

Loi et intégration des variables aléatoires

*

On se place dans un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

1 Loi des variables aléatoires

1.1 Définition

Soit (E, \mathcal{E}) un espace mesurable et $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$ une v.a. Alors X *transporte* la mesure de probabilité \mathbb{P} en une nouvelle mesure de probabilité P_X sur (E, \mathcal{E}) , définie par

$$\forall B \in \mathcal{E}, \quad P_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B).$$

Définition 1. La mesure de probabilité P_X s'appelle la *loi* de X . On note parfois $P_X = \mathcal{L}(X)$.

Remarque 2.

- 1) P_X est la “*distribution*” des valeurs de X . Elle répond à la question “avec quelle probabilité X prend-elle une valeur (resp. un ensemble de valeurs) donné(e) ?
- 2) En intégration, on dit que P_X est la “*mesure image*” de \mathbb{P} par X .

Deux v.a.r. sont égales *presque sûrement* si elles sont égales \mathbb{P} p.p. (on dira plutôt \mathbb{P} p.s.). Ainsi,

$$\begin{aligned} X = Y \text{ p.s.} &\Leftrightarrow \exists N \in \mathcal{A} \text{ tel que } \mathbb{P}(N) = 0 \text{ et } \forall \omega \notin N, X(\omega) = Y(\omega) \\ &\Leftrightarrow \mathbb{P}(X = Y) = 1. \end{aligned}$$

Si $X = Y$ p.s., leur loi est la même (mais la réciproque est fausse).

Exemple 3. On dit qu’une v.a. X suit une loi de Bernoulli si elle prend deux valeurs (p.s.), i.e. si sa loi est portée par deux points. On précise souvent cela comme suit.

Une v.a. X suit une loi de Bernoulli de paramètre $p \in [0, 1]$ si elle est (p.s.) à valeurs dans $\{0, 1\}$ et $\mathbb{P}(X = 1) = p$. Et donc $\mathbb{P}(X = 0) = 1 - p$. On a $P_X = (1 - p)\delta_0 + p\delta_1$. On note $X \sim B(1, p)$.

Une v.a. Y suit une loi de Rademacher si $P_Y = \frac{1}{2}(\delta_{-1} + \delta_1)$. De manière équivalente $Y = 2X - 1$ où $X \sim B(1, \frac{1}{2})$.

Exemple 4. Soit $\lambda > 0$ et X une v.a.r. On dit que X suit une loi de Poisson¹ de paramètre λ si

$$\forall k \in \mathbb{N}, \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

La loi P_X , portée par \mathbb{N} , est alors $P_X = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \delta_k$. On note $X \sim \mathcal{P}(\lambda)$.

1. La loi de Poisson est la *loi des événements rares*. Elle apparaît comme une limite adéquate de lois binômiales.

1.2 Existence d'une v.a. de loi donnée

Beaucoup d'énoncés d'exercices commenceront par *Soit X une v.a. de loi μ* où (E, \mathcal{E}) est un espace mesurable fixé et μ une mesure de probabilité sur (E, \mathcal{E}) . Remarquons ici qu'il est toujours possible de trouver une telle X . Il suffit (même si c'est artificiel) de considérer comme espace de probabilité $\Omega \equiv E$, $\mathcal{A} \equiv \mathcal{E}$ et $\mathbb{P} = \mu$, et pour X , l'identité de $\Omega = E$ dans E .

En revanche, de nombreuses v.a. distinctes peuvent avoir la même loi (sans même être nécessairement définies sur le même espace de probabilité).

Exemple 5. i. Si $X \sim B(1, \frac{1}{2})$, $1 - X \sim B(1, \frac{1}{2})$.

ii. Si $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \in \mathbb{N}$, $q_n = \sum_{k=0}^n p_k$, $n \in \mathbb{N}$, ($q_{-1} = 0$) et $I_n = [q_{n-1}, q_n[$ alors la v.a. $\tilde{X} \equiv \sum_{n \in \mathbb{N}} n \mathbf{1}_{I_n}$ définie sur $(\Omega, \mathcal{A}, \mathbb{P}) \equiv ([0, 1[, \mathcal{B}([0, 1]), \lambda_1(\cdot \cap [0, 1])$) suit une loi de Poisson de paramètre λ .

1.3 Fonction de répartition d'une v.a.

Définition 6. Soit X une v.a.r. La fonction de répartition F_X de X est la fonction de répartition de sa loi P_X i.e.

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

Remarque 7. La fonction de répartition *détermine la loi*.

Exemple 8. Une v.a.r. X suit une loi exponentielle² de paramètre $\lambda > 0$ (on note $X \sim \text{Exp}(\lambda)$), si $P_X = f_\lambda \bullet \lambda_1$, $f_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0, \infty[}(x)$. On notera souvent formellement, pour dire cela,

$$\mathbb{P}(X \in dx) = \lambda e^{-\lambda x} \mathbf{1}_{[0, \infty[}(x) dx.$$

On a alors

$$F_X(t) = (1 - e^{-\lambda t}) \mathbf{1}_{[0, \infty[}(t).$$

1.4 Loïs discrètes, loïs continues

Définition 9. Les *loïs de probabilité* sont les mesures de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (ou $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$).

Ce sont donc les distributions des valeurs de toutes les v.a. (resp. v.ä.) possibles.

On rappelle que toute loi de probabilité μ sur \mathbb{R} se décompose de manière unique comme une combinaison convexe de trois lois de probabilité, l'une discrète, la seconde (absolument) continue, la troisième diffuse et singulière

$$\mu = \alpha \mu_d + \beta \mu_{a.c.} + \gamma \mu_s, \quad \alpha, \beta, \gamma \in \mathbb{R}_+ \text{ t.q. } \alpha + \beta + \gamma = 1.$$

Un exemple de mesure singulière a été étudié en D.M. Dans la suite du cours, on se consacrera essentiellement à savoir manipuler (sans hésitation) les deux autres *extrêmes*, les cas discrets et (absolument) continus.

Loi discrètes Une loi de probabilité P_X est une loi discrète si elle est portée par une suite $(x_n)_{n \in \mathbb{N}^*} \subset \mathbb{R}^d$. Pour la définir, on se donne des poids $(p_n)_{n \in \mathbb{N}^*} \geq 0$ tels que $Z \equiv \sum_{n=1}^{+\infty} p_n < +\infty$. Alors

$$P_X = \frac{1}{Z} \sum_{n=1}^{+\infty} p_n \delta_{x_n}.$$

2. La loi exponentielle joue un grand rôle dans l'étude des *processus de Markov à temps continu* (des v.a. qui évoluent dans le temps et dont le futur ne dépend -aléatoirement- que du présent). Vous en connaissez peut-être deux : le mouvement Brownien et le processus de Poisson (D.M. 2 ?)

Loi continues Un v.â. $X = (X_1, \dots, X_d)$ à valeurs \mathbb{R}^d suit une loi continue s'il existe $f \in \mathbb{L}_+^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d) \equiv \mathbb{L}_+^1(\mathbb{R}^d)$ telle que $P_X = \frac{1}{Z} f \bullet \lambda_d$ où $Z = \int_{\mathbb{R}^d} f(x) dx$. Cela signifie que, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathbb{P}(X \in A) = \frac{1}{Z} \int_A f(x) dx,$$

ce que l'on note *formellement*

$$\mathbb{P}(X \in dx) = f(x) \frac{dx}{Z}.$$

Exemple 10. Loi uniforme sur $[a, b]$, $a, b \in \mathbb{R}$, $b > a$: $P_X(dx) = \mathbf{1}_{[a,b]} \frac{dx}{(b-a)}$.

Exemple 11. Loi gaussienne (centrée réduite) ou **loi normale**³ : $\mathbb{P}(X \in dx) = e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$.

La valeur de $Z = \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$ n'est pas si simple à déterminer. On présente ci-dessous une manière d'y parvenir qui illustre élégamment le théorème de Fubini et la formule de changement de variables (voir l'appendice ??).

La fonction $g(x, y) = e^{-\frac{x^2+y^2}{2}} \in \mathcal{L}_+^1(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ ($\underbrace{dx dy}_{=\lambda_2(dx, dy)}$) est positive, radiale et on peut en séparer les facteurs. Ainsi, par le théorème de Fubini,

$$\int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy = \left(\int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx \right)^2 < \infty.$$

Et de plus, comme $\lambda_2(\mathbb{R}_+ \times \{0\}) = 0$,

$$\int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy = \int_{\mathbb{R}^2 \setminus (\mathbb{R}_+ \times \{0\})} e^{-\frac{x^2+y^2}{2}} dx dy.$$

C'est sur cette dernière intégrale qu'on peut utiliser le changement de variable en coordonnées polaires

$$\Phi : \begin{array}{ll} D =]0, +\infty[\times]0, 2\pi[& \rightarrow \Delta = \mathbb{R}^2 \setminus (\mathbb{R}_+ \times \{0\}) \\ (\rho, \theta) & \mapsto (\rho \cos \theta, \rho \sin \theta) \end{array}$$

C'est un C^1 difféomorphisme entre ouverts. Grâce au théorème d'inversion locale, il suffit de vérifier que 1) D est ouvert 2) Φ est une bijection de D sur Δ , 3) Φ est C^1 sur D , 4) $J_\Phi(\rho, \theta) \equiv \det \text{Jac}(\Phi)(\rho, \theta) \neq 0$ sur D . On a

$$\text{Jac}(\Phi) = \begin{pmatrix} \frac{\partial \Phi}{\partial \rho} & \frac{\partial \Phi}{\partial \theta} \\ \cos \theta & -\rho \sin \theta \\ \sin \theta & \rho \cos \theta \end{pmatrix}$$

et donc $J_\Phi(\rho, \theta) = \rho$. Il ne reste qu'à achever le calcul :

$$Z^2 = \int_D g \circ \Phi(\rho, \theta) |J_\Phi(\rho, \theta)| d\rho d\theta = 2\pi \int_0^{+\infty} e^{-\rho^2/2} \rho d\rho = 2\pi.$$

3. cette loi, appelée aussi parfois Maxwellienne par les physiciens, est très importante. On verra (théorème central limite) qu'elle apparaît pour décrire les *fluctuations* autour de la moyenne.

1.5 Loi des vecteurs aléatoires

Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire à valeurs \mathbb{R}^d .

Sa fonction de répartition est la fonction $F_X(t_1, \dots, t_d) = \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d)$ définie sur \mathbb{R}^d .

Proposition 12. La fonction de répartition détermine la loi des vecteurs aléatoires i.e. pour tous v.à. X et Y à valeurs \mathbb{R}^d , non nécessairement définis sur le même espace de probabilité, $F_X = F_Y \Rightarrow P_X = P_Y$.

Démonstration. Il suffit d'appliquer le corollaire ?? en remarquant que $\mathcal{C} = \{] - \infty, t_1] \times \dots \times] - \infty, t_d] : t_1, \dots, t_d \in \mathbb{R} \}$ est un π -système engendrant $\mathcal{B}(\mathbb{R}^d)$. \square

Définition 13 (Marginales). Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire à valeurs \mathbb{R}^d . La i -ème marginale P_{X_i} de P_X est la loi de X_i :

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad P_{X_i}(A) = \mathbb{P}(X \in \overbrace{\mathbb{R} \times \dots \times \mathbb{R} \times A}^{d \text{ facteurs}} \times \mathbb{R} \times \dots \times \mathbb{R}) = \mathbb{P}(X_i \in A)$$

i-ème

P_X détermine les d marginales $(P_{X_i})_{i=1, \dots, d}$ mais la réciproque est fausse.

Exemple 14. Soient $X = (X_1, X_2)$ et $Y = (Y_1, Y_2)$ à valeurs dans $\{0, 1, 2\}^2$. Leurs lois (distinctes) représentées ci-dessous par leurs *tables de fréquences* ont les mêmes marginales.

$X_1 = i \setminus X_2 = j$	0	1	2	margin. X_1
0	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{3}{8}$
1	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{7}{16}$
2	0	0	$\frac{3}{16}$	$\frac{3}{16}$
margin. X_2	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{1}{2}$	1

$Y_1 = i \setminus Y_2 = j$	0	1	2	margin. Y_1
0	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{3}{8}$
1	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{7}{16}$
2	0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$
margin. Y_2	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{1}{2}$	1

2 Intégration des variables aléatoires

2.1 Espérance

Nous renvoyons à l'appendice ?? pour un bref survol de la définition de l'intégrale de Lebesgue. Les *trois théorèmes de convergence* y sont rappelés, formulés en terme d'espérance.

Définition 15. Si X est une v.a.r. intégrable, $\int_{\Omega} X d\mathbb{P}$ est appelée *espérance* de X et notée $\mathbb{E}(X)$. On note $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ l'ensemble des v.a.r. intégrables et $\mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ l'espace de Banach obtenu en prenant le quotient de \mathcal{L}^1 par la relation d'équivalence $X = Y$ p.s.. Sa norme est notée $\|X\|_1 = \mathbb{E}(|X|)$.

2.2 Inégalités de Markov et de Jensen

Inégalité de Markov : Soit $X \geq 0$ p.s. une v.a. et $a > 0$. Alors

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

Cette inégalité, anodine au premier abord, est en fait souvent utile en calcul des probabilités. Remarquez qu'elle n'a d'intérêt que pour a suffisamment grand.

Démonstration. $a\mathbb{P}(X \geq a) = \int_{\{X \geq a\}} a d\mathbb{P} \leq \int_{\{X \geq a\}} X d\mathbb{P} \leq \mathbb{E}(X)$. \square

Inégalité de Jensen : Soit X une v.a.r. intégrable et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe telle que $\varphi \geq 0$ ou bien $\varphi(X)$ est intégrable. Alors

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

La convexité joue un rôle primordial en théorie des probabilités. L'inégalité de Jensen permet de considérer l'intégrale par rapport à une mesure de probabilité comme une généralisation naturelle des combinaisons convexes.

Démonstration. Si $\varphi : I \rightarrow \mathbb{R}$ est convexe sur un intervalle I contenant x_0 les taux d'accroissement $x \in I \setminus \{x_0\} \mapsto \frac{\varphi(x) - \varphi(x_0)}{x - x_0} \in \mathbb{R}$ sont croissants et donc φ est dérivable à gauche et à droite en tout point et,

$$\forall x < x_0 < x', \quad \frac{\varphi(x) - \varphi(x_0)}{x - x_0} \leq \varphi'(x_{0-}) \leq \varphi'(x_{0+}) \leq \frac{\varphi(x') - \varphi(x_0)}{x' - x_0}.$$

Il s'ensuit que, pour tout x ,

$$\varphi(x) \geq \varphi(x_0) + \varphi'(x_{0+})(x - x_0).$$

(Remarquez que n'importe quelle valeur c entre $\varphi'(x_{0-})$ et $\varphi'(x_{0+})$ aurait tout aussi bien fait l'affaire). On applique cela à $x = X(\omega)$ et on intègre en \mathbb{P} . Il vient

$$\mathbb{E}(\varphi(X)) \geq \varphi(x_0) + c(\mathbb{E}(X) - x_0)$$

ce qui permet de conclure en choisissant $x_0 = \mathbb{E}(X)$. □

2.3 Lemme de transport

Lemme 16. Soit X un v.à. à valeurs \mathbb{R}^d de loi P_X et $\Psi : \mathbb{R}^d \rightarrow \mathbb{R}$ une application borélienne. Alors,

$$\Psi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P}) \Leftrightarrow \Psi \in \mathbb{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_X).$$

Et, si cela est vérifié, ou si $\Psi \geq 0$, alors

$$\mathbb{E}(\Psi(X)) = \int_{\mathbb{R}^d} \Psi(x) P_X(dx). \tag{1}$$

La moralité de ce lemme est que, pour intégrer les fonctions de X , la loi de X contient toute l'information nécessaire.

Remarque 17. Ce résultat s'étend sans difficulté à toute mesure image (non nécessairement une probabilité).

Remarque 18 (Caractérisation de la loi). Pour déterminer la loi d'un v.à. X , fonction d'autres v.à. dont la loi est connue, on utilise souvent le lemme de transport de la manière suivante. On montre (en utilisant le thm de Fubini, un changement de variable ou d'autres techniques lorsqu'il y a lieu; cf TD) que, pour toute $\Psi : \mathbb{R}^d \rightarrow \mathbb{R}$ borélienne bornée (abrégé bo.bo.),

$$\mathbb{E}(\Psi(X)) = \int_{\mathbb{R}^d} \Psi(x) \mu(dx)$$

où μ est une mesure (nécessairement de probabilité; pourquoi?). On en tire alors $P_X = \mu$.

Bien évidemment, prendre $\Psi = \mathbf{1}_B$, $B \in \mathcal{B}(\mathbb{R}^d)$, suffirait, mais considérer une fonction Ψ *générique* permet de se concentrer sur l'essentiel (les techniques à utiliser).

Preuve du lemme de transport. On va établir (1) pour toute Ψ borélienne positive. Cela permet de conclure, puisqu'alors, pour toute $\Psi : \mathbb{R}^d \rightarrow \mathbb{R}$ borélienne,

$$\begin{aligned} \Psi \in \mathbb{L}^1(\mathbb{R}^d, P_X) &\Leftrightarrow \left(\int \Psi_+ dP_X < +\infty \text{ et } \int \Psi_- dP_X < +\infty \right) \\ &\Leftrightarrow \left(\int \underbrace{\Psi_+(X)}_{=(\Psi(X))_+} d\mathbb{P} < +\infty \text{ et } \int \underbrace{\Psi_-(X)}_{=(\Psi(X))_-} d\mathbb{P} < +\infty \right) \Leftrightarrow \Psi(X) \in \mathbb{L}^1(\Omega, \mathbb{P}). \end{aligned}$$

Et $\Psi = \Psi_+ - \Psi_-$ donne la formule (1) pour Ψ quelconque à partir de celles pour Ψ_+ et Ψ_- .

Le cas Ψ borélienne positive se traite *par approximation par les fonctions simples*. Cet argument, dans la suite résumé par ces quelques mots, est souvent utile et s'appuie sur le lemme fondamental ???. Une fois n'est pas coutume, nous le détaillerons ici.

On commence par le cas où $\Psi = \mathbf{1}_B$, $B \in \mathcal{B}(\mathbb{R}^d)$. Alors $\Psi(X) = \mathbf{1}_{\{X \in B\}}$ et

$$\mathbb{E}(\Psi(X)) = \mathbb{P}(X \in B) = P_X(B) = \int_{\mathbb{R}^d} \mathbf{1}_B(x) P_X(dx).$$

(1) s'étend alors par linéarité aux fonctions simples. Enfin, si $\Psi \geq 0$ borélienne, elle est limite d'une suite croissante $(\Psi_n)_n$ de fonctions simples. Il reste à utiliser le théorème de convergence monotone pour \mathbb{P} d'une part et pour P_X d'autre part :

$$\mathbb{E}(\Psi(X)) = \sup_n \uparrow \mathbb{E}(\Psi_n(X)) = \sup_n \uparrow \int_{\mathbb{R}^d} \Psi_n(x) P_X(dx) = \int_{\mathbb{R}^d} \Psi(x) P_X(dx).$$

□

2.4 Caractérisation de la loi : restriction de la classe des fonctions test

La remarque 18 introduisait une technique à acquérir pour déterminer la loi d'un v.â. (qui vous aidera aussi en M1 à déterminer des *espérances ou lois conditionnelles*) : tester la loi contre les fonctions “bo.bo.” i.e. les fonctions $\mathcal{L}_b^0(\mathbb{R}^d)$.

Il nous faudra plus tard restreindre la classe des fonctions tests aux fonctions continues à support compact. C'est le but de la proposition qui suit, en particulier utile lorsque $\mu = P_X$ est la loi d'un v.â. et ν est une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Proposition 19. Soient μ et ν deux mesures finies sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Les propositions suivantes sont équivalentes :

- i. $\mu = \nu$;
- ii. $\forall \Psi \in \mathcal{L}_b^0(\mathbb{R}^d), \quad \int \Psi(x) \mu(dx) = \int \Psi(x) \nu(dx)$;
- iii. $\forall \Psi \in C_b(\mathbb{R}^d), \quad \int \Psi(x) \mu(dx) = \int \Psi(x) \nu(dx)$;
- iv. $\forall \Psi \in C_c(\mathbb{R}^d), \quad \int \Psi(x) \mu(dx) = \int \Psi(x) \nu(dx)$;

Remarque 20. On pourrait encore restreindre aux fonctions test $C_c^\infty(\mathbb{R}^d)$. C'est une conséquence de la régularisation par convolution que nous aborderons au prochain chapitre.

Démonstration. $i) \Rightarrow ii) \Rightarrow iii) \Rightarrow iv)$ est trivial.

Montrons maintenant $iii) \Rightarrow i)$. On utilise le lemme d'Urysohn ?? pour approximer l'indicatrice d'un ouvert. Soit U un ouvert propre de \mathbb{R}^d (i.e. $U \neq \mathbb{R}^d, \emptyset$). Alors, pour $\varepsilon > 0$ assez petit,

$$F_\varepsilon \equiv \{x \in \mathbb{R}^d : d(x, U^c) \geq \varepsilon\}$$

est un fermé non vide disjoint de U^c , lui-même fermé $\neq \emptyset$. La fonction $f_\varepsilon(x) \equiv \frac{d(x, U^c)}{d(x, U^c) + d(x, F_\varepsilon)}$ est continue vaut 0 sur U^c et 1 sur F_ε . De plus, $F_\varepsilon \subset F_{\varepsilon'}$ si $0 < \varepsilon' \leq \varepsilon$ entraîne $d(\cdot, F_{\varepsilon'}) \leq d(\cdot, F_\varepsilon)$. La suite $(f_{\frac{1}{n}})_n$ définie pour n assez grand est donc croissante et, U étant ouvert, $(f_{\frac{1}{n}})_n \uparrow \mathbf{1}_U$ quand $n \rightarrow +\infty$. Par convergence monotone,

$$\mu(U) = \sup_n \int f_{\frac{1}{n}}(x) \mu(dx) = \sup_n \int f_{\frac{1}{n}}(x) \nu(dx) = \nu(U).$$

Et cela est encore vérifié pour $U = \mathbb{R}^d$ (en prenant $\Psi = \mathbf{1}$) et $U = \emptyset$ (trivial). On conclut grâce au corollaire ?? puisque l'ensemble des ouverts est un π -système qui engendre $\mathcal{B}(\mathbb{R}^d)$.

Pour $iv) \Rightarrow i)$, on procède de même en remplaçant U par $U^{(n)} = U \cap B(0, n[$ (boule ouverte). Les $f_\varepsilon^{(n)}$ sont alors $C_c(\mathbb{R}^d)$. \square

3 Les espaces $\mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$

La plupart des résultats énoncés dans la première section sont valables pour tout espace mesuré (E, \mathcal{E}, μ) et ont été vus au premier semestre dans ce cadre. Néanmoins, certains résultats (comme la croissance des normes par exemple) utilisent que la mesure \mathbb{P} est une mesure de probabilité. On emploiera les notations probabilistes. Le but est ici de parcourir rapidement les résultats, quitte à simplifier parfois un peu la trame. En particulier, on ne hiérarchisera pas nécessairement les résultats en propositions, théorèmes, etc...

3.1 Définitions et premières propriétés

Soit $1 \leq p < +\infty$ et X une v.a.r. On dit que $X \in \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ si $\mathbb{E}(|X|^p) < +\infty$.

En quotientant $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ par la relation d'équivalence $X = Y$ p.s., on obtient l'espace vectoriel $\mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$. On notera plus concisément \mathbb{L}^p lorsque l'espace de probabilité est fixé (comme ici). On pose $\|X\|_p \equiv \mathbb{E}(|X|^p)^{\frac{1}{p}}$.

Le cas $p = +\infty$:

$$X \in \mathcal{L}^\infty(\Omega, \mathcal{A}, \mathbb{P}) \text{ ssi (définition) } \exists \lambda \in [0, \infty) \text{ t.q. } |X| \leq \lambda \text{ p.s.}$$

On définit en quotientant comme précédemment l'espace $\mathbb{L}^\infty(\Omega, \mathcal{A}, \mathbb{P})$ et on le note \mathbb{L}^∞ . On pose $\|X\|_\infty \equiv \inf\{\lambda \in [0, \infty) : |X| \leq \lambda \text{ p.s.}\}$.

Inégalité de Hölder Soient $1 \leq p, q \leq \infty$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Alors

$$\left. \begin{array}{l} X \in \mathbb{L}^p \\ Y \in \mathbb{L}^q \end{array} \right\} \implies XY \in \mathbb{L}^1 \quad \text{et} \quad \|XY\|_1 \leq \|X\|_p \|Y\|_q$$

Conséquence 21. (Inégalité de Minkowski).

Pour tout $1 \leq p \leq \infty$,

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

$\|\cdot\|_p$ est une norme sur $\mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$.

On montre que $\mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de Banach. L'espace \mathbb{L}^2 est un espace de Hilbert (réel) pour le produit scalaire (réel) $\langle X, Y \rangle_2 \equiv \mathbb{E}(XY)$. Le produit scalaire sur l'espace de Hilbert complexe $\mathbb{L}^2(\Omega; \mathbb{C})$ (cf plus loin) est $\langle X, Y \rangle_2 \equiv \mathbb{E}(X \bar{Y})$. L'**inégalité de Cauchy-Schwarz** est alors

$$|\langle X, Y \rangle_2| \leq \|X\|_2 \|Y\|_2.$$

Pour $1 \leq p < \infty$ (attention $p \neq \infty!!!$), le dual de \mathbb{L}^p est \mathbb{L}^q (si $q = \frac{p}{p-1}$ i.e. $\frac{1}{p} + \frac{1}{q} = 1$). Pour montrer ce résultat, il faut utiliser le théorème de Radon-Nikodym.

Les normes p sont croissantes : si $1 \leq r \leq s$, alors

$$\|X\|_r \leq \|X\|_s. \quad \text{Et donc } \mathbb{L}^s \subset \mathbb{L}^r.$$

Cela utilise l'inégalité de Jensen, et donc en particulier que \mathbb{P} est une probabilité.

Exercice 1. Montrer que $\|X\|_p \uparrow \|X\|_\infty$ quand $p \uparrow \infty$.

▼ On suppose ici que $\|X\|_\infty < \infty$ (à vous de traiter le cas $\|X\|_\infty = \infty$). Par homogénéité (i.e. en considérant $|X|/\|X\|_\infty$ au lieu de X), on se ramène à $0 \leq X \leq 1$ et $\|X\|_\infty = 1$. Mais alors $\forall 0 < \lambda < 1$, $\mathbb{P}(X > \lambda) > 0$. Ainsi,

$$1 \geq \mathbb{E}(X^p)^{\frac{1}{p}} \geq \lambda \mathbb{P}(X > \lambda)^{\frac{1}{p}},$$

grâce à l'inégalité de Markov. En remarquant que $\mathbb{P}(X > \lambda)^{\frac{1}{p}} \rightarrow 1$ quand $p \rightarrow +\infty$, on obtient

$$\lambda \leq \liminf_{p \rightarrow +\infty} \|X\|_p \leq \limsup_{p \rightarrow +\infty} \|X\|_p \leq 1.$$

Et ceci pour tout $\lambda < 1$, ce qui permet de conclure. ▲

Terminologie : $\mathbb{E}(|X|^p)$ est dit *moment absolu d'ordre p* . Si $n \in \mathbb{N}^*$, $\mathbb{E}(X^n)$ est le $n^{\text{ième}}$ *moment* de $X \in \mathbb{L}^n$.

3.2 Le $p^{\text{ième}}$ moment absolu via la fonction de répartition

Connaître le comportement à l'infini de $\mathbb{P}(|X| > t) = 1 - F_{|X|}(t)$ (on parle de queue de la loi de X) permet de savoir si X a un moment absolu d'ordre p fini (i.e. est dans \mathbb{L}^p si $1 \leq p < \infty$). C'est ce que signifie la proposition suivante.

Proposition 22. Soit X une v.a.r. et $0 < p < \infty$. Alors

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

De plus,

$$\mathbb{E}(|X|) < +\infty \iff \text{Pour un (ou tout) } \varepsilon > 0, \quad \sum_{n=1}^\infty \mathbb{P}(|X| > \varepsilon n) < +\infty$$

3.3 Variance et covariance

3.3.1 Cas des v.a.r.

Soit $X \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$. Alors $X \in \mathbb{L}^1$, et, comme \mathbb{P} est une probabilité, $\mathbb{E}(X) \in \mathbb{L}^2$. Ainsi, $X - \mathbb{E}(X) \in \mathbb{L}^2$.

Définition 23. i. La **variance** de $X \in \mathbb{L}^2$ est la quantité

$$\text{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

ii. L'**écart type** de X est $\sigma(X) = \sqrt{\text{var}(X)}$.

iii. La **covariance** de X et Y (toutes deux dans \mathbb{L}^2) est

$$\text{Cov}(X, Y) = \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Remarquez que $\text{var}(X) = \text{Cov}(X, X)$.

3.3.2 Extension aux vecteurs aléatoires

Soit $\vec{X} = (X_1, \dots, X_d)$ un v.à. sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs \mathbb{R}^d .

Définition 24. On dit que \vec{X} est de puissance $p^{\text{ième}}$ intégrable et on note $\vec{X} \in \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^d)$ ou $\mathcal{L}^p(\Omega; \mathbb{R}^d)$ si chacune des composantes X_i est dans $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ ou, de manière équivalente, si $\mathbb{E}(\|\vec{X}\|^p) < \infty$ (pour n'importe quel choix de norme sur \mathbb{R}^d). $\mathcal{L}^p(\Omega; \mathbb{C}) \simeq \mathcal{L}^p(\Omega; \mathbb{R}^2)$.

Remarquez que $\mathcal{L}^p(\Omega; \mathbb{R}^d) \simeq (\mathcal{L}^p(\Omega))_d$.

On définit comme précédemment l'espace $\mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^d)$ des classes de v.à. modulo l'égalité p.s.

L'espérance de $\vec{X} \in \mathbb{L}^1(\Omega; \mathbb{R}^d)$ est le vecteur $\mathbb{E}(\vec{X}) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d)) \in \mathbb{R}^d$.

La **matrice de (variance-)covariance** de $\vec{X} \in \mathbb{L}^2(\Omega; \mathbb{R}^d)$ est $\text{Cov}(\vec{X}) = (\text{Cov}(X_i, X_j))_{i,j=1,\dots,d} \in M_d(\mathbb{R})$.

Propriétés 25. i. L'application $(X_1, X_2) \in \mathbb{L}^2(\Omega; \mathbb{R}^2) \mapsto \text{Cov}(X_1, X_2) \in \mathbb{R}$ est une forme bilinéaire continue, symétrique, positive.

ii. $\text{var}(X_1 + X_2) = \text{var}(X_1) + 2 \text{Cov}(X_1, X_2) + \text{var}(X_2)$.

iii. Si $\vec{X} \in \mathbb{L}^2(\Omega; \mathbb{R}^d)$, $\text{Cov}(\vec{X}) \in M_d(\mathbb{R})$ est une matrice symétrique, positive.

iv. Pour tout $\vec{X} \in \mathbb{L}^2(\Omega; \mathbb{R}^d)$ et toute $A \in M_{n,d}(\mathbb{R})$, le v.à. $A\vec{X}$ est dans $\mathbb{L}^2(\Omega; \mathbb{R}^n)$. Et on a

$$\mathbb{E}(A\vec{X}) = A \mathbb{E}(\vec{X}) \quad \text{et} \quad \text{Cov}(A\vec{X}) = A \text{Cov}(\vec{X}) A'.$$

(On a identifié un vecteur au vecteur colonne de ses coordonnées dans la base canonique).

v. **Inégalité de Tchebichev.** Soit $X \in \mathbb{L}^2(\Omega)$. Alors, pour tout $a > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\text{var}(X)}{a^2}.$$

3.4 Moyenne et variance de lois usuelles

3.4.1 Loi binômiale

X suit une loi binômiale⁴ $B(n, p)$, $n \in \mathbb{N}^*$, $p \in]0, 1[$, si

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

Alors $\mathbb{E}(X) = np$ et $\text{Var}(X) = np(1-p)$.

3.4.2 Loi de Poisson

Soit $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$. Alors $\mathbb{E}(X) = \text{Var}(X) = \lambda$.

3.4.3 Loi uniforme

Soit X de loi uniforme sur $[a, b]$, $a < b$ (noté $X \sim \mathcal{U}([a, b])$). Alors $\mathbb{E}(X) = \frac{a+b}{2}$, $\text{Var}(X) = \frac{(b-a)^2}{12}$.

4. On verra plus tard que c'est la loi de la somme de n v.a. indépendantes de même loi de Bernoulli $B(1, p)$.

3.4.4 Loi gaussienne

X suit une loi gaussienne de moyenne $m \in \mathbb{R}$ et de variance $\sigma^2 > 0$ si

$$\mathbb{P}(X \in dx) = \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \frac{dx}{\sqrt{2\pi\sigma^2}}.$$

Comme leurs noms l'indiquent, on a alors $\mathbb{E}(X) = m$ et $\text{Var}(X) = \sigma^2$.

3.4.5 Premiers pas vers les lois gaussiennes vectorielles

Soit $m = (m_1, \dots, m_d)$ et $\sigma > 0$. Le v.à. $X = (X_1, \dots, X_d)$ suit une loi gaussienne $\mathcal{N}(m, \sigma^2 I_d)$ (I_d désignant la matrice identité) si

$$\mathbb{P}(X \in (dx_1, \dots, dx_d)) = \exp\left(-\frac{\sum_{i=1}^d (x_i - m_i)^2}{2\sigma^2}\right) \frac{dx_1 \dots dx_d}{(\sqrt{2\pi}\sigma)^d}.$$

Alors $\mathbb{E}(X) = m$ et $\text{Cov}(X) = \sigma^2 I_d$.

4 Fonction caractéristique

Définition 26. Soit X un vecteur aléatoire sur $(\omega, \mathcal{A}, \mathbb{P})$ à valeurs dans \mathbb{R}^d . La fonction caractéristique de X , notée φ^X est la fonction à valeurs complexes

$$t \in \mathbb{R}^d \mapsto \varphi^X(t) = \mathbb{E}(e^{i\langle t, X \rangle}) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} d\mathbb{P}_X(x).$$

La fonction caractéristique est une fonction continue à valeurs complexes, de module majoré par 1, telle que $\varphi^X(0) = 1$.

Si la loi de X admet une densité f par rapport à la mesure de Lebesgue sur \mathbb{R}^d alors

$$\varphi^X(t) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} f(x) dx = \hat{f}(t).$$

Théorème 27. Si X et Y sont deux vecteurs aléatoires sur $(\omega, \mathcal{A}, \mathbb{P})$ à valeurs dans \mathbb{R}^d de loi \mathbb{P}^X et \mathbb{P}^Y telles que $\varphi^X = \varphi^Y$ alors $\mathbb{P}^X = \mathbb{P}^Y$.

Plusieurs preuves sont possibles dont une utilisant le théorème des classes monotones fonctionnel.

Théorème 28. Formule d'inversion de Fourier

Soit X un vecteur aléatoire sur $(\omega, \mathcal{A}, \mathbb{P})$ à valeurs dans \mathbb{R}^d tel que φ^X soit intégrable par rapport à la mesure de Lebesgue. Alors X admet une densité continue bornée f^X par rapport à la mesure de Lebesgue sur \mathbb{R}^d , donnée, pour tout $x \in \mathbb{R}^d$ par

$$f^X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t, x \rangle} \varphi^X(t) dt.$$

Proposition 29. Soit X une variable aléatoire sur $(\omega, \mathcal{A}, \mathbb{P})$ de fonction caractéristique $\varphi^X = \varphi$ de loi \mathbb{P}^X .

– Si $\mathbb{E}(|X|^m) < \infty$, alors φ^X est m fois dérivable, de dérivée k -ième ($k \leq m$)

$$\varphi^{(k)}(t) = \mathbb{E}(i^k X^k e^{itX}).$$

En particulier $\varphi^{(k)}(0) = \mathbb{E}(i^k X^k)$.

- Réciproquement, si φ est k fois dérivable en 0, alors X admet des moments d'ordre plus petit ou égaux à $2n \leq k$.

Définition 30. Soit X une variable aléatoire sur $(\omega, \mathcal{A}, \mathbb{P})$, on appelle transformée de Laplace (ou fonction génératrice des moments) la fonction $L^X(s) = \mathbb{E}(e^{sX})$ définie pour les valeurs de s telles que e^{sX} soit intégrable.

Proposition 31. Soit X une variable aléatoire sur $(\omega, \mathcal{A}, \mathbb{P})$ telle que e^{tX} est intégrable pour t dans un intervalle contenant 0. Alors la transformée de Laplace est définie sur cet intervalle contenant 0, analytique au voisinage de 0 et

$$L^X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}(X^n)$$

pour t dans ce voisinage.