Chapitre 2

Les grands outils probabilistes

Dans ce chapitre, nous introduisons les notions d'espérance, variance, fonction caractéristique, et nous apprenons à les manipuler (calculs). Nous verrons leur utilité aux chapitres suivants.

2.1 Espérance, variance, écart-type

2.1.1 Définition

L'*espérance* d'une variable aléatoire $X : (\Omega, \mathcal{A}, \mathbf{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})$ ou $X : (\Omega, \mathcal{A}, \mathbf{P}) \to (\mathbb{N}, \mathcal{P}(\mathbb{N})$ est la moyenne de toutes ses valeurs possibles pondérées par la probabilité $\mathbf{P}$, i.e. la valeur qu'on peut espérer que X prenne.

Commençons par les variables aléatoires à valeurs entières.

Définition 2.1.1 *Si $X : (\Omega, \mathcal{A}, \mathbf{P}) \to (\mathbb{N}, \mathcal{P}(\mathbb{N})$ est une v.a. $\mathbf{P}$-intégrable, son* espérance *est définie par*

$$E(X) := \int_\Omega X(\omega) d\mathbf{P}(\omega) = \sum_{\mathbb{N}} k\, \mathbf{P}_X(k) = \sum_{\mathbb{N}} k\, \mathbf{P}(X = k)\,.$$

Si X est une v.a. positive, son espérance est encore définie comme ci-dessus, mais peut éventuellement être infinie. Si Ω est discret (cas fréquent), ceci se réécrit

$$E(X) := \sum_{\omega \in \Omega} X(\omega) \mathbf{P}(\omega) = \sum_{\mathbb{N}} k\, \mathbf{P}_X(k) = \sum_{\mathbb{N}} k\, \mathbf{P}(X = k)\,.$$

Un exemple : on lance deux dés équilibrés. $\Omega = \{1, \ldots, 6\}^2$ muni de la probabilité uniforme $\mathbf{P}$. On regarde la somme des deux chiffres : $X(i, j) = i + j$. Alors $E(X) = \ldots ... = 7$. (Exo).

Traitons maintenant le cas d'une v.a. à valeurs réelles.

Définition 2.1.2 *Si $X : (\Omega, \mathcal{A}, \mathbf{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})$ est une v.a. $\mathbf{P}$-intégrable, son* espérance *est définie par*

$$E(X) := \int_\Omega X(\omega) d\mathbf{P}(\omega) = \int_{\mathbb{R}} x\, d\mathbf{P}_X(x)\,.$$

Si X est une v.a. positive, son espérance est encore définie comme ci-dessus, mais peut éventuellement être infinie.

Exemple : $X : \Omega = [0, 1] \to \Omega$ définie par $X(\omega) = \omega$, $\mathbf{P} = \mathbf{P}_X$ loi uniforme sur $[0, 1]$. Alors $E(X) = \frac{1}{2}$.

L'espérance n'existe pas toujours, ou n'est pas forcément finie. Dans ce cas, on dit que X n'est pas intégrable. Dans le vocabulaire des analystes, l'espérance de X existe et est finie ssi $X \in L^1(\Omega, \mathbf{P})$. Notez aussi que $\|X\|_{L^1(\Omega)} = E(|X|)$.

Lorsque X est une v.a. positive, $E(X)$ a toujours un sens, mais peut être infinie.

Exercice 2.1.3 *Montrez qu'une v.a. X qui suit une loi de Cauchy sur $\mathbb{R}$ n'est pas intégrable. Son espérance n'est pas définie.*

Remarque 2.1.4 On parle d'espérance de v.a. car les v.a. sont ce qui intéresse le plus les probabilistes, mais en réalité, l'espérance est une quantité qui ne dépend que de la loi de X, et pas de la v.a. Autrement dit, deux v.a. de même loi ont même espérance, et on peut parler de la *valeur moyenne* d'une mesure de probabilité μ (définie comme $\int_{\mathbb{R}} x d\mu(x)$) plutôt que de l'espérance d'une v.a. X de loi $\mathbf{P}_X = \mu$.

Les propriétés élémentaires de l'espérance sont essentiiellement celles de l'intégrale, en ajoutant le fait qu'on intègre par rapport à une mesure de masse 1. Soient donc des v.a. à valeurs dans $\mathbb{R}$ ou $\mathbb{N}$, définies sur $(\Omega, \mathcal{A}, \mathbf{P})$.

1. $E(X + Y) = E(X) + E(Y)$
2. $E(\lambda X) = \lambda E(X)$
3. $E(\mathbf{1}_A) = \mathbf{P}(A)$
4. X intégrable ssi $|X|$ intégrable.
5. Si X intégrable et $\mathbf{P}_X(\mathbb{R}_+) = \mathbf{P}(X \geq 0) = 1$ alors $E(X) \geq 0$

Exercice 2.1.5 *Démontrer ces propriétés. Indication pour la dernière propriété : intégrer sur les ensembles $\{\omega \in \Omega,\ X(\omega) \geq 0\}$ et $\{\omega \in \Omega,\ X(\omega) < 0\}$.*

Reformulons la proposition 1.5.10.

Proposition 1.5.10 *La variable aléatoire réelle X a pour densité la fonction mesurable positive f si et seulement si pour toute fonction $h : \mathbb{R} \to \mathbb{R}$ mesurable bornée, on a*

$$E(h(X)) = \int_{\mathbb{R}} h(x) f(x) d\lambda(x)$$

2.1.2 Exemples

v.a. discrètes

Si X suit une loi uniforme sur $\{0, 1\}$, $E(X) = \frac{1}{2}$.
Si X suit une loi de Bernoulli $\mathcal{B}(p)$ alors $E(X) = p$.
Si X suit une loi binomiale $\mathcal{B}(n, p)$, alors $E(X) = np$.
Si X suit une loi de Poisson $\mathcal{P}(\lambda)$ alors $E(X) = \lambda$.
Si X suit une loi géométrique $(\mathbf{P}(X = k) = p^k(1 - p))$ alors $E(X) = \frac{p}{1-p}$.

v.a. continues

Si X suit une loi uniforme sur $[a, b]$, alors $E(X) = \frac{a+b}{2}$.
Si X suit une loi normale centrée réduite $E(X) = 0$
Si X suit une loi normale de moyenne m $E(X) = m$.
Si X suit une loi exponentielle de paramètre a $E(X) = \frac{1}{a}$.
Si X suit une loi de Cauchy, alors X n'est pas intégrable ($E(X)$ n'est pas définie.)

NB la loi exp modélise les durées de vie : $\frac{1}{a}$ représente la durée de vie de la particule considérée.

2.1.3 Variance, écart-type

Les quantités introduites dans ce paragraphe servent beaucoup en statistiques.

Définition 2.1.6 *Soit X une v.a. intégrable. à valeurs entières ou réelles. Sa* variance *est la quantité $Var(X) = E\left((X - E(X))^2\right)$. Lorsque X est intégrable, la variance de X existe toujours dans $\mathbb{R}_+ \cup \{+\infty\}$. L'*écart-type *de X est $\sigma(X) = \sqrt{var(X)}$.*

Justifier si nécessaire cette définition.

Remarque 2.1.7 Si X est une v.a.r. de carré intégrable, alors par l'inégalité de Cauchy-Schwartz, on déduit $E(|X|) \leq \sqrt{E(X^2)}\sqrt{E(1)}$. Or $E(1) = \mathbf{P}_X(\mathbb{R}) = 1$. Donc $E(X^2) < \infty$ implique $E(|X|) < \infty$, i.e. X intégrable (et aussi bien sûr $\mathrm{Var}(X) < \infty$). Autrement dit, en théorie des probabilités, on a l'inclusion $L^2(\Omega, \mathcal{A}, \mathbf{P}) \subset L^1(\Omega, \mathcal{A}, \mathbf{P})$. Cette inclusion bien pratique n'est pas vérifiée en analyse sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, ni plus généralement sur les espaces mesurés de mesure infinie.

Cette inclusion est stricte. En effet, il existe des exemples de variables aléatoires intégrables mais dont la variance est infinie.

La variance représente la moyenne des carrés des écarts à la moyenne. On mesure ainsi la façon dont X prend des valeurs lointaines ou proches de la moyenne $E(X)$, souvent ou non. On aurait pu s'intéresser à la quantité $E(|X-E(X)|)$ a priori, mais les carrés sont plus aisés à manipuler dans les calculs que les $|.|$.

On vérifie que

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

Proposition 2.1.8 *$Var(X) = 0$ ssi $X = E(X)$ p.s. ssi X est constante p.s.*

Démonstration : Exo □

Variance des lois classiques

Exercice 2.1.9 *Calculer les variances de v.a. suivant les lois classiques introduites plus haut. Ce type d'exercice calculatoire fera à coup sûr l'objet de questions considérées (par moi) comme très faciles le jour du partiel ou de l'examen.*

Indication : pour certaines lois, il est utile de calculer d'abord $E(X(X-1))$ puis $\mathrm{Var}(X) = E(X(X-1) + E(X) - (E(X))^2$

X suit une loi uniforme sur $\{0,1\}$: $\mathrm{Var}(X) = \frac{1}{4}$
X suit une loi $\mathcal{B}(p)$ sur $\{0,1\}$: $\mathrm{Var}(X) = p(1-p)$.
X suit une loi binomiale $\mathcal{B}(n,p)$: $\mathrm{Var}(X) = np(1-p)$.
X suit une loi de Poisson $\mathcal{P}(\lambda)$:$\mathrm{Var}(X) = \lambda$.
X suit une loi géométrique $\mathcal{G}(\alpha)$: $\mathrm{Var}(X) = \frac{p}{(1-p)^2}$
X suit une loi uniforme sur l'intervalle $[a,b]$: $\mathrm{Var}(X) = \frac{(a-b)^2}{12}$
X suit une loi normale $\mathcal{N}(m,\sigma^2)$: $\mathrm{Var}(X) = \sigma^2$
X suit une loi exponentielle de paramètre a : $\mathrm{Var}(X) = \frac{1}{a^2}$

2.1.4 Corrélation de deux variables aléatoires

Définition 2.1.10 *Soient X et Y deux v.a. réelles intégrables. La* covariance *de X et Y, ou coefficient de corrélation, est définie par*

$$Cov(X,Y) := E\left((X-E(X))(Y-E(Y)\right) = E(XY) - E(X)E(Y)$$

Vérifiez l'égalité de droite.
Lorsque Cov(X,Y) = 0, on dit que X *et* Y *sont non corrélées.*

Lemme 2.1.11 *Si X et Y sont indépendantes, alors elles sont non corrélées.*

Démonstration : Voir chapitre 3. □

Remarque 2.1.12 La réciproque est fausse. Voir ch 3

2.1.5 Moments

Définition 2.1.13 *On dit que X admet un moment d'ordre $k \geq 1$, si $|X|^k$ est intégrable, soit encore $E(|X|^k) < \infty$. Son moment d'ordre k est alors $E(X^k)$.*

En analyse, on dirait que X est dans $L^k(\Omega, \mathbf{P})$.

2.1.6 Inégalités classiques

Les inégalités de ce paragraphe sont fondamentales et élémentaires. Les preuves sont à retenir au moins autant, si ce n'est plus, que les énoncés.

Lemme 2.1.14 (Inégalité de Markov) *Soient X une v.a. réelle et $a > 0$ un réel positif. Alors pour tout réel $p > 0$ on a*

$$P(|X| \geq a) \leq \frac{E(|X|^p)}{a^p}.$$

Démonstration : $E(|X|) = \int_\Omega |X| d\mathbf{P} = \int_{\{|X|\geq a\}} |X|\, d\mathbf{P} + \int_{\{|X|<a\}} |X|\, d\mathbf{P}$. Or vu l'ensemble sur lequel on intègre, on a $\int_{\{|X|\geq a\}} |X|\, d\mathbf{P} \geq a\mathbf{P}(|X| \geq a)$. Et le dernier terme $\int_{\{|X|<a\}} |X|\, d\mathbf{P}$ est positif. Autrement dit on a $E(|X|) \geq a\mathbf{P}(|X| \geq a)$.

Pour l'inégalité avec p, le raisonnement est identique (exo). □

Lemme 2.1.15 (Inégalité de Bienaymé-Tchebytchev) *Soit X une v.a.r. de carré X^2 intégrable. Alors pour tout $a > 0$, on a*

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$$

Démonstration : Appliquer l'inégalité de Markov à $Y = X - E(X)$ avec $p = 2$. □

Cette inégalité illustre bien comment la variance d'une variable aléatoire permet de contrôler les écarts à la moyenne de la v.a.

2.2 Fonctions génératrices

Dans cette section, on présente un outil élémentaire très utile dans l'étude des v.a. à valeurs dans $\mathbb{N}$. On verra une généralisation aux v.a. r. au paragraphe suivant.

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Remarquons que par définition de l'espérance, on a

$$E(s^X) = \sum_{n=0}^{\infty} \mathbf{P}(X=n)s^n \quad \text{pour tout} \quad s \geq 0$$

Définition 2.2.1 *La fonction génératrice de X est la fonction g définie par*

$$g_x(s) = \sum_{n=0}^{\infty} \mathbf{P}(X=n)s^n = E(s^X) \quad \textit{pour tout } 0 \leq s \leq 1$$

Posons $q_n = \mathbf{P}(X = n)$. On a alors $q_0 = g_X(0)$, $g_X(1) = 1$, et $\lim_{s\to 1} g_X(s) = g_X(1) = 1$ par convergence monotone. On vérifie que sur $[0,1]$, g_X est convexe. La série entière g_X a un rayon de convergence $R \geq 1$, puisque $g_X(1) = 1$. Donc g_X est C^∞ sur $[0,1[$. Ses dérivées successives vérifient $g_X'(s) = \sum_{n=1}^\infty nq_n s^{n-1}$, $g_X''(s) = \sum_{n=2}^\infty n(n-1)q_n s^{n-2}$ ... On en déduit aisément

Proposition 2.2.2 *La fonction génératrice vérifie pour tout $n \geq 0$, $\mathbf{P}(X=n) = \frac{1}{n!}g_X^{(n)}(0)$. En particulier, la fonction génératrice d'une variable aléatoire détermine la loi de X.*

Donnons des exemples de calculs de fonctions génératrices.

Loi binomiale $B(n,p)$.

$$g(s) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s^k = (ps + (1-p))^n$$

Loi de Poisson $\mathcal{P}(\lambda)$

$$g(s) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k s^k}{k!} = e^{\lambda(s-1)}$$

Loi géométrique $\mathcal{G}(a)$

$$g(s) = \sum_{k=0}^{\infty} (1-a)a^k s^k = \frac{1-a}{1-as}$$

Calcul des moments

Rappelons que $E(|X|^p) < \infty$ implique $E(|X|^q) < \infty$ pour tout $q \leq p$. (Pourquoi ? Exo !) Notons aussi qu'une v.a. à valeurs entières est positive !

Proposition 2.2.3 *Soit X une v.a. à valeurs dans $\mathbb{N}$.*
- *$E(X) < \infty$ ssi g_X est dérivable à gauche en 1, et dans ce cas, $E(X) = g'_X(1)$.*
- *$E(X^2) < \infty$ ssi g_X est deux fois dérivable à gauche en 1, et dans ce cas, $E(X(X-1)) = g''_X(1)$.*

Démonstration : voir TD ou ouvrages classiques □

Exercice 2.2.4 *Calculer espérance et variance de v.a. de loi binomiale, Poisson, ou géométrique en utilisant la proposition ci-dessus.*

Remarque 2.2.5 Un grand intérêt des fonctions génératrices est que la fonction génératrice d'une somme de variables aléatoires indépendantes est le produit des fonctions génératrices. Ceci sera détaillé dans le cas des fonctions caractéristiques, plus générales.

2.3 Fonctions caractéristiques

Intégrale de fonctions à valeurs complexes

Vous avez vu en intégration la définition et les propriétés de l'intégrale de fonctions à valeurs réelles. Mais tout s'étend sans aucun problème aux fonctions à valeurs complexes.

Soit $f : \Omega \to \mathbb{C}$ une fonction qui s'écrit donc $f = u + iv$, avec u et v deux fonctions de Ω dans $\mathbb{R}$. Si μ est une mesure sur Ω, on dit que f est intégrable pour μ si $\int_\Omega |f(\omega)|\, d\mu < +\infty$. Remarquons que f est intégrable ssi $\mathrm{R}ef$ et $\mathrm{I}mf$ sont intégrables, puisque

$$\max(|\mathrm{R}ef|, |\mathrm{I}mf|) \leq |f| \leq |\mathrm{R}ef| + |\mathrm{I}mf|$$

On définit alors l'intégrale de f par

$$\int_\Omega f\, d\mu = \int u\, d\mu + i \int v\, d\mu\,.$$

Cette intégrale est un nombre complexe.

Les calculs se font... comme on pense !

$$\int_0^1 e^{iax} dx = \int_0^1 \cos ax + i \sin ax dx = \left[\frac{e^{iax}}{ia}\right]_0^1 = \frac{e^{ia} - 1}{ia}.$$

Si $f : \mathbb{R} \to \mathbb{C}$ est continue, alors $u = \mathrm{Re}f$ et $v = \mathrm{Im}f$ sont aussi continues. Elles ont donc des primitives, qu'on note U et V. Alors $F = U + iV$ est une primitive de f. Et bien sûr $\int_a^b f(x)dx = \int_a^b u(x)dx + i\int_a^b v(x)dx = [U + iV]_a^b = [F]_a^b = F(b) - F(a)$.

2.3.1 Fonction caractéristique d'une variable aléatoire réelle

Définition 2.3.1 *Soit μ une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (i.e. $\mu(\mathbb{R}) < +\infty$). Alors la* transformée de Fourier de la mesure μ *est la fonction $\hat{\mu} : \mathbb{R} \to \mathbb{C}$ définie pour tout $t \in \mathbb{R}$ par*

$$\hat{\mu}(t) = \int_R e^{itx} d\mu(x).$$

La fonction $\hat{\mu}$ est bien définie car $x \to e^{itx}$ est continue donc mesurable, et bornée donc intégrable (car μ est supposée de masse totale finie).

Définition 2.3.2 *Soit $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une v.a. réelle. Sa* fonction caractéristique, *notée φ_X, est la transformée de Fourier de la loi P_X de X. Autrement dit pour tout $t \in \mathbb{R}$*

$$\varphi_X(t) = \hat{P}_X(t) = \int_\mathbb{R} e^{itx} dP_X(x) = E(e^{itX})\,.$$

Remarquez que la convention est légèrement différente de celle utilisée pour la transformée de Fourier de fonctions en analyse.

Remarque 2.3.3 Si X est une v.a. à densité f, alors sa fonction caractéristique n'est rien d'autre, au signe près, que la transformée de Fourier de f : $\varphi_X(t) = \hat{f}(-t)$.

Proposition 2.3.4 *Soit X une v.a. réelle.*
1. *$\varphi_X(0) = 1$*
2. *$|\varphi_X(t)| \leq 1$ pour tout $t \in \mathbb{R}$.*
3. *$\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$*
4. *φ_X est à valeurs réelles si et seulement si $P_X = P_{-X}$.*
5. *φ_X est uniformément continue.*
6. *Soit $p \geq 1$. Si $E(|X|^p) < +\infty$, alors φ_X est de classe C^p et $\varphi_X^{(k)}(0) = i^k E(X^k)$ pour tout $1 \leq k \leq p$.*

Démonstration : $*$ D'abord $\varphi_X(0) = E(1) = 1$.
$*$ $|\varphi_X(t)| \leq \int_\mathbb{R} |e^{itx}| dP_X(x) = 1$.
$*$ Si $Y = aX + b$, avec $a \neq 0$ et $b \in \mathbb{R}$,

$$\varphi_Y(t) = E(e^{itY}) = E(e^{itaX+itb}) = e^{itb}E(e^{itaX}) = e^{itb}\varphi_{aX}(t) = e^{itb}\varphi_X(at).$$

$*$ $\overline{\varphi_X(t)} = \overline{E(e^{itX})} = E(e^{-itX}) = E(e^{it(-X)}) = \varphi_{-X}(t)$. Par conséquent, si $\overline{\varphi_X} = \varphi_X$, on a $\varphi_{-X} = \varphi_X$, d'où $P_X = P_{-X}$ (d'après le théorème 2.3.6 ci-dessous). Réciproquement si $P_X = P_{-X}$ alors bien sûr $\varphi_X = \varphi_{-X} = \overline{\varphi_X}$.
$*$ La fonction φ_X est uniformément continue si pour tout $\varepsilon > 0$ il existe $\alpha > 0$ tq pour tous s, t, si $|t - s| < \alpha$ alors $|\varphi_X(t) - \varphi_X(s)| \leq \varepsilon$. Calculons donc

$$|\varphi_X(t) - \varphi_X(s)| \leq \int_\mathbb{R} |e^{itx} - e^{isx}| dP_X(x) = \int |e^{i(t-s)x} - 1| dP_X(x)\,.$$

Quand $u \to 0$, $|e^{iux} - 1| \to 0$. Donc $|e^{i(t-s)x} - 1| \to 0$ quand $|s - t| \to 0$, uniformément en s et t. Vous pouvez détailler le "passage aux ε". En intégrant par rapport à la probabilité P_X, le résultat en découle.
$*$ Soit $p \geq 1$ et supposons que $E(|X|^p) < \infty$. On a $\frac{d}{dt}e^{itx} = ixe^{itx}$ d'où $|\frac{d}{dt}e^{itx}| = |x|$. Or $|x| \leq \max(1, |x|^p)$, donc l'hypothèse nous dit que $x \to |x|$ est intégrable. Le théorème de dérivation sous le signe $\int$ implique que $\varphi_X(t)$ est dérivable, de dérivée $\varphi'_X(t) = E(iXe^{itX})$. Pour les dérivées d'ordre supérieur, le raisonnement est le même. Par récurrence, on obtient $\varphi_X^{(k)}(t) = i^k E(X^k e^{itX})$. En particulier, $\varphi_X^{(k)}(0) = i^k E(X^k)$. □

Rappelons cette proposition classique utile pour les lois à densité.

Proposition 2.3.5 (Riemann-Lebesgue) *Si $f : \mathbb{R} \to \mathbb{R}$ est intégrable, alors sa transformée de Fourier $\mathcal{F}f$ tend vers 0 à l'infini.*

En particulier, si f est la densité de la loi P_X d'une variable aléatoire, alors $\lim_{t\to\pm\infty} \varphi_X(t) = \lim_{t\to\mp\infty} \hat{f}(t) = 0$.

Démonstration : Ceci a été fait en intégration. Principe de la preuve : on vérifie d'abord que c'est vrai pour les fonctions indicatrices d'intervalles bornés. Par linéarité, c'est vrai pour les fonctions en escalier, qui sont denses dans l'ensemble des fonctions intégrables. On en déduit le résultat par densité (à justifier proprement). □

Exemples (à détailler en exercice)

Si X est une v.a.r. qui suit une loi de Bernoulli $\mathcal{B}(p)$, alors $\varphi_X(t) = E(e^{itX}) = pe^{it} + 1 - p$.

Si X est une v.a. uniforme sur $[-a,a]$ alors

$$\varphi_X(t) = \int_{-a}^{a} \frac{e^{itx}}{2a} dx = \frac{1}{a}\left[\frac{e^{itx}}{it}\right]_{-a}^{a} = \frac{\sin ta}{ta}.$$

Si X est une v.a. dont la loi est une loi normale centrée réduite, elle a pour densité $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. On a donc

$$\varphi_X(t) = \int_{\mathbb{R}} \frac{e^{itx - x^2/2}}{\sqrt{2\pi}} dx .$$

Comment calculer cette intégrale ? D'abord on remarque que $\frac{d}{dt} e^{itx-x^2/2}\frac{1}{\sqrt{2\pi}} = ixe^{itx-x^2/2}\frac{1}{\sqrt{2\pi}}$ est intégrable. La proposition précédente nous permet de déduire que $\varphi'_X(t) = \int_{\mathbb{R}} \frac{ixe^{itx-x^2/2}}{\sqrt{2\pi}} dx$. Par intégration par parties, on obtient (détaillez) $\varphi'_X(t) = -t\varphi_X(t)$ et $\varphi_X(0) = 1$. Cette équation différentielle (cf cours sur les équations différentielles) a pour unique solution

$$\varphi_X : t \mapsto \varphi_X(t) = e^{-t^2/2}.$$

Si X est une v.a. dont la loi est une loi normale de moyenne m et de variance σ^2, sa loi P_X a pour densité $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2}$. Un changement de variable nous ramène au cas précédent, et on trouve

$$\varphi_X(t) = \exp(itm - \frac{\sigma^2 t^2}{2}) .$$

Autre méthode (Revuz [?]).

Si X est une v.a. de loi $\mathcal{N}(m,\sigma^2)$, alors $Y = X - m$ a pour loi une loi normale centrée $\mathcal{N}(0,\sigma^2)$. (démonstration immédiate à l'aide des fonctions de répartition ou du changement de variable $y = x - m$). En particulier, ce changement de variable donne immédiatement $\varphi_X(t) = e^{itm}E(e^{itY})$. Il nous reste à calculer $E(e^{itY})$. Si $z \in \mathbb{R}$, alors un changement de variable élémentaire donne

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{zx} e^{-x^2/\sigma^2} dx = \exp(\sigma^2 z^2/2) .$$

Maintenant, si $z \in \mathbb{C}$, on a $|e^{zx}| \leq e^{|z||x|}$. On en déduit (détailler un peu) que l'intégrale $\frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{zx} e^{-x^2/\sigma^2} dx$ est bien définie et holomorphe en $z \in \mathbb{C}$. Cette fonction de $z \in \mathbb{C}$ coïncide sur l'axe réel avec la fonction holomorphe $z \mapsto \exp(\sigma^2 z^2/2)$. Le théorème de prolongement analytique assure qu'elles sont égales sur $\mathbb{C}$. Prendre $z = it$ donne alors le résultat voulu.

Théorème 2.3.6 *Soient μ et ν deux mesures de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si $\hat{\mu} = \hat{\nu}$, alors $\mu = \nu$. Soient X et Y deux v.a. réelles. Alors leurs lois P_X et P_Y sont égales si et seulement si $\varphi_X = \varphi_Y$.*

Ce théorème nous donne l'unicité d'une loi de fonction caractéristique donnée. Autrement dit la fonction caractéristique d'une v.a. caractérise la loi de la v.a. Attention, je vous rappelle qu'il existe plein de v.a. qui n'ont rien à voir entre elles (pas le même espace Ω de départ par exemple), et qui ont même loi.

Ce théorème peut se démontrer de diverses façons. Il découle par exemple du lemme et du théorème ci-dessous (cf Revuz Probas).

Introduisons une notation. Si m est une mesure de probabilité sur $\mathbb{R}$, pour tous $a < b$, on pose

$$H_m(a,b) = m(]a,b[) + \frac{m(\{a\}) + m(\{b\})}{2} .$$

Lemme 2.3.7 *Soient m_1 et m_2 deux probabilités sur $\mathbb{R}$ telles que $H_{m_1} = H_{m_2}$. Alors $m_1 = m_2$.*

Démonstration : m_1 et m_2 étant finies, elles ont au plus un nombre dénombrable d'atomes (i.e. de singletons de mesure $m_i(\{x\}) > 0$. Considérons la classe des intervalles (ouverts, fermés, semi-ouverts) dont les extrémités ne sont pas des atomes ni de m_1 ni de m_2. Cette classe est stable par intersection finie, et engendre la tribu. Le corollaire 1.2.4 donne le résultat voulu. □

Théorème 2.3.8 (Inversion de Fourier) *Soit m une mesure de probabilité sur $\mathbb{R}$. Alors pour tout $a < b$, on a*

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \hat{m}(t) = H_m(a,b) .$$

Si $\hat{m}$ est intégrable, alors m a une densité continue p donnée par

$$p(x) = \frac{1}{2\pi} \int \hat{m}(t) e^{-itx} dt .$$

Démonstration : Le théorème de Fubini et un calcul élémentaire (changement de variable et parité) montrent que l'intégrale de gauche vaut

$$\int_{\mathbb{R}} \left(\int_{-T(x-a)}^{T(x-a)} \frac{\sin t}{t} dt - \int_{-T(x-b)}^{T(x-b)} \frac{\sin t}{t} dt \right) dm(x)$$

A noter qu'il s'agit d'intégrales au sens de Riemann : Si $x < a$, alors $\int_{-T(x-a)}^{T(x-a)} \cdots = -\int_{[-T|x-a|, T|x-a|]} \cdots$

La fonction $x \mapsto \int_0^x \frac{\sin t}{t} dt$ est bornée et tend vers $\pi/2$ en $+\infty$. Donc, après étude du signe de $x - a$ et de $x - b$, on vérifie que la fonction dans les $(\ldots)$ tend vers $\pi - \pi = 0$ si $x > \max(a,b)$, vers $\pi + \pi = 2\pi$ si $a < x < b$, vers $0 = -\pi + \pi$ si $x < a$, vers π si $x = a$ ou $x = b$. Autrement dit, la fonction de x entre parenthès converge vers $\pi(\mathbf{1}_{\{a\}} + \mathbf{1}_{\{b\}} + 2\mathbf{1}_{]a,b[})$, en restant uniformément bornée. Le théorème de convergence dominée de Lebesgue s'applique alors pour donner le résultat de convergence voulu.

Montrons maintenant la deuxième partie du théorème. e Le théorème de Fubini permet de voir que l'intégrale de l'énoncé vaut encore

$$\int_a^b \left(\int_{-T}^{T} e^{-itx} \hat{m}(t)\, dt \right) dx$$

Si $\hat{m}$ est intégrable, lorsque $T \to \infty$, l'intégrale entre parenthèses converge vers $\int_{\mathbb{R}} e^{-itx}\hat{m}(t)dt$, en restant majorée par $\|\hat{m}\|_1$. Le théorème de convergence dominée de Lebesgue donne alors, à la limite, $H_m(a,b) = \frac{1}{2\pi}\int_a^b \int_{\mathbb{R}} e^{-itx}\hat{m}(t)dtdx$, ce qu'il fallait démontrer. □

Corollaire 2.3.9 *Soit X une variable aléatoire. Si sa fonction caractéristique φ_X est intégrable alors X est à densité (continue).*

Démonstration : La deuxième partie du théorème donne l'implication : si φ_X est intégrable, alors X a une densité (continue). □

Remarque 2.3.10 Une variable aléatoire à densité n'a pas forcément une fonction caractéristique intégrable.

2.3.2 Fonction caractéristique d'une variable aléatoire vectorielle

Rappelons que si x et y sont deux vecteurs de $\mathbb{R}^d$, leur produit scalaire, noté de façons très variées, vérifie

$$< x, y >= (x|y) = x.y = x_1 y_1 + \ldots + x_d y_d .$$

Définition 2.3.11 *Soit μ une mesure finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Sa transformée de Fourier $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ est définie pour tout $u \in \mathbb{R}^d$ par*

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{iu.x} d\mu(x).$$

Si X est une v.a. vectorielle à valeurs dans $\mathbb{R}^d$, sa fonction caractéristique *$\varphi_X : \mathbb{R}^d \to \mathbb{C}$ est la transformée de Fourier de sa loi P_X.*

Soient μ une mesure sur $\mathbb{R}^p$ et ν une mesure sur $\mathbb{R}^q$. Remarquez que

$$\widehat{(\mu \times \nu)} = \hat{\mu}\hat{\nu}.$$

En effet, en utilisant le théorème de Fubini, on a pour $u = (u_1, u_2)$ avec $u_1 \in \mathbb{R}^p$ et $u_2 \in \mathbb{R}^q$

$$\begin{aligned} \widehat{(\mu \times \nu)}(u) &= \int_{\mathbb{R}^{p+q}} e^{iu.x} d(\mu \times \nu)(x_1, x_2) = \int_{\mathbb{R}^p} \int_{\mathbb{R}^q} e^{iu_1 x_1} e^{iu_2 x_2} d\nu(x_2) d\mu(x_1) \\ &= \int_{\mathbb{R}^p} e^{iu_1 x_1} d\mu(x_1) \int_{\mathbb{R}^q} e^{iu_2 x_2} d\nu(x_2) = \hat{\mu}(u_1)\hat{\nu}(u_2) \end{aligned}$$

Ceci sera très important par la suite.

Théorème 2.3.12 *Soient μ et ν deux mesures de probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Si $\hat{\mu} = \hat{\nu}$, alors $\mu = \nu$. Soient X et Y deux v.a. à valeurs dans $\mathbb{R}^d$. Alors leurs lois P_X et P_Y sont égales si et seulement si $\varphi_X = \varphi_Y$.*

Démonstration : Admis. Voir Master I, ou Revuz. □

Exemple de calcul d'une fonction caractéristique vectorielle : $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$ et P = Lebesgue sur $[0, 1]$. Soit $X : \Omega \to \mathbb{R}^2$ définie par $X(\omega) = (1_{[0,1/2]}(\omega), \omega)$. Notons $X_1(\omega) = 1_{[0,1/2]}(\omega)$ et $X_2(\omega) = \omega$ ses coordonnées. Si $u = (u_1, u_2) \in \mathbb{R}^2$, on calcule (détaillez)

$$\begin{aligned} \varphi_X(u) &= E(e^{iu.X}) = E(e^{i(u_1,u_2).(X_1(\omega),X_2(\omega))}) = E(e^{iu_1 X_1(\omega) + iu_2 X_2(\omega)}) \\ &= \int_\Omega e^{iu_1 X_1(\omega) + iu_2 X_2(\omega)} \, dP(\omega) = \ldots = \frac{e^{iu_1}}{iu_2}(e^{iu_2/2} - 1) + \frac{e^{iu_2} - e^{iu_2/2}}{iu_2} \end{aligned}$$

Proposition 2.3.13 *Soit X une v.a. à valeurs dans $\mathbb{R}^d$. Alors*

1. $\varphi_X(0) = 1$
2. $|\varphi_X(u)| \leq 1$ *pour tout* $u \in \mathbb{R}^d$
3. *Si* $A \in M_{d,n}(\mathbb{R})$ *et* $b \in \mathbb{R}^n$ *alors* $\varphi_{A.X+b}(v) = e^{iv\dot{b}} \varphi_X({}^tAv)$ *pour* $v \in \mathbb{R}^n$.
4. φ_X *est uniformément continue*
5. *Si* $n = (n_1, \ldots, n_d)$ *et* $E(|X_1|^{n_1} \ldots |X_d|^{n_d}) < +\infty$ *alors* $\frac{\partial^n \varphi_X(0)}{\partial^n u} = i^{|n|} E(X_1^{n_1} \ldots X_d^{n_d})$
6. *Si* $\hat{\mu} = \hat{\nu}$ *alors* $\mu = \nu$ *et* X *et* Y *ont même loi si et seulement si* $\varphi_X = \varphi_Y$.

Démonstration : Démontrons seulement 3). Si $Y = A.X + b \in \mathbb{R}^n$ et $v \in \mathbb{R}^n$, alors $\varphi_Y(v) = E(e^{iv.Y}) = E(e^{iv.(AX+b)}) = e^{iv.b} E(e^{iv.AX}) = e^{iv.b} E(e^{i(A^t v).X}) = e^{iv.b} \varphi_X(A^t v)$. □

Théorème 2.3.14 (Théorème de Levy) *Soit $(P_n)_{n \in \mathbb{N}}$ une suite de mesures de probabilité sur $\mathbb{R}^d$ telles que $\widehat{P}_n$ converge simplement vers une fonction ϕ continue en 0. Alors la suite $(P_n)_{n \in \mathbb{N}}$ converge étroitement vers une mesure m telle que $\hat{m} = \phi$.*

Ce théorème est du niveau M1. Nous commenterons cet énoncé à nouveau lorsque nous aurons vu ce que signifie la convergence étroite.

Chapitre 3

Indépendance

3.1 Produit de lois

Ce paragraphe a dû être traité en détail dans le module intégration 2.

Soient $(X_1, \mathcal{A}_1, \mathcal{P}_1)$ et $(X_2, \mathcal{A}_2, \mathcal{P}_2)$ deux espaces de probabilité.

La *tribu produit* sur $X_1 \times X_2$ est la tribu engendrée par les ensembles $A_1 \times A_2$, $A_i \in \mathcal{A}_i$. En particulier, elle contient $\mathcal{A}_1 \times \mathcal{A}_2$, mais elle est plus grosse. Elle est notée $\mathcal{A}_1 \otimes \mathcal{A}_2$. Par exemple, la tribu borélienne $\mathcal{B}(\mathbb{R}^2)$ est également la tribu produit $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

La *mesure produit* $\mathbf{P}_1 \otimes \mathbf{P}_2$ est l'unique mesure qui vérifie pour tous $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$

$$\mathbf{P}_1 \otimes \mathbf{P}_2(A_1 \times A_2) = \mathbf{P}_1(A_1)\mathbf{P}_2(A_2)\,.$$

$\mathbf{P}_1$ et $\mathbf{P}_2$ sont les mesures marginales de $\mathbf{P}_1 \otimes \mathbf{P}_2$. L'existence et l'unicité de la mesure produit (i.e. le fait qu'elle peut être prolongée, et de manière unique, en une mesure sur tous les ensembles de la tribu produit, sont admis ici. Voir cours intégration 2.

Par définition de $\mathbf{P}_1 \otimes \mathbf{P}_2$, si $g(x, y) = \mathbf{1}_{A_1 \times A_2}(x, y) = \mathbf{1}_{A_1}(x)\mathbf{1}_{A_2}(y)$ alors l'application $x \mapsto \int_\mathbb{R} g(x, y) d\mathbf{P}_2(y) = \mathbf{1}_{A_1}(x)\mathbf{P}_2(A_2)$ est mesurable, l'application $y \mapsto \int_\mathbb{R} g(x, y) d\mathbf{P}_1(x) = \mathbf{1}_{A_2}(y)\mathbf{P}_1(A_1)$ est mesurable, et on a

$$\int_{\mathbb{R}^2} g d(\mathbf{P}_1 \otimes \mathbf{P}_2) = \int_\mathbb{R} \left(\int_\mathbb{R} g(x, y) d\mathbf{P}_2(y) \right) d\mathbf{P}_1(x) = \int_\mathbb{R} \left(\int_\mathbb{R} g(x, y) d\mathbf{P}_1(x) \right) d\mathbf{P}_2(y).$$

On peut montrer que ceci se généralise à toutes les fonctions $g(x, y) = \mathbf{1}_A(x, y)$, avec $A \in \mathcal{B}_1 \times \mathcal{B}_2$ (voir cours intégration 2). Ensuite, cela se généralise aisément à toutes les fonctions simples, puis par un argument de convergence monotone à toutes les fonctions mesurables positives. Le théorème de convergence dominée permet d'obtenir cela pour toutes les fonctions g qui sont $\mathbf{P}_1 \otimes \mathbf{P}_2$ intégrables. En résumé, on a :

Théorème 3.1.1 (Théorème de Fubini) *Soient $\mathbf{P}_1$ et $\mathbf{P}_2$ deux probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Soit $\mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2$ la mesure produit. Soit $g : (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) \to \mathbb{R}$ une fonction mesurable qui est $\mathbf{P}$-intégrable (resp. positive). Alors les fonctions $y \mapsto \int_\mathbb{R} g(x, y) d\mathbf{P}_1(x)$ et $x \mapsto \int_\mathbb{R} g(x, y) d\mathbf{P}_2(y)$ sont mesurables, et on a*

$$\int_{\mathbb{R}^2} g d(\mathbf{P}_1 \otimes \mathbf{P}_2) = \int_\mathbb{R} \left(\int_\mathbb{R} g(x, y) d\mathbf{P}_2(y) \right) d\mathbf{P}_1(x) = \int_\mathbb{R} \left(\int_\mathbb{R} g(x, y) d\mathbf{P}_1(x) \right) d\mathbf{P}_2(y).$$

Démonstration : Voir cours intégration 2 □

En pratique, lorsqu'on travaille avec des fonctions mesurables positives, ce théorème dit qu'intégrer par rapport à la mesure produit $\mathbf{P}_1 \otimes \mathbf{P}_2$, c'est pareil que d'intégrer d'abord suivant une coordonnée, puis suivant l'autre. C'est comme cela qu'on calcule en pratique $\int_{\mathbb{R}^2} g \, d(\mathbf{P}_1 \otimes \mathbf{P}_2)$.

3.2 Couple de variables aléatoires

3.2.1 Couple de variables aléatoires, densité marginale

Soit $X : (\Omega, \mathcal{A}, \mathbf{P}) \to (\mathbb{R}^{d_1}, \mathcal{B}(\mathbb{R}^{d_1}))$ et $Y : (\Omega, \mathcal{A}, \mathbf{P}) \to (\mathbb{R}^{d_2}, \mathcal{B}(\mathbb{R}^{d_2}))$. Le *couple* (X, Y) est simplement la variable aléatoire $Z = (X, Y) : \omega \in \Omega \mapsto Z(\omega) = (X_1(\omega), X_2(\omega)) \in \mathbb{R}^{d_1+d_2}$. La *loi conjointe* de X et Y est la loi du couple $Z = (X, Y)$.