n \to 0,$$

mais $\mathbb{E}(|X_n|) = 1$.

5.3 Méthode des moments

Théorème 36. *(admis)*
Si μ est une mesure de probabilité sur $\mathbb{R}$ admettant des moments de tout ordre notés m_k, et si la série entière $\sum \frac{m_k}{k!} z^k$ a un rayon de convergence positif. Alors μ est l'unique mesure de probabilité ayant (m_k) comme moments. On dit que μ est déterminée par ses moments.

Théorème 37. *(admis)*
Soit X déterminée par ses moments et X_n une suite ayant tout moment, avec :

$$\forall k, \mathbb{E}[X_n^k] \to \mathbb{E}[X^k].$$

Alors $X_n \xrightarrow{e} X$.

5.4 Théorème de Levy

Définition. Une suite de v.a.r. (X_n) est dite *tendue* si pour tout $\varepsilon > 0$, il existe $K > 0$ tel que :

$$\forall n, \mathbb{P}(|X_n| > K) < \varepsilon.$$

Proposition 38. *(admise)*
Toute suite réelle tendue admet une sous-suite qui converge en loi.

Lemme 4. *Soit X_n une suite de v.a.r. et X telles que pour tout $t \in \mathbb{R}, \phi_n(t) \to \phi(t)$. Alors la suite est tendue.*

Démonstration. Soit $\varepsilon > 0$. Par théorème de Fubini, on a pour tout $u > 0$:

$$\frac{1}{u}\int_{-u}^{u}(1 - \phi_n(t))dt = \int_{-\infty}^{\infty} \frac{1}{u}\int_{-u}^{u}(1 - e^{itX})dt\mu_n(dx) \geq 2\int_{|X|\geq 2/u}\left(1 - \frac{1}{|ux|}\right)\mu_n(dx).$$

Par ailleurs, le théorème de convergence dominée nous dit que $\frac{1}{u}\int_{-u}^{u}(1 - \phi_n(t))dt$ converge vers $\frac{1}{u}\int_{-u}^{u}(1 - \phi(t))dt$. Or ϕ est continue en 0. avec $\phi(0) = 1$. Donc on peut trouver u tel que, pour n assez grand :

$$\mathbb{P}(|X_n| \geq 1/u) \leq \frac{1}{u}\int_{-u}^{u}(1 - \phi_n(t))dt \leq \varepsilon.$$

On obtient donc une constante $C > 0$ et un entier N, tels que :

$$\forall n \geq N, \mathbb{P}(|X_n| \geq C) \leq \varepsilon.$$

Or la famille finie $(X_n)_{n\leq N}$ est tendue. Donc on obtient finalement que la suite globale (X_n) est tendue. □

Théorème 39. *Théorème de Levy*

Soit (X_n) une suite de v.a.r. et X une variable réelle. Alors :

$$X_n \xrightarrow{e} X \text{ ssi pour tout } t \in \mathbb{R},\ \phi_{X_n}(t) \to \phi_X(t).$$

Démonstration. Le sens direct est clair. Montrons dons l'autre sens. Grâce au lemme précédent, on sait que la suite est tendue. Soit (X_{n_p}) une sous-suite de (X_n) qui converge en loi vers la loi μ. On a alors grâce au sens direct du théorème de Levy, que $\phi_{X_{n_p}}(t) \to \psi$ avec ψ la fonction caractéristique de la loi μ. Par hypothèse, on a donc $\psi = \phi$. Ainsi (X_{n_p}) converge en loi vers X.

On conclut grâce à la proposition qui suit. □

Proposition 40. *(admise)*

Soit (X_n) une suite de variables aléatoires réelles tendue et X une v.a.r. Si toutes les sous-suites de (X_n) convergent en loi vers X, alors la suite (X_n) converge en loi vers X.

5.5 Loi des grands nombres

Lemme 5. *Si $p \geq 1$ et $X \in L^p$ positive, alors :*

$$\mathbb{E}[X^p] = \int_0^\infty pt^{p-1}\mathbb{P}(X > t)dt.$$

Démonstration. Par théorèmes de Fubini-Tonelli et de transfert, on obtient :

$$\begin{aligned}
\int_0^\infty pt^{p-1}\mathbb{P}(X > t)dt &= \int_0^\infty pt^{p-1} \int_t^\infty \mathbb{P}_X(dx)dt \\
&= \int_{\mathbb{R}^2} pt^{p-1}\mathbf{1}_{]t,\infty[}(x)\mathbb{P}_X(dx)dt \\
&= \int_0^\infty \int_0^x pt^{p-1}dt\mathbb{P}_X(dx) \\
&= \int_0^\infty x^p P_X(dx) = \mathbb{E}[X^p].
\end{aligned}$$

□

Théorème 41. *Soit (X_i) une suite de variables aléatoires i.i.d. deux à deux indépendantes. Alors :*

$$\frac{1}{n}\sum_{k\leq n} X_k \to \mathbb{E}[X_1] \quad p.s.$$

Démonstration. On peut se ramener au cas de variables aléatoires positives. De plus, si on pose $Y_k = X_k\mathbf{1}_{X_k\leq k}$ et $T_n = \sum_{k\leq n} Y_k$, alors on a déjà l'indépendance deux à deux des Y_k et $Y_k \in L^\infty$.

Soit $a > 1$ fixé et on pose $u_n = [a^n]$. Soit $\varepsilon > 0$. Alors :

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

Par inégalité de Bienaymé-Chebychev, on a :

$$\Sigma_\varepsilon := \sum_n \mathbb{P}\left(\left|\frac{T_{u_n} - \mathbb{E}[T_{u_n}]}{u_n}\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\sum_n \frac{1}{u_n}\mathbb{E}[X_1^2 \mathbf{1}_{X_1 \leq n}].$$

Soit $K = 2a/(a-1)$ et $x > 0$. Soit $N = \min\{n, u_n \geq x\}$. Alors $a^N \geq x$ et comme pour $y \geq 1$, on a toujours $y \leq 2[y]$, il vient que :

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

Ainsi en appliquant le théorème de Fubini-Tonelli et par l'inégalité précédente, on a :

$$\Sigma_\varepsilon \leq \frac{1}{\varepsilon^2}\mathbb{E}\left[X_1^2 \mathbf{1}_{X_1 > 0} \sum_{u_n \geq X_1} \frac{1}{u_n}\right] \leq \frac{K}{\varepsilon^2}\mathbb{E}[X_1] < \infty.$$

Ainsi par le lemme de Borel-Cantelli, on a $\mathbb{P}(\limsup \frac{1}{u_n}|T_{u_n} - \mathbb{E}[T_{u_n}]| > 0) = 0$. Cela veut justement dire que presque surement $(T_{u_n} - \mathbb{E}[T_{u_n}])/u_n$ tend vers 0.

De plus, par convergence monotone, $\mathbb{E}[Y_k] \to \mathbb{E}[X_1]$.

Donc par le lemme de Cesaro, $\mathbb{E}[T_{u_n}]/u_n \to \mathbb{E}[X_1]$. Donc $T_{u_n}/u_n \to \mathbb{E}[X_1]$ p.s.

Ensuite par le lemme précédent, on a :

$$\sum_k \mathbb{P}(X_k \neq Y_k) = \sum_k \mathbb{P}(X_k > k) \leq \int_0^\infty \mathbb{P}(X_1 > t)dt = \mathbb{E}[X_1] < \infty.$$

Ainsi, par le lemme de Borel-Cantelli, on a $\mathbb{P}(\limsup\{X_k \neq Y_k\}) = 0$. Ainsi, presque surement il existe $n(x)$ tel que si $n \geq n(x)$, en posant $S_n = \sum_{k \leq n} X_k$, on obtient :

$$(S_n - T_n)(x) = S_{n(x)}(x) - T_{n(x)}(x).$$

Donc $\lim \frac{S_n - T_n}{n} = 0$ presque surement. On obtient donc le bon résultat pour X_{u_n}. Puis par encadrement, on obtient finalement la loi des grands nombres. □

6 Applications de la Loi des Grands Nombres

Théorème 42. *Soit (X_n) une suite de v.a.i.i.d. à valeurs dans $[0, \infty[$.*

Soit $T_n = X_1 + ... + X_n$ et $N_t = \sup\{n, T_n \leq t\}$ (par exemple, il s'agit du nombre de clients qui sont passer en caisse avant le temps t).

Alors $\frac{N_t}{t} \to \frac{1}{\mathbb{E}[X_1]}$ p.s.

Démonstration. Par la Loi des Grands Nombres, $\frac{T_n}{n}$ converge presque sûrement vers $\mathbb{E}[X_1]$. Par définition :

$$\frac{N_t}{T_{N_t}} \geq \frac{N_t}{t} \geq \frac{N_t}{T_{N_t+1}} = \frac{N_t+1}{T_{N_t+1}}.\frac{N_t}{N_t+1}.$$

Or T_n est fini presque sûrement. Donc N_t tend vers ∞.

Alors presque sûrement N_t/T_{N_t} tend vers $1/\mathbb{E}[X_1]$ et $N_t/(N_t+1)$ tend vers 1. Par théorème d'encadrement des suites, on a le résultat. □

Remarque. Si on a seulement T_n/n qui converge en probabilité vers $\mathbb{E}[X_1]$ et $N_t \to \infty$ p.s., alors il n'est pas forcément vrai que :

$$\frac{N_t}{t} \to \frac{1}{\mathbb{E}[X_1]} p.s.$$

Théorème 43. *(Glivenko-Cantelli)*
Soit (X_n) une suite de variables aléatoires i.i.d. réelles. Soit :

$$F_{n,\omega}(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k(\omega)\leq x} \quad ; \quad F(x) = \mathbb{P}(X_1 \leq x),$$

Alors $\|F_{n,\omega} - F\|_\infty \to 0$.

Démonstration. • Traitons d'abord le cas de variables aléatoires uniformes sur $[0,1]$. Ainsi $F(x) = x$ pour $x \in [0,1]$.
De plus, si $x < 0, F_{n,\omega}(x) = F(x) = 0$ et si $x > 1, F_{n,\omega}(x) = F(x) = 1$.
Soit $q \in [0,1] \cap \mathbb{Q}$. On sait par la Loi des Grands Nombres que $F_{n,\omega}(q) \to q$ presque sûrement. Soit $k \in \mathbb{N}^*$ et $x \in [0,1]$. On pose $E_k = \lceil kx \rceil$. Ainsi :

$$|F_{n,\omega}(x) - x| \leq \left|F_{n,\omega}(x) - F_{n,\omega}\left(\frac{E_k}{k}\right)\right| + \left|F_{n,\omega}\left(\frac{E_k}{k}\right) - \frac{E_k}{k}\right| + \left|\frac{E_k}{k} - x\right|.$$

Or les $F_{n,\omega}$ sont croissantes. Alors :

$$|F_{n,\omega}(x) - x| \leq F_{n,\omega}\left(\frac{\lceil kx \rceil}{k}\right) - F_{n,\omega}\left(\frac{\lfloor kx \rfloor}{k}\right) + \max_{0\leq i\leq k}\left|F_{n,\omega}\left(\frac{i}{k}\right) - \frac{i}{k}\right| + \frac{1}{k}.$$

Ainsi :

$$\|F_{n,\omega} - F\|_\infty \leq \max_{0\leq i\leq k-1}\left[F_{n,\omega}\left(\frac{i+1}{k}\right) - F_{n,\omega}\left(\frac{i}{k}\right)\right] + \max_{0\leq i\leq k}\left|F_{n,\omega}\left(\frac{i}{k}\right) - \frac{i}{k}\right| + \frac{1}{k}.$$

Enfin, $\limsup \|F_{n,\omega} - F\|_\infty = 0$. On a donc bien la nullité de la limite voulue.

• Rappelons que si U est une variable aléatoire uniforme sur $[0,1]$, alors

$$F^{-1}(U) = \inf\{x \in \mathbb{R}, F(x) \geq U\}$$

a pour fonction de répartition F. De plus, le résultat cherché ne dépend que de la fonction de répartitions : on peut donc supposer les X_k sous cette forme. Considérons alors une suite (U_n) de variables aléatoires uniformes dans $[0,1]$. Ainsi :

$$\sup_{x\in\mathbb{R}} |F_{n,\omega}(x) - F(x)| = \sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{U_k\leq F(x)} - F(x)\right| \leq \sup_{y\in[0,1]}\left|\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{U_k\leq y} - y\right|.$$

Le membre de droite tend vers 0 par le premier point. On a donc le résultat dans le cas général.

□

6.1 Théorème central limite - TCL

Lemme 6. *Si $z_1, ..., z_n, y_1, ..., y_n$ sont des nombres complexes de module ≤ 1, alors :*

$$\left|\prod_{i=1}^{n} z_i - \prod_{i=1}^{n} y_i\right| \leq \sum_{i=1}^{n} |z_i - y_i|.$$

Démonstration. La preuve se fait par récurrence immédiate en utilisant l'égalité qui suit :

$$\prod_{i=1}^{n} z_i - \prod_{i=1}^{n} y_i = \left(\prod_{i=1}^{n-1} z_i - \prod_{i=1}^{n-1} y_i\right) z_n + \prod_{i=1}^{n-1} y_i(z_n - y_n).$$

□

Remarque. Par la formule de Taylor avec reste intégral, on a :

$$e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} = \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is} ds = \frac{i^{n+1}}{n!} \left(\frac{x^n}{i} - \frac{n}{i} \int_0^x (x-s) e^{is} ds \right).$$

Ainsi :

$$\left| e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} \right| \leq \frac{2|x|^n}{n!}.$$

Cette inégalité nous permet ainsi d'écrire pour une variable aléatoire X qui est dans $L^2(\mathbb{P})$ le développement limité à l'ordre 2 de sa fonction caractéristique.

Théorème 44. *Soit X_n une suite i.i.d. dans $L^2(\mathbb{P})$ de moyenne m et de variance σ^2.*

On définit $S_n = \sum_{k=1}^{n} X_k$. Alors :

$$\frac{S_n - Nm}{\sigma\sqrt{n}} \xrightarrow{e} \mathcal{N}(0,1).$$

Démonstration. Quitte à étudier $(X_n - m)/\sigma$, on peut supposer $m = 0$ et $\sigma = 1$. Posons alors ϕ la fonction caractéristique des X_n et :

$$Y_n = \frac{S_n}{\sqrt{n}}.$$

Calculons déjà ϕ_l la fonction caractéristique de la gaussienne centrée réduite.

$$\phi_l(t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{itx} e^{-\frac{x^2}{2}} dx = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{(x-it)^2}{2}} dx = e^{-\frac{t^2}{2}}.$$

Par indépendance, on a $\phi_{Y_n}(t) = \phi(t/\sqrt{n})^n$. Ainsi :

$$\phi_{Y_n}(t) - e^{-\frac{t^2}{2}} = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n - \left(e^{-\frac{t^2}{2n}}\right)^n.$$

Par le lemme précédent, on obtient :

$$|\phi_{Y_n}(t) - e^{-\frac{t^2}{2}}| \leq \sum_{k=1}^{n} |1 - \frac{t^2}{2n} + o(1/n) - e^{-\frac{t^2}{2n}}| \leq o(1).$$

□

6.2 Application du TCL

Soit une suite (X_n) de v.a.r. binomiales de paramètre p. Posons $S_n = X_1 + ... + X_n$. Alors :

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{(e)} \mathcal{N}(0,1).$$

Supposons $p = 1/2$ et $n = 100$. On veut calculer :

$$\mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq 1 \right).$$

Le calcul direct nous donne :

$$\mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq 1 \right) = (1/2)^{100} \sum_{k \geq 55} \binom{100}{k} = 0.184.$$

Via le TCL, il suffit de calculer $\mathbb{P}(\mathcal{N}(0,1) > 1) = 0.15$.

7 Lois conditionnelles

7.0.1 Cas discret

Définition. Soit X et Y deux variables aléatoires sur $(\Omega, \Sigma, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$ et $(F, \mathcal{F})$. On suppose que E est au plus dénombrable. Pour tout $A \in \mathcal{F}$ et pour tout $i \in E$, on pose $p(i, A) = \mathbb{P}(Y \in A | X = i)$.

Proposition 45. *L'application $p : E \times \mathcal{F} \to [0, 1]$ vérifie :*
- *$\forall A \in \mathcal{F}, i \mapsto p(i, A)$ est mesurable.*
- *$\forall i \in E, A \mapsto p(i, A)$ est une loi de probabilité de $\mathcal{F}$.*

Définition. On appelle *loi conditionnelle* de Y sachant $X = i$ la fonction $p(i, \cdot)$, et le *noyau de transition* est la fonction p.

Proposition 46. *La loi conditionnelle p est caractérisée par :*

$$\mathbb{E}(f(X, Y)) = \int_E \int_F f(x, y) p(x, dy) \mu_X(dx),$$

où μ_X est la loi de X et f mesurable bornée quelconque sur $E \times F$.

Démonstration. On a déjà :

$$p(A, B) = \sum p(X = i, Y \in B) = \sum p(Y \in B | X = i) \mathbb{P}(X = i) = \int_A \int_B p(x, dy) \mu_X(dx).$$

Réciproquement, si la formule est vraie, on prend $f(x, y) = \mathbf{1}_A(x) \mathbf{1}_B(y)$, ce qui nous donne : $\mathbb{P}(x \in A, y \in B) = \int_A \int_B (p(x, y) \mu_X(dx)$. En particulier :

$$\mathbb{P}(Y \in B | X = i) = \int_B \frac{p(i, dy) \mu_X(\{i\})}{\mathbb{P}(X = i)} = p(i, B).$$

Donc p est la loi conditionnelle de Y sachant X. □

Exemple. *Si X_i sont de loi $Bin(1, p)$, posons $Y = \sum_{i=1}^n X_i$ qui est de loi binomiale de paramètre (n, p). Alors :*

$$\mathbb{P}(X_1 = 0 | Y = k) = 1 - \frac{k}{n}.$$

7.0.2 Cas général

Remarque. On ne suppose plus que $(E, \mathcal{E})$ est dénombrable. On veut alors définir une loi conditionnelle comme dans le cas discret.

Définition. Un *noyau de transition* sur $E \times \mathcal{F}$ est une application $p : E \times \mathcal{F} \to [0, 1]$ telle que :
- $\forall A \in \mathcal{F}, i \mapsto p(i, A)$ est mesurable.
- $\forall i \in E, A \mapsto p(i, A)$ est une loi de probabilité de $\mathcal{F}$.

On dit que le noyau p est la *loi conditionnelle* de Y sachant X si :

$$\forall f : E \times F \to \mathbb{R} \text{ mesurable bornée}, \mathbb{E}(f(X, Y)) = \int_E \int_F f(x, y) p(x, dy) \mu_X(dx).$$

Théorème 47. *Si X et Y sont deux variables aléatoires à valeurs dans un espace polonais (métrique, séparable, complet). On munit E, F muni de leur tribu borélienne $\mathcal{E}, \mathcal{F}$. Alors il existe un noyau de transition p sur $E \times \mathcal{F}$ unique μ_X-ps qui est la loi conditionnelle de Y sachant X.*

Théorème 48. *Si (X,Y) a une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$, qu'on note $h(x,y)$, alors la loi conditionnelle de Y sachant X est donné par le noyau $p(x,dy) = p(x,y)dy$ où :*

$$p(x,y) = \frac{h(x,y)}{k(x)},$$

si $k(x) \neq 0$, et qui s'annule sinon, et avec :

$$k(x) = \int h(x,y)dy.$$

Démonstration. Clairement $p(x,dy)$ est un noyau et :

$$\mathbb{E}(f(X,Y)) = \int\int f(x,y)h(x,y)dxdy = \int\int f(x,y)p(x,y)dy\mu_X(dx).$$

□

Exemple. *Soit X de loi uniforme $[0,1]$, et Y tel que Y sachant X soit de loi uniforme $[0,X]$.*

On cherche la loi de (X,Y). Pour cela, on considère f mesurable bornée. Alors :

$$\begin{aligned}\mathbb{E}[f(X,Y)] &= \int_x\int_y f(x,y)\frac{1}{x}\mathbf{1}_{0\leq y\leq x}dy\mathbf{1}_{[0,1]}(x)dx \\ &= \int\int f(x,y)\frac{1}{x}\mathbf{1}_{0\leq y\leq x\leq 1}dxdy.\end{aligned}$$

Ainsi la loi de (X,Y) est $\frac{1}{x}\mathbf{1}_{0\leq y\leq x\leq 1}dxdy$.

Alors la loi de Y est :

$$\int \frac{1}{x}\mathbf{1}_{0\leq y\leq x\leq 1}dx = \int_y^1 \frac{dx}{x} = -\log(y)\mathbf{1}_{[0,1]}(y).$$

Enfin, on obtient la loi de X sachant Y qui est :

$$\frac{-1}{x\log y}\mathbf{1}_{0\leq y\leq x\leq 1}dx.$$

Exemple. *Soit X et Y de loi exponentielle de paramètre λ. Quelle est la loi de Y sachant $Z = X+Y$?*

Tout d'abord la loi de (Y,Z) est donnée par : $\lambda^2 e^{-\lambda z}\mathbf{1}_{0\leq y\leq z}dydz$. De plus, celle de Z qui n'est que l'intégrale de cette dernière par rapport à y, est :

$$\lambda^2 z e^{-\lambda z}\mathbf{1}_{0\leq z}dz.$$

En divisant ces deux lois, on obtient alors la loi de Y sachant Z qui vaut donc :

$$\frac{1}{z}\mathbf{1}_{0\leq y\leq z}dy.$$

8 Espérance conditionnelle

Le but de cette partie est de définir $\mathbb{E}[X|Y]$ ou $\mathbb{E}[X|\mathcal{F}]$ pour une tribu $\mathcal{F}$. Elle sera une variable aléatoire $\sigma(Y)$-mesurable.

8.1 Cas discret

Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, et Y une variable aléatoire de $Y \in E$ avec E un espace au plus dénombrable. On peut supposer quitte à réduire E, que pour tout $y \in E, \mathbb{P}(Y = y) > 0$.

Définition. On définit $\mathbb{E}[X|Y]$ comme la variable aléatoire qui vaut avec probabilité $\mathbb{P}(Y = y)$:

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

Exemple. *Soit X un dé et Y la parité de X. Alors :*

$$\mathbb{E}[X|Y] = \begin{cases} 3 \text{ si } Y = 1 \\ 4 \text{ sinon} \end{cases}$$

Proposition 49. $\mathbb{E}[\ |\ \mathbb{E}[X|Y]\ |\ ] \leq \mathbb{E}[\ |\ X\ |\ ]$.

Démonstration.

$$\begin{aligned} \mathbb{E}(|\mathbb{E}[X|Y]|) &= \sum |\mathbb{E}[X\mathbf{1}_{Y=y}]| \\ &\leq \sum \mathbb{E}[|X|\mathbf{1}_{Y=y}] \\ &\leq \mathbb{E}[|X|] \end{aligned}$$

□

Théorème 50. *La variable aléatoire $\mathbb{E}[X|Y]$ est caractérisée par :*

$$\forall Z \text{ bornée } \sigma(Y)\text{-mesurable, } \mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X|Y]Z].$$

Démonstration.

- Comme Z est $\sigma(Y)$-mesurable, il existe h mesurable telle que $Z = h(Y)$. Alors :

$$\begin{aligned} \mathbb{E}(Z\mathbb{E}(X|Y)) &= \sum_y h(y)\mathbb{E}(X\mathbf{1}_{Y=y}) \\ &= \sum_y \mathbb{E}(h(Y)X\mathbf{1}_{Y=y}) \\ &= \mathbb{E}(XZ) \end{aligned}$$

- Soit une variable $\tilde{X}$ une variable aléatoire qui vérifie cette propriété. Soit :

$$Z = \mathbf{1}_{\tilde{X} - \mathbb{E}[X|Y] > 0},$$

$\sigma(Y)$-mesurable. Alors :

$$\mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X|Y]],$$

ce qui nous donne $\mathbb{E}[(\tilde{X} - \mathbb{E}[X|Y])\mathbf{1}_{\tilde{X} - \mathbb{E}[X|Y] > 0}] = 0$. Alors :

$$\tilde{X} \leq \mathbb{E}[X|Y] \text{ p.s.}$$

De la même façon, on obtient l'inégalité inverse.

□

8.2 Cas intégrable

Théorème 51. *Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. Alors il existe une unique variable aléatoire dans $L^1(\Omega, \mathcal{G}, \mathbb{P})$ qui vérifie :*

$$\forall Z \ \mathcal{G}\text{-mesurable bornée, } \mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]].$$

Démonstration.

- **Existence :** Supposons $X \geq 0$. Posons Q la mesure sur $\mathcal{G}$ définie par pour tout $A \in \mathcal{G}$, $Q(A) = \mathbb{E}[X\mathbf{1}_A]$. Alors $Q << \mathbb{P}$. Par le théorème de Radon-Nykodym, il existe une fonction f mesurable bornée telle que $Q = f\mathbb{P}$. Donc il existe une variable aléatoire $\tilde{X}$ $\mathcal{G}$-mesurable de loi $f\mathbb{P}$.
  De plus, $\mathbb{E}[\tilde{X}] = \mathbb{E}[X] < \infty$. Donc $\tilde{X} \in L^1$, et comme :
  $$Q(A) = \mathbb{E}[\tilde{X}\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A],$$
  par approximation par des fonctions étagées, cette égalité est vraie pour tout Z $\mathcal{G}$-mesurable.
  Si X est quelconque, on applique le résultat à sa partie positive et sa partie négative, et on définit :
  $$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}].$$
- **Unicité :** Soit une variable $\tilde{X}$ une variable aléatoire qui vérifie cette propriété. Soit :
  $$Z = \mathbf{1}_{\tilde{X} - \mathbb{E}[X|Y] > 0},$$
  $\sigma(Y)$-mesurable. Alors :
  $$\mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X|Y]],$$
  ce qui nous donne $\mathbb{E}[(\tilde{X} - \mathbb{E}[X|Y])\mathbf{1}_{\tilde{X} - \mathbb{E}[X|Y] > 0}] = 0$. Alors :
  $$\tilde{X} \leq \mathbb{E}[X|Y] \ p.s.$$
  De la même façon, on obtient l'inégalité inverse.

□

Proposition 52. *Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ positive. Alors $\mathbb{E}[X|\mathcal{G}] \geq 0$.*

Démonstration. Supposons qu'il existe $a > 0$ tel que $\mathbb{P}(\mathbb{E}[X|\mathcal{G}] < -a) > 0$.
Posons $B = \{\mathbb{E}[X|\mathcal{G}] < -a\} \in \mathcal{G}$. Alors :

$$0 > \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_B] = \mathbb{E}[X\mathbf{1}_B] \geq 0,$$

ce qui est impossible. D'où la positivité de $\mathbb{E}[\cdot|\mathcal{G}]$ □

Proposition 53. *Si X est $\mathcal{G}$-mesurable, alors $\mathbb{E}[X|\mathcal{G}] = X$.*

Démonstration. Cela provient directement de l'unicité de $\mathbb{E}[X|\mathcal{G}]$. □

Proposition 54. *L'application $X \mapsto \mathbb{E}[X|\mathcal{G}]$ est linéaire.*

Démonstration. Soit X, Y deux variables aléatoires $L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $a \in \mathbb{R}$. Soit $Z \in L^1(\Omega, \mathcal{G}, \mathbb{P})$. Alors :

$$\begin{aligned}
\mathbb{E}[(a\mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}])Z] &= a\mathbb{E}[\mathbb{E}[X|\mathcal{G}]Z] + \mathbb{E}[\mathbb{E}[Y|\mathcal{G}]Z], \text{ par linéarité de l'espérance} \\
&= a\mathbb{E}[XZ] + \mathbb{E}[YZ], \text{ par définition} \\
&= \mathbb{E}[(aX + Y)Z]
\end{aligned}$$

Par unicité, $\mathbb{E}[aX + Y|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}]$. □

Proposition 55. *Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Alors :*

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

Démonstration. La fonction $\mathbf{1}$ est une variable aléatoire de $L^1(\Omega, \mathcal{G}, \mathbb{P})$. Alors :

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}] = \mathbb{E}[X\mathbf{1}] = \mathbb{E}[X].$$

□

Proposition 56. *Soit $X, X' \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ avec $X \geq X'$. Alors $\mathbb{E}[X|\mathcal{G}] \geq \mathbb{E}[X'|\mathcal{G}]$.*

Démonstration. La variable aléatoire $Y = X - X'$ est positive et est dans $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Donc par positivité de l'espérance conditionnelle $\mathbb{E}[X - X'|\mathcal{G}] \geq 0$. Enfin par linéarité de l'espérance conditionnelle, on a :

$$\mathbb{E}[X|\mathcal{G}] - \mathbb{E}[X'|\mathcal{G}] \geq 0.$$

□

Proposition 57. *Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Alors :*

$$|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}].$$

Démonstration. On sait que $-|X| \leq X \leq |X|$. On a alors par la propriété précédente :

$$\mathbb{E}[-|X||\mathcal{G}] \leq \mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[|X||\mathcal{G}].$$

Puis comme $\mathbb{E}[-|X||\mathcal{G}] = -\mathbb{E}[|X||\mathcal{G}]$, on obtient :

$$|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}].$$

□

8.3 Cas positif

Théorème 58. *Soit $X \in [0, \infty]$. Soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. Alors il existe une unique variable aléatoire positive qui vérifie :*

$$\forall Z \ \mathcal{G}\text{-mesurable positive, } \mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]].$$

Démonstration.

- **Existence :** On a que $X \wedge n$ est borné. On définit

$$\mathbb{E}[X|\mathcal{G}] = \lim \mathbb{E}[X \wedge n|\mathcal{G}].$$

La limite existe par le théorème précédent et par théorème de convergence monotone. Soit $Z \geq 0$ $\mathcal{G}$-mesurable. Alors :

$$\begin{aligned}\mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]] &= \mathbb{E}[Z \lim \mathbb{E}[X \wedge n|\mathcal{G}]] \\ &= \lim \mathbb{E}[(Z \wedge n)\mathbb{E}[X \wedge n|\mathcal{G}]] \\ &= \lim \mathbb{E}[(Z \wedge n)] \\ &= \mathbb{E}[ZX]\end{aligned}$$

- **Unicité :** Soit deux variables aléatoires Y, Z convenant. Soit $a, b \in \mathbb{Q}$. Posons $B_{a,b} = \{Y \leq a < b \leq Z\} \in \mathcal{G}$, puis $B = \cup_{a,b \in \mathbb{Q}} B_{a,b}$. Ainsi :

$$\mathbb{E}[(Z - Y)\mathbf{1}_B] = \mathbb{E}[X\mathbf{1}_B] - \mathbb{E}[X\mathbf{1}_B] = 0.$$

Donc $(Z - Y)\mathbf{1}_B = 0$. Donc $Z \leq Y$ p.s. Par symétrie, on obtient l'unicité.

□

Proposition 59. *Soit $X, Y \geq 0$ et $a \geq 0$. Alors :*

$$\mathbb{E}[aX + Y|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}].$$

Démonstration. Soit Z $\mathcal{G}$-mesurable positive. Alors :

$$\begin{aligned}\mathbb{E}[(a\mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}])Z] &= a\mathbb{E}[\mathbb{E}[X|\mathcal{G}]Z] + \mathbb{E}[\mathbb{E}[Y|\mathcal{G}]Z], \text{ par linéarité de l'espérance}\\ &= a\mathbb{E}[XZ] + \mathbb{E}[YZ], \text{ par définition}\\ &= \mathbb{E}[(aX+Y)Z]\end{aligned}$$

Par unicité, $\mathbb{E}[aX + Y|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}]$. □

Proposition 60. *Si X est $\mathcal{G}$ mesurable, alors $\mathbb{E}[X|\mathcal{G}] = X$.*

Démonstration. X est une variable aléatoire qui convient et par unicité de l'espérance conditionnelle on obtient l'égalité. □

8.4 Propriétés de l'espérance conditionnelle

Proposition 61. *(Théorème de convergence monotone)*
Soit X_n une suite de variables aléatoires positives croissante convergeant vers X. Alors $\mathbb{E}[X_n|\mathcal{G}]$ converge vers $\mathbb{E}[X|\mathcal{G}]$.

Démonstration. □

Proposition 62. *(Lemme de Fatou)*
Soit X_n une suite de variable aléatoire positive. Alors :

$$\mathbb{E}[\underline{\lim} X_n|\mathcal{G}] \leq \underline{\lim}\mathbb{E}[X_n|\mathcal{G}].$$

Démonstration. □

Proposition 63. *(Théorème de convergence dominée)*
Soit $X_n \in L^1$ avec l'existence de $Z \in L^1$ telle que $|X_n| \leq Z$. Alors :

$$\mathbb{E}[\lim X_n|\mathcal{G}] = \lim \mathbb{E}[X_n|\mathcal{G}].$$

Démonstration. □

Proposition 64. *(Inégalité de Jensen)*
Pour tout $\phi : \mathbb{R} \to \mathbb{R}^+$ convexe, et pour toute variable positive X :

$$\mathbb{E}[\phi(X)|\mathcal{G}] \geq \phi(\mathbb{E}[X|\mathcal{G}]).$$

Démonstration. Posons $E_\phi = \{(a,b) \in \mathbb{R}^2, \forall x \in \mathbb{R}, \phi(x) \geq ax + b\}$. Alors :

$$\phi(x) = \sup_{(x,y)\in E_\phi \cap \mathbb{Q}^2} (ax+b).$$

Ainsi :

$$\begin{aligned}\mathbb{E}[\phi(X)|\mathcal{G}] &= \mathbb{E}[\sup_{(x,y)\in E_\phi \cap \mathbb{Q}^2} (aX+b)|\mathcal{G}]\\ &\geq \sup_{(x,y)\in E_\phi \cap \mathbb{Q}^2} (a\mathbb{E}[X|\mathcal{G}] + b)\\ &\geq \phi(\mathbb{E}[X|\mathcal{G}])\end{aligned}$$

□

Proposition 65. *(Inégalité de Markov)*
Soit X positive et $t > 0$. Alors :

$$\mathbb{P}(X \geq t|\mathcal{G}) \leq \frac{1}{t}\mathbb{E}[X|\mathcal{G}].$$

Démonstration. On a $t\mathbf{1}_{X\geq t} \leq X$. Ainsi :

$$t\mathbb{P}(X \geq t|\mathcal{G}) = t\mathbb{E}[X \geq t|\mathcal{G}] \leq \mathbb{E}[X \geq t|\mathcal{G}].$$

□

Proposition 66. *Soit X une v.a.r. et Y $\mathcal{G}$-mesurable. Alors :*

$$\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}].$$

Démonstration. Soit Z $\mathcal{G}$-mesurable. Alors :

$$\mathbb{E}[\mathbb{E}[YX|\mathcal{G}]Z] = \mathbb{E}[ZYX] = \mathbb{E}[Z(Y\mathbb{E}[X|\mathcal{G}])].$$

Or $Y\mathbb{E}[X|\mathcal{G}]$ est $\mathcal{G}$-mesurable. Par unicité, on obtient l'égalité voulue. □

Proposition 67. *Soit $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$ et $X \geq 0$ intégrable. Alors :*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}_1]|\mathcal{G}_2] = \mathbb{E}[X|\mathcal{G}_1] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1].$$

Démonstration. Remarquons que $\mathbb{E}[X|\mathcal{G}_1]$ est $\mathcal{G}_2$-mesurable. Donc :

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}_1]|\mathcal{G}_2] = \mathbb{E}[X|\mathcal{G}_1]$$

Montrons l'autre égalité. Soit Y une variable aléatoire $\mathcal{G}_1$-mesurable donc $\mathcal{G}_2$-mesurable. Alors :

$$\mathbb{E}[Y\mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1]] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{G}_2]] = \mathbb{E}[YX].$$

Par unicité, on obtient l'autre égalité. □

Proposition 68. *Soit deux sous-tribus $\mathcal{G}_1$ et $\mathcal{G}_2$ de $\mathcal{F}$. Alors :*

$$\mathcal{G}_1 \sqcup \mathcal{G}_2 \textit{ ssi } \forall X\ \mathcal{G}_1\textit{-mesurable positive, } \mathbb{E}[X|\mathcal{G}_2] = \mathbb{E}[X].$$

Démonstration. Supposons qu'elles sont indépendantes. Alors considérons d'abord $X = \mathbf{1}_B$. Alors pour tout $A \in \mathcal{G}_2$:

$$\mathbb{E}[\mathbf{1}_B\mathbf{1}_A] = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{E}[\mathbb{E}[\mathbf{1}_B]\mathbf{1}_A].$$

Cette égalité reste vraie en remplaçant $\mathbf{1}_A$ par une variable aléatoire $\mathcal{G}_2$-mesurable, par densité. On a donc $\mathbb{E}[X|\mathcal{G}_2] = \mathbb{E}[X]$, pour les indicatrices puis par densité pour toutes les variables positives $\mathcal{G}_1$-mesurable.

Supposons réciproquement que pour tout $A \in \mathcal{G}_1$, $\mathbb{E}[\mathbf{1}_A|\mathcal{G}_2] = \mathbb{P}(A)$. Alors $\forall(A, B) \in \mathcal{G}_1 \times \mathcal{G}_2$:

$$\mathbb{P}(A \cap B) = \mathbb{E}[\mathbf{1}_A\mathbf{1}_B] = \mathbb{E}[\mathbb{E}[\mathbf{1}_A|\mathcal{G}_2]\mathbf{1}_B] = \mathbb{E}[\mathbb{P}(A)\mathbf{1}_B] = \mathbb{P}(A)\mathbb{P}(B).$$

□

8.5 Cas L^2

Théorème 69. *Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, alors $\mathbb{E}[X|\mathcal{G}]$ est la projection orthogonale de X sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$.*

Démonstration. Par l'inégalité de Jensen, $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]^2] \leq \mathbb{E}[X^2] < \infty$. Donc $\mathbb{E}[X|\mathcal{G}] \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. Soit $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. Montrons que $X - \mathbb{E}[X|\mathcal{G}]$ est orthogonal à Z.

Or pour tout Z borné, on a par définition de l'espérance conditionnelle la nullité du produit scalaire. Les fonctions bornées sont denses dans $L^2(\Omega, \mathcal{G}, \mathbb{P})$, donc :

$$\forall Z \in L^2(\Omega, \mathcal{G}, \mathbb{P}), \mathbb{E}[Z(X - \mathbb{E}[X|\mathcal{G}])] = 0.$$

□

9 Vecteurs Gaussiens

9.1 Cas 1D - Rappels

Définition. On dit que X est une variable aléatoire gaussienne de moyenne m et de variance σ^2 - ce qu'on note $X \sim \mathcal{N}(m, \sigma^2)$ - si X a pour densité :

$$x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

avec $m \in \mathbb{R}$ et $\sigma > 0$.

Proposition 70. *Pour une variable gaussienne $X \sim \mathcal{N}(m, \sigma^2)$, on a :*

$$\phi(t) = e^{itm} e^{-\frac{t^2\sigma^2}{2}}.$$

Démonstration. En effet, soit $t \in \mathbb{R}$. Alors :

$$\phi_(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$
$$= e^{itm} e^{-\frac{t^2\sigma^2}{2}}.$$

□

Proposition 71. *Si $X \sim \mathcal{N}(m, \sigma^2)$ et $Y \sim \mathcal{N}(m', \sigma'^2)$ sont indépendantes, alors :*

$$X + Y \sim \mathcal{N}(m + m', \sigma^2 + \sigma'^2).$$

Démonstration. Cela provient directement de $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$ et du théorème de Lévy. □

Proposition 72. *Si $X \sim \mathcal{N}(0, 1)$, alors $\mathbb{E}[X^{2n+1}] = 0$ et $\mathbb{E}[X^{2n}] = \frac{(2n)!}{2^n n!}$.*

Démonstration. La densité de X est paire. Donc $\mathbb{E}[X^{2n+1}] = \int_{\mathbb{R}} x^{2n+1}\phi(x)dx = 0$.

En fait, $\mathbb{E}[X^{n+1}] = n\mathbb{E}[X^{n-1}]$, par intégration par parties. On obtient donc la deuxième formule. □

9.2 Cas général

Définition. On dit qu'une famille de variables aléatoires $X = (X_1, ..., X_n)$ est un *vecteur gaussien* si pour tout $t \in \mathbb{R}^n$, la variable aléatoire $\langle t, X\rangle$ est gaussienne.

Remarque. Chaque composante d'un vecteur gaussien est une variable aléatoire gaussienne, mais elles ne sont pas forcément indépendantes.

Exemple. *Soit ε de loi $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ et $X \sim \mathcal{N}(0,1)$. Alors $(X, \varepsilon X)$ n'est pas un vecteur gaussien, pourtant ces composantes sont gaussiennes.*

Remarque. Via la proposition 71, si on a des variables aléatoires gaussiennes indépendantes $X_1, ..., X_n$, alors le vecteur associé est gaussien.

Définition. Soit X un vecteur gaussien. On note :
- $m = \mathbb{E}[X] \in \mathbb{R}^n$ son *vecteur moyenne.*
- $D = Cov(X) = (\mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j])_{i,j} \in M_n(\mathbb{R})$ sa *matrice de covariance.*

On note alors $X \sim \mathcal{N}(m, D)$.

Remarque. D est positive car $\langle t, Dt\rangle = Var(\langle t, X\rangle) \geq 0$.

Proposition 73. *Soit $X \sim \mathcal{N}(m, D), M \in \mathbb{R}^n$ et $A \in M_n(\mathbb{R})$.*
Alors $M + AX \sim \mathcal{N}(M + Am, ADA^t)$.

Démonstration. Le fait qu'il s'agisse d'un vecteur gaussien est évident par la définition même d'un vecteur gaussien. Montrons alors les formules de sa moyenne et de sa matrice de covariance. On pose $Y = M + AX$. Alors :
- $\mathbb{E}[M + AX]_i = M_i + \sum_j a_{i,j}\mathbb{E}[X_j] = (M + Am)_j$.
- Vérifions donc la matrice de covariance :

$$\begin{aligned}
Cov(Y_i, Y_j) &= \mathbb{E}[(M_i + (AX)_i)(M_j + (AX)_j)] - \mathbb{E}[(M_i + (AX)_i)]\mathbb{E}[(M_j + (AX)_j)] \\
&= M_iM_j + M_i\mathbb{E}[(AX)_j] + M_j\mathbb{E}[(AX)_i] + \mathbb{E}[(AX)_i(AX)_j] \\
&\quad - M_iM_j - M_i\mathbb{E}[(AX)_j] - M_j\mathbb{E}[(AX)_i] - \mathbb{E}[(AX)_i]\mathbb{E}[(AX)_j] \\
&= \sum_{k,l} a_{i,k}a_{j,l}\left(\mathbb{E}[X_kX_l] - \mathbb{E}[X_k]\mathbb{E}[X_l]\right) \\
&= \sum_{k,l} a_{i,k}a_{j,l}d_{k,l} \\
&= \sum_{k} a_{i,k}(DA^t)_{k,j} = (ADA^t)_{i,j}
\end{aligned}$$

□

Proposition 74. *Soit $m \in \mathbb{R}^n$ et $D \in M_n(\mathbb{R})$ positive. Alors il existe un vecteur gaussien tel que $Y \sim \mathcal{N}(m, D)$.*

Démonstration. Il existe $A = \sqrt{D}$ et on définit $U = (U_1, ..., U_n) \sim \mathcal{N}(0, Id_n)$. On pose alors $Y = m + AU$ un vecteur gaussien, qui convient par la propriété précédente. □

Proposition 75. *Soit $X \sim \mathcal{N}(m, D)$. Alors sa fonction caractéristique est :*

$$\phi_X(t) = \exp\left(-\frac{1}{2}\langle t, Dt\rangle + i\langle t, m\rangle\right).$$

Démonstration. La démonstration est basée sur l'indépendance des coordonnées du vecteur gaussien et de la fonction caractéristique d'une variable gaussienne. La rédaction est laissée au lecteur. □

Théorème 76. *Soit $X \sim \mathcal{N}(m, D)$. On a alors : D est diagonale ssi les X_i sont indépendantes.*

Démonstration. Le sens réciproque est évident. Montrons alors le sens direct.

On prend $Y \sim \mathcal{N}(m, D)$ avec $Y_i \sim \mathcal{N}(m_i, d_{ii})$ indépendantes. Alors :

$$\mathbb{E}[e^{i\langle t,Y\rangle}] = \prod e^{it_j m_j - t_j^2 d_{jj}/2} = e^{i\langle t,m\rangle - \frac{1}{2}\langle t,Dt\rangle},$$

qui est la fonction caractéristique de $\mathcal{N}(m, D)$. Par le théorème de Lévy, X et Y ont même loi. □

Proposition 77. *Soit $X \sim \mathcal{N}(m, D)$ et $\det D \neq 0$, alors X a pour densité :*

$$\frac{1}{\sqrt{(2\pi)^n \det D}} \exp\left(\frac{1}{2}\langle X - m, D^{-1}(X - m)\rangle\right) dx.$$

Démonstration. Soit $U \sim \mathcal{N}(0, I_n)$. On sait que $X = m + AU$ avec $A = \sqrt{D}$. Alors :

$$\begin{aligned}
\mathbb{E}[f(X)] = \mathbb{E}[f(m + AU)] &= \int_{\mathbb{R}^n} f(m + Aw) \prod_i \frac{1}{\sqrt{2\pi}} e^{\frac{-w_i^2}{2}} dw \\
&= \int_{\mathbb{R}^n} f(x) \frac{1}{\sqrt{2\pi}^n} e^{\frac{-1}{2}\langle A^{-1}(x-m), A^{-1}(x-m)\rangle} \det A^{-1} dx \\
&= \frac{1}{\sqrt{(2\pi)^n \det D}} \int_{\mathbb{R}^n} f(x) \exp\left(\frac{1}{2}\langle x - m, D^{-1}(x - m)\rangle\right) dx,
\end{aligned}$$

car A est symétrique. □

9.3 Conditionnement d'un vecteur gaussien

Théorème 78. *Soit $(X, Y_1, ..., Y_n)$ gaussien alors :*

- *pour tout $d \leq n$, il existe $a, b_1, ..., b_d \in \mathbb{R}$ tels que :*

$$\mathbb{E}[X|Y_1 + ... + Y_n] = a + \sum_i b_i Y_i.$$

- *$Z = X - \mathbb{E}[X|Y_1, ..., Y_d]$ est centrée et indépendante de $Y_1, ..., Y_d$.*
- *$X|(Y_1 = y_1, ..., Y_d = y_d)$ est gaussien.*

Démonstration. On travaille dans $H = Vect(1, Y_1, ..., Y_d) \subset L^2$ fermé de dimension finie. Soit $W = p_H(X)$. Donc $W = a + \sum_i b_i Y_i$.

De plus $Z = X - W \perp H$. Alors :

- $\mathbb{E}[Z] = \langle Z, 1\rangle = 0$. Donc Z centrée.
- $\mathbb{E}[ZY_i] = 0 = \mathbb{E}[Z]\mathbb{E}[Y_i]$. On sait que (Z, Y_i) est gaussien. Alors Z et Y_i sont indépendants.

De plus $\mathbb{E}[X|Y_1, ..., Y_d] = \mathbb{E}[Z|Y_1, ..., Y_d] + \mathbb{E}[W|Y_1, ..., Y_d] = 0 + (a + \sum_i b_i Y_i)$. Il ne reste donc plus qu'à démontrer le dernier point.

$$\begin{aligned}
\mathbb{E}[e^{itX}|Y_1 = y_1, ..., Y_d = y_d] &= \mathbb{E}[e^{it(Z+\mathbb{E}[X|Y=y])}|Y_1 = y_1, ..., Y_d = y_d] \\
&= e^{it\mathbb{E}[X|Y_1=y_1,...,Y_d=y_d]}\mathbb{E}[e^{itZ}] \\
&= e^{it\mathbb{E}[X|Y_1=y_1,...,Y_d=y_d]}e^{-\frac{t^2\mathbb{E}[Z^2]}{2}},
\end{aligned}$$

qui est la fonction caractéristique d'une variable aléatoire gaussienne. □

Exemple. *Soit* $X \sim \mathcal{N}\left[(0,0,0), \begin{pmatrix} 4 & 1 & 1 \\ 1 & 7 & 2 \\ 1 & 2 & 11 \end{pmatrix}\right].$

On cherche à calculer $Z = \mathbb{E}[X_1|X_2, X_3]$ qui est donc de la forme $a + b_2X_2 + b_3X_3$.

Or $\mathbb{E}[Z] = 0 = a$. De plus $\mathbb{E}[ZX_2] = \mathbb{E}[X_1X_2] = 1 = 7b_2 + 2b_3$ et $\mathbb{E}[ZX_3] = 1 = 2b_2 + 11b_3$.

On conclut en résolvant le système ainsi mis en évidence.