Université Paul Sabatier — Deuxième semestre 2014-2015
L3 FONDA : Intégration 2 et Probabilités

Espaces de probabilités et variables aléatoires

En probabilités les espaces s'appellent le plus souvent Ω, les tribus $\mathcal{F}$ ou $\mathcal{A}$ et les mesures $\mathbb{P}$.

1 Définitions et exemples

1.1 Expérience aléatoire

On cherche à décrire une *expérience aléatoire*
- dont le résultat appartient à un ensemble bien déterminé et connu à l'avance,
- mais ce résultat lui-même n'est pas connu à l'avance avec certitude.

On peut modéliser une telle expérience par un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ où
- Ω est l'ensemble des résultats possibles (dit *espace d'épreuves*). Un élement $\omega \in \Omega$ est donc un résultat possible de l'expérience (dit une *épreuve*, ou un *évènement élémentaire*).
- $\mathcal{A}$: ensemble des *évènements* observables i.e. dont on sait dire après l'expérience s'ils sont réalisés ou non. $A \in \mathcal{A}$ est donc un *évènement.*
- $\mathbb{P}$ est une mesure de probabilité sur $\mathcal{A}$ décrivant l'aléa de l'expérience.

Exemple 1. Jet de deux dés.
$\Omega = \{1, \dots, 6\}^2$. $\mathrm{Card}(\Omega) = \sharp(\Omega) = 36$.
$\mathcal{A} = \mathcal{P}(\Omega)$ (comme souvent lorsque Ω est fini).
Exemple d'évènement : $A = \{\text{obtenir au moins un } 6\} = \{(1,6); (2,6); \dots; (6,6); (6,5); (6,4); \dots; (6,1)\}$.
$A^c = \{\text{n'obtenir aucun } 6\}$.
$\mathbb{P}$ est l'équiprobabilité i.e. $\mathbb{P}(\{(i,j)\}) = \frac{1}{36}$ pour tout résultat possible $\omega = (i,j)$, $i, j \in \{1, \dots, 6\}$.

$$\mathbb{P}(A) = \frac{\text{nombre de résultats favorables}}{\text{nombre de résultats possibles}} = \frac{\sharp(A)}{\sharp(\Omega)} = \frac{11}{36} = 1 - \mathbb{P}(A^c) = 1 - (\frac{5}{6})^2.$$

Exemple 2. Jeu de pile ou face.

Une partie : $\Omega = \{0, 1\}$. $\mathcal{A} = \mathcal{P}(\Omega) = \big\{\emptyset; \{0\}; \{1\}; \{0; 1\}\big\}$. $\omega = \left| \begin{array}{l} 0 \text{ si le résultat est Face} \\ 1 \text{ si le résultat est Pile} \end{array} \right.$
$\mathbb{P}(\{1\}) = p, \mathbb{P}(\{0\}) = 1 - p$, $0 < p < 1$. $p = \frac{1}{2}$ si la pièce est équilibrée.

n parties (qu'on suppose équilibrées).
$\Omega = \{0, 1\}^n$. $\mathcal{A} = \mathcal{P}(\Omega)$. (rm : $\sharp(\mathcal{P}(\Omega) = 2^{2^n}$).
$\mathbb{P}$ est l'équiprobabilité : $\forall (x_1, \dots, x_n) \in \{0,1\}^n, \mathbb{P}(\{(x_1, \dots, x_n)\}) = \frac{1}{2^n}$.

1.2 Espace probabilisés

Définition 3. Soit $(\Omega, \mathcal{A})$ un espace mesurable. Une probabilité ou une mesure de probabilité sur $(\Omega, \mathcal{A})$ est une mesure positive $\mathbb{P}$ sur $\mathcal{A}$ telle que $\mathbb{P}(\Omega) = 1$. L'espace $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé, $\mathbb{P}$ une loi de probabilité ou simplement une loi.

Si μ est une mesure finie sur $(\Omega, \mathcal{A})$, $\quad 0 < \mu(\Omega) < \infty$, alors $\mathbb{P} =:= \frac{\mu}{\mu(\Omega)}$ est une mesure de probabilité.

Les propriétés suivantes sont la traduction des propriétés des mesures dans le cas de mesure de masse 1. Si $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé et $A, B, A_n, n \in \mathbb{N}$ appartiennent à $\mathcal{A}$,

i. $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$,

ii. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,

iii. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,

iv. $\mathbb{P}(\cup_n A_n) \leq \sum_n \mathbb{P}(A_n)$,

v. si la suite $(A_n)_n$ est croissante ou déctroissante pour l'inclusion,

$$\mathbb{P}(\cup A_n) = \lim_n \mathbb{P}(A_n), \mathbb{P}(\cap A_n) \qquad = \lim_n \mathbb{P}(A_n).$$

vi. Inégalité de Bonférroni,

$$\forall k \geq 2, \quad \sum_{n=1}^{k} \mathbb{P}(A_n) - \sum_{n=2}^{k} \sum_{m=1}^{n-1} \mathbb{P}(A_n \cap A_m) \leq \mathbb{P}(\cup_{n=1}^{k} A_n) \leq \sum_{n=1}^{k} \mathbb{P}(A_n).$$

Exemple 2 (suite) Jeu de pile ou face infini
(ce modèle sera "construit" à la fin du chapitre).
On veut pouvoir considérer l'évènement

$$A = \{ \text{ on n'obtient que des piles indéfiniment } \}.$$

Intuitivement, $0 \leq \mathbb{P}(A) \leq \mathbb{P}(A_n)$ où $A_n = \{$ on n'obtient que des piles lors des n premiers lancers $\}$ puisque $A = \cap_{p \in \mathbb{N}^*} A_p \subset A_n$. Or, $\mathbb{P}(A_n) = \frac{1}{2^n} \to 0$ quand $n \to \infty$. D'où $\mathbb{P}(A) = 0$. On verra qu'on peut écrire

$$\mathbb{P}(\cap_{n \in \mathbb{N}^*} \downarrow A_n) = \lim_n \downarrow \mathbb{P}(A_n).$$

Moins artificiellement, la construction du jeu de pile ou face infini permet de répondre à des questions du type

Quelle est la probabilité de tirer (au moins une fois) 20 piles d'affilée ? ou 200 fois ?

1.3 Quelques exemples

i. Si $(\Omega, \mathcal{A})$ est un espace mesurable, $x \in \Omega$, δ_c est une mesure de probabilité associée à la constante c.

ii. Si $\Omega = \{1, 2, 3, 4, 5, 6\}$ muni de la tribu de ses parties,

$$\mathbb{P} := \frac{1}{6} \sum_{i=1}^{6} \delta_i$$

est une mesure de probabilité qui sert à modéliser le jeu de dés.

iii. Si $\Omega = \{0, 1\}$ muni de la tribu de ses parties, $0 \leq p \leq 1$,

$$\mathbb{P} := p\delta_1 + (1-p)\delta_0$$

est une mesure de probabilité qu sert à modéliser le résultat d'un lancé d'une pièce biaisée.

iv. Si $(\Omega, \mathcal{A})$ est un espace mesuré, $(x_n)_n$ une suite de points de Ω et $(p_n)_n$ une suite de nombres réels positifs tels que $\sum_n p_n = 1$,

$$\mathbb{P} := \sum_n p_n \delta_n$$

est une mesure de probabilité. Si $\Omega = \{x_n, \quad n \in \mathbb{N}\}$ et $\mathcal{A} := \mathcal{P}(\Omega)$ alors elle est absolument continue par rapport à la mesure de comptage.

Les Cas particuliers suivants sont très utilisés au moins dans les exercices.

- Si $x_n = n$ et $p_n = e^{-\lambda}\frac{\lambda^n}{n!}$ $\mathbb{P}$ est la loi de Poisson. Elle modélise par exemple le nombre de sinistres pris en charge par une compagnie d'assurance.
- x_n et $p_n = (1-p)p^n$ $n \geq 1$, $\mathbb{P}$ est la loi de géométrique. Elle modélise le nombre de lancé nécessaires pour obtenir un pile.
- $x_k = k$, $p_k = C_n^k p^k (1-p)^{n-k}$ pour $0 \leq k \leq n$et $p \in [0,1]$ $\mathbb{P}$ est la loi Binomiale $\mathcal{B}(n,p)$. Elle modélise le nombre de piles obtenus lors de n lancers de pièces.

v. $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ et f une fonction mesurble positive telle que $\int_{\mathbb{R}} f dm = 1$, $\mathbb{P}(A) = \int \mathbf{1}_A f dm$.

Définition 4. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probablisé, un élément A de $\mathcal{A}$ est un événement.
Un événement a lieu presque surement si et seulement si $\mathbb{P}(A) = 1$.

Dans l'exemple du jet de dés, $\{1,3,5\}$ est un événement.
Si $\Omega = \mathbb{R}^2$, $\mathcal{A} = \mathbb{R}$ et $\mathbb{P}$ est la mesure uniforme sur le disque $D = \{(x,y)|,\ x^2 + y^2 = 1\}$,

$$\mathbb{P}(A) := \frac{1}{\pi} \int \mathbf{1}_{A \cap D} dm$$

(notée $\mathcal{U}(D)$) alors $D \setminus \{(x,0),\quad x \in \mathbb{R}\}$ a lieu presque surement. Si vous lancez une flechette, presque surement vous ne toucherez pas l'axe horizontal.

1.4 Variables aléatoires

Définition 5. Une variable aléatoire est une fonction mesurable sur $(\Omega, \mathcal{A}, \mathbb{P})$.

Les variables aléatoires (v.a. en abrégé) sont très souvent notées avec des lettres capitales. Dans ce cours, elles seront soit à valeur dans $\mathbb{R}$ ou dans $\mathbb{R}^d$ et appellées v.a. réelles ou vectorielles.
Si on lance n fois une pi'ece, le nombre de piles obtenus défini une variable aléatoire.
Si X et Y sont deux v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ telles que $\{\omega,\quad X(\omega) = Y(\omega)\}$ ait lieu $\mathbb{P}$ presque surement, on écrit $X = Y$ $\mathbb{P}$ ps.

i. $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}([0,1])$ et $\mathbb{P} = m$ $\mathbb{P}$ est appelée la probabilité uniforme sur $[0,1]$ notée $\mathcal{U}([0,1])$ et l'application identité de $(\Omega, \mathcal{A})$ dans lui même est mesurable, c'est une v.a.

ii. Sur $(\Omega, \mathcal{A}, \delta_x)$ où $x \in \Omega$, toute v.a. est δ_x constante presque surement.

En pratique l'espace $(\Omega, \mathcal{A}, \mathbb{P})$ est un peu mythique. Une loi peut toujours être représentée comme une mesure image par une application mesurable (l'identité est une v.a.). Cela signifie qu'une mesure de probabilité est toujours la loi d'une variable. La représentation d'une loi par une v.a. n'est pas unique. La loi de Bernoulli peut être reprşentée par*
- $\Omega = \{0,1\}$ muni de la tribu de ses parties, et $\mathbb{P} = p\delta_1 + (1-p)\delta_0$,
- $\Omega = [0,1]$ muni de sa tribu borélienne, $\mathbb{P}$ la mesure de Lebesgue sur $[0,1]$ et $X = \mathbf{1}_{[0,p]}$, la mesure image $\mathbb{P}^X$ est la loi de Bernoulli de paramètre p.

Définition 6. Une loi est dite discrète si c'est une combinaison linéaire finie ou dénombrable de masses de Dirac,

$$\mathbb{P} := \sum_i p_i \delta_{x_i};$$

une loi discrète ne prend presque sûrement qu'un nombre au plus dénombrable de valeurs.
Si $\mathbb{P}$ est absolument continue par rapport à μ et si X est de loi $\mathbb{P}$, X admet pour densité f par rapport à $\mathbb{P}$ où $f = \frac{d\mathbb{P}}{\mu}$. Si μ est la mesure de Lebesgue, on dit que f est de densité f.

1.5 Fonctions de répartitions

Définition 7. Soit X une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$. La fonction de répartition de X, notée F^X est celle de la loi $\mathbb{P}^X$, c'est-à dire

$$F^X(t) = \mathbb{P}^X(]-\infty, t]) = \mathbb{P}(X \leq t), \quad t \in \mathbb{R}.$$

Elle vérifie
- $0 \leq F^X \leq 1$,
- F^X est continue à droite, limitée à gauche,
- $\lim_{t\to-\infty} F^X(t) = 0$, $\lim_{t\to\infty} F^X(t) = 1$.

Réciproquement, si F est une fonction vérifiant les propriétés ci-dessus alors F est la fonction de répartition d'une loi de probabilité.

Proposition 8. La fonction de répartition caractérise la loi, $F^X = F^Y$ si et seulement si $\mathbb{P}^X = \mathbb{P}^Y$.

Une fonction de répartition admet au plus un nombre dénombrable de points de discontinuité.

Une va.a. de fonction de répartition dérivable de dérivée f et égale à l'intégrale de sa dérivée est à densité f.

Exemple 9.

i. La fonction F définie par $F(t) := \begin{cases} 1 - e^{-\theta t}, & \text{si } t > 0, \\ 0, & \text{sinon} \end{cases}$ où $\theta > 0$ est la fonction de répartition de la loi exponentielle de densité $\theta e^{-\theta t}\mathbf{1}_{]0,+\infty[}(t), \quad t \in \mathbb{R}$,

ii. $F = \mathbf{1}_{[c,\infty[}$ est la fonction de répartition de la masse de Dirac en c, δ_c.

iii. La fonction de répartition de la loi binomiale $\mathcal{B}(n, p)$, $n \in \mathbb{N}^*$, $p \in [0, 1]$ est

$$\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} \mathbf{1}_{[k,+\infty[}(t),$$

iv. La fonction de répartition F de la loi gaussienne $\mathcal{N}(0, 1)$ de densité $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ n'est pas calculable explictement mais est tabulée. Elle vérifie si X est une v.a. de fonction de répartition F, alors $Y = \sigma X + m$, $\sigma > 0$, $m \in \mathbb{R}$ est une v.a. de fonction de répartition $F^Y(t) = F(\frac{t-m}{\sigma})$, $\quad t \in \mathbb{R}$. La loi de Y est notée $\mathcal{N}(m, \sigma^2)$ est appelée loi normale de moyenne m de vraiance σ^2 et a pour densité $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma}}, \quad x \in \mathbb{R}$.

v. La fonction $F(t) := t\mathbf{1}_{[0,1[} + \mathbf{1}_{[0,\infty[}$, $\quad t \in \mathbb{R}$ est la fonction de répartition de la mesure uniforme sur $[0, 1]$.

Proposition 10. Soit F une fonction de répartition, sa fonction quantile est

$$F^{\leftarrow}(u) := \inf\{x \in \mathbb{R} | F(x) \geq u\}, \quad u \in]0, 1[.$$

Si U est une v.a. de loi uniforme sur $[0, 1]$ alors $F^{\leftarrow}(U)$ a pour fonction de répartition F.

2 Vecteurs aléatoires

Dans ce paragraphe d est un entier supérieur ou égal à 2.

Définition 11. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Une v.a. $X = (X_1, ..., X_d)$ à valeurs dans $\mathbb{R}^d$ muni de sa tribu borélienne est un vecteur aléatoire.

Définition 12. La fonction de répartition de X est la fonction

$$t = (t_1, ..., t_d) \in \mathbb{R}^d \mapsto F^X(t) := \mathbb{P}(X_1 \leq t_1, ..., X_d \leq t_d).$$

La loi de X_i est appelée la $i$ième marginale sa fonction de répartition est

$$F^{X_i}(t) = \lim_{t_j \to \infty, j \neq i} \mathbb{P}(X_1 \leq t_1, ..., X_d \leq t_d).$$

La loi d'un vecteur détermine celles de ses marginales, la réciproque est fausse en générale.

Exemple 13. 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.

$X_1 = i \backslash X_2 = j$	0	1	2	marg. 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	$\frac{3}{16}$	$\frac{3}{16}$
marg. X_2	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{1}{2}$	1

$Y_1 = i \backslash Y_2 = j$	0	1	2	marg. Y_1
0	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{3}{8}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{7}{16}$
2	0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$
marg. Y_2	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{1}{2}$	1

Exemple 14. La loi gaussienne centrée réduite vectorielle de dimension 2 a pour densité

$$\frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}, \quad \forall (x, y) \in \mathbb{R}^2.$$

Si $Z = (X, Y)$ est une v.a. bidiemnsionnelle de densité $f(x, y)$, alors
– X est une v.a. de densité $f^X(x) := \int_{\mathbb{R}} f(x, y)dy$,
– Y est une v.a. de $f^X(x) := \int_{\mathbb{R}} f(x, y)dy$.

Proposition 15. Soient (X, Y) deux vecteurs aléatoires définis sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$, alors $\mathbf{P}^X = \mathbf{P}^Y$ si et seulement si $F^X = F^Y$.

3 Moyenne et inégalités

Définition 16. Soit X est une v.a.r. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$, si X intégrable (i.e. $\int_\Omega |X(\omega)|d\mathbb{P}(\omega) < +_infty$, $\int_\Omega X d\mathbb{P}$ est appelée *espérance* de X et notée $\mathbb{E}(X)$.

On dit que X est centrée si elle est intgrable et $\mathbb{E}(X) = 0$.

Lemme 17. Théorème de transport Soit X un v.$\vec{a}$. à valeurs $\mathbb{R}^d$ de loi P_X et $\Psi : \mathbb{R}^d \to \mathbb{R}$ 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 18. Ce résultat s'étend sans difficulté à toute mesure image (non nécessairement une probabilité).

Remarque 19. i. Si X est un v.a. de $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)$ et A un borélien de $\mathbb{R}^d$, alors $\mathbf{1}_A(X)$ est mesurable et

$$\mathbb{E}(\mathbf{1}_A(X)) = \mathbb{P}(X \in A).$$

ii. Soit X est un vecteur aléatoire à valeurs dans $\mathbb{R}^d$, de densité f, soit $h : \mathbb{R}^d \to \mathbb{R}^d$ une bijection C^1 de jacobien $J_h(x) \neq 0$ pour tout x, alors $Y = h(X)$ est un V.a. de densité

$$g(y) := \frac{1}{|J_h(y)|} f \circ h^{-1}(y), \quad y \in \mathbb{R}.$$

En effet en utilisant successivement le théorème de transport, le théorème de Fubini transport puis théorème de transport

$$\begin{aligned}\mathbb{E}(\Phi \circ h(X)) &= \int_{\mathbb{R}^d} \Phi \circ h(x) d\mathbb{P}^X(x), \\ &= int_{\mathbb{R}^d} \Phi \circ h(x) f(x) dm(x) \\ &= \int_{\mathbb{R}^d} \Phi(y) \frac{1}{|J_h(y)|} f \circ h^{-1}(y) dm(y).\end{aligned}$$

Si la densité f est nulle hors d'un ouvert U de $\mathbb{R}^d$, la même formule s'applique si h est définie sur U.

Dans la pratique la loi de X se décompose le plus souvent en une partie absoluement continue par rapport à la mesure de Lebegue et une partie absoluement continue par rapport à la mesure de comptage.

Si $\mathbb{P}^X$ est absoluement continue par rapport à la mesure de Lebesgue de densité f, sous les condition d'intégrabilité du Théorème de transport,

$$\mathbb{E}(\Phi(X)) = \int_{\mathbb{R}^d} \Phi(x) f(x) dm(x).$$

Si au contraire $\mathbb{P}^X := \sum_n p_n \delta_{x_n}$,

$$\mathbb{E}(\Phi(X)) = \sum_n f(x_n) p_n.$$

Exemple 20. i. Si X suit une Bernoulli de paramètre $1/2$, $\mathbb{E}(X) = \frac{1}{2}$,

ii. Si X suit une loi binomiale $\mathcal{B}(n,p)$ $\mathbb{E}(X) = np$,

iii. Si X suit une loi exponentielle de paramètre θ, $\mathbb{E}(X) = \frac{1}{\theta}$,

iv. Si X suit une loi de Poisson de paramètre λ, $\mathbb{E}(X) = \lambda$,

v. Si X suit une loi normale de moyenne m et de variance σ^2, $\mathbb{E}(X) = m$.

vi. si$X$ suit une loi de Cauchy, X n'est pas intégrable.

Remarque 21 (Caractérisation de la loi)**.** Pour déterminer la loi d'un v.$\vec{a}$. X, fonction d'autres v.$\vec{a}$. 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 $\Psi : \mathbb{R}^d \to \mathbb{R}$ 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 $\Psi = \mathbb{1}_B$, $B \in \mathcal{B}(\mathbb{R}^d)$, 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 $\Psi : \mathbb{R}^d \to \mathbb{R}$ borélienne,

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

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 $\Psi = \mathbb{1}_B$, $B \in \mathcal{B}(\mathbb{R}^d)$. Alors $\Psi(X) = \mathbb{1}_{\{X \in B\}}$ et

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

(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 :

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

□

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

Inégalité de Jensen : Soit X une v.a.r. intégrable et $\varphi : \mathbb{R} \to \mathbb{R}$ 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 $\varphi : I \to \mathbb{R}$ est convexe sur un intervalle I contenant x_0 les taux d'accroissement $x \in I \backslash \{x_0\} \mapsto \frac{\varphi(x)-\varphi(x_0)}{x-x_0} \in \mathbb{R}$ 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 intégre 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 $x_0 = \mathbb{E}(X)$. □

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

$$\mathrm{var}(X) = \mathbb{E}\Big((X - \mathbb{E}(X))^2\Big) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

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

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

$$\mathrm{Cov}(X, Y) = \mathbb{E}\Big((X - \mathbb{E}(X))\,(Y - \mathbb{E}(Y))\Big) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y).$$

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

3.3.2 Extension aux vecteurs aléatoires

Soit $\vec{X} = (X_1, \ldots, X_d)$ un v.ā. 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.ā. 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 $\mathrm{Cov}(\vec{X}) = (\mathrm{Cov}(X_i,X_j))_{i,j=1,\dots,d} \in M_d(\mathbb{R})$.

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

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

iii. Si $\vec{X} \in \mathbb{L}^2(\Omega;\mathbb{R}^d)$, $\mathrm{Cov}(\vec{X}) \in M_d(\mathbb{R})$ 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.ā. $A\vec{X}$ est dans $\mathbb{L}^2(\Omega;\mathbb{R}^n)$. Et on a

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

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

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

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

4 Fonction caractéristique

Une autre caractérisation d'une mesure de probabilité utile en pratique est la transformée de Fourier.

Définition 26. Soit X un vecteur aléatoire sur $(\Omega,\mathcal{A},\mathbb{P})$ à valeur dans $\mathbb{R}^d$. La fonction caractéristique de X ou de la loi de X est

$$\varphi^X(t) := \mathbb{E}\left(e^{i\langle t,X\rangle}\right) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} dP^X(x), \quad t_i n\mathbb{R}^d$$

La fonction caractéristique est à valeurs complexes, de module majoré pae 1, et $\varphi^X(0) = 1$. Si la loi de X a une densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, alors

$$\varphi^X(t) = \hat{f}(-t), \quad t \in \mathbb{R}^d.$$

Théorème 27. Si X et Y sont deux v.a. de loi P^X et P^Y telles que $\varphi^X = \varphi^Y$ alors $P^X = P^Y$.

Théorème 28. Soit X une v.a. telle 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 donnée par

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

4.1 Génération des moments via la fonction caractéristique

Théorème 29. Soit X une v.a.r. de loi P_X et de fonction caractéristique φ_X.

i. Si $\mathbb{E}(|X|^n) < +\infty$, alors φ_X est n fois dérivable et on a

$$\forall 0 \leq k \leq n,\, \forall t \in \mathbb{R}, \quad \varphi_X^{(k)}(t) = i^k \int_{\mathbb{R}} x^k e^{itx} P_X(dx). \tag{2}$$

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

ii. Réciproquement, si, pour un n **pair**, φ_X est n fois dérivable en 0 , alors $\mathbb{E}(|X|^n) < +\infty$.

Démonstration. i. L'identité (2) se montre par récurrence : c'est une conséquence immédiate du théorème de dérivation sous le signe somme (c.v. dominée). Cela a déjà été démontré pour la T.F. d'une fonction $\mathbb{L}^1$.

ii. Par récurence sur $k = n/2 \geq 1$. Le cas $k = 1$ contient essentiellement toute la difficulté. Supposons donc $\phi = \phi_X$ deux fois dérivable en 0. Par la formule de Taylor-Young,

$$\phi(t) - \phi(0) - \phi'(0)t - \phi''(0)\frac{t^2}{2} = \circ(t^2)$$

quand t tend vers 0. En appliquant cela à t et $-t$ et en sommant, on obtient donc

$$\frac{\phi(t) + \phi(-t) - 2\phi(0)}{t^2} \to \phi''(0). \tag{3}$$

Remarquons que

$$\frac{2\phi(0) - \phi(t) - \phi(-t)}{t^2} = \mathbb{E}\left(\frac{2 - (e^{itX} + e^{-itX})}{t^2}\right) = 2\,\mathbb{E}\left(\frac{1 - \cos(tX)}{t^2}\right) \in \mathbb{R}_+$$

ce qui montre, à la limite quand $t \to 0$, que $-\phi''(0) \in [0, +\infty[$. Enfin, le lemme de Fatou pour $\frac{1-\cos(tX)}{t^2} \geq 0$ permet d'affirmer que

$$+\infty > -\phi''(0) = 2\lim_{t\to 0}\mathbb{E}\left(\frac{1 - \cos(tX)}{t^2}\right) \geq 2\,\mathbb{E}\left(\lim_{t\to 0}\frac{1 - \cos(tX)}{t^2}\right) = \mathbb{E}(X^2).$$

C'est le résultat souhaité pour $k = 1$.

Supposons maintenant le résultat vrai pour $n = 2k$ et montrons le pour $n = 2k + 2$. On suppose donc que $\phi^{(2k+2)}(0)$ existe. Soit $\psi = \phi^{(2k)}$. ψ est donc dérivable autour de 0 et $\psi''(0)$ existe : on peut donc appliquer (3) à ψ. Mais, comme $\phi^{(2k)}(0)$ existe, on sait par hypothèse de récurrence que $\mathbb{E}(X^{2k}) < +\infty$. Et donc d'après la partie i., pour tout t, $\psi(t) = (-1)^k\mathbb{E}(X^{2k}e^{itX})$. On peut alors conclure comme précédemment :

$$\begin{aligned} +\infty > -(-1)^k\psi''(0) &= \lim_{t\to 0}(-1)^k\frac{2\psi(0) - \psi(t) - \psi(-t)}{t^2} \\ &= 2\lim_{t\to 0}\mathbb{E}\left(X^{2k}\frac{1 - \cos(tX)}{t^2}\right) \geq 2\,\mathbb{E}\left(X^{2k}\lim_{t\to 0}\frac{1 - \cos(tX)}{t^2}\right) = \mathbb{E}(X^{2k+2}). \end{aligned}$$

□

Exemple 30. Si $X \sim \mathcal{N}(0, \sigma^2)$, alors, pour tout $n \in \mathbb{N}$, $\mathbb{E}(X^{2n}) = \frac{\sigma^{2n}}{2^n}\frac{(2n)!}{n!}$.
Pour le voir, on développe $\Phi_X(t) = e^{-\frac{\sigma^2 t^2}{2}}$ *en série entière. On obtient*

$$\Phi_X(t) = \sum_{k=0}^{+\infty}(-1)^k\frac{\sigma^{2k}}{2^k}\frac{t^{2k}}{k!} = \sum_{n=0}^{+\infty}\Phi_X^{(n)}(0)\frac{t^n}{n!}.$$

D'où l'on tire $\Phi_X^{(2k)}(0) = (-1)^k\frac{\sigma^{2k}}{2^k}\frac{(2k)!}{k!}$ *(et* $\Phi_X^{(2k+1)}(0) = 0$*).*

4.2 Autres fonctions pouvant caractériser la loi d'une variable aléatoire

4.2.1 Transformée de Laplace

Définition 31. Si X est une variable aléatoire à valeurs dans $\mathbb{R}^d$, sa transformée de Laplace est

$$L^X(s) = \mathbb{E}(e^{\langle s,X\rangle})$$

définie pour les valeurs de s telles que $e^{\langle s,X\rangle}$ soit intégrable.

Proposition 32. Si X est une variable aléatoire réelle telle que e^{tX} soit intégrable pour tout $t \in I$ où I est un intervalle ouvert contenant 0, alors L^X est définie sur I, analytique au voisinage de 0, X admet des moments de tous ordre

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

En particulier, $(L^X)^{(n)}(0) = \mathbb{E}(X^n)$.
De plus si Y est une variable aléatoire telle que $L^X = L^Y$ sur J un intervalle ouvert contenant 0 alors $\mathbb{P}_X = \mathbb{P}_Y$.

Exemple 33. Soit X une variable aléatoire admettant Φ_1 comme densité par rapport à la mesure de Lebesgue, alors la transformée de Laplace de X^2 est $L^{X^2}(t) = \frac{1}{\sqrt{1-t}}, \quad t < 1$.

4.2.2 Fonction génératrice des moments

Définition 34. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$, la fonction génératrice des moments G^X est définie par $G^X(s) = \sum_{n=0}^{\infty} s^n \mathbb{P}(\{\omega, X(\omega) = n\}) = \mathbb{E}(s^X), \quad s \in]-1,1]$.

Proposition 35.

i. La fonction G^X est analytique sur $]-1,1[$.

ii. Si X et Y sont deux variables aléatoires à valeurs dans $\mathbb{N}$, telles que $G^X(s) = G^Y(s)$, $s \in]-1,1[$, alors $\mathbb{P}_X = \mathbb{P}_Y$.