Chapitre 1

Formalisme des probabilités

1.1 Rappels de cours

Le formalisme de la théorie des probabilités utilise les outils de la théorie de la mesure en adoptant un vocabulaire spécifique aux probabilités. Il s'agit donc dans un premier temps de préciser ce nouveau langage avant d'introduire des concepts propres à cette théorie. Pour une justification du choix du formalisme et de sa signification en tant que modèle de la réalité, on pourra consulter, par exemple, l'ouvrage [9], chapitre 5.

1.1.1 Variable aléatoire et loi d'une variable aléatoire

Une **probabilité** $\mathbb{P}$ sur un espace mesurable $(\Omega, \mathcal{F})$ est une mesure positive sur $(\Omega, \mathcal{F})$ telle que $\mathbb{P}(\Omega) = 1$. On dit alors que $(\Omega, \mathcal{F}, \mathbb{P})$ est **un espace de probabilité.** Ω est appelé l'espace des **réalisations** ou des **épreuves.** Les éléments de la tribu $\mathcal{F}$ sont **les événements de Ω.** Une propriété vraie $\mathbb{P}$-presque-partout est dite **$\mathbb{P}$-presque-sûre,** en abrégé $\mathbb{P}$-p.s. ou plus simplement p.s.

Dans la suite $(\Omega, \mathcal{F}, \mathbb{P})$ désignera toujours un espace de probabilité et $(E, \mathcal{E})$ un espace mesurable.

Une application $(\mathcal{F}, \mathcal{E})$-mesurable X de Ω dans E s'appelle une **variable aléatoire sur** $(\Omega, \mathcal{F}, \mathbb{P})$ **à valeurs** E , en abrégé v.a. à valeurs E.

Si $E = \mathbb{R}$ (resp. $E = \overline{\mathbb{R}}$, $E = \mathbb{R}^d$ où $d \in \mathbb{N}^*$) on dit que X est une v.a. **réelle** (resp. **numérique, de dimension** d), en abrégé v.a.r., v.a.n., v.a. dans $\mathbb{R}^d$.

D'après la théorie de la mesure on a :

Proposition 1.1. *Une application X de Ω dans $\mathbb{R}^d$ est une v.a. si, et seulement si, ses composantes sont des v.a.r.*

L'image de la probabilité $\mathbb{P}$ par la v.a. X est une probabilité sur $(E, \mathcal{E})$, notée $\mathbb{P}_X$, et appelée la **loi (de probabilité) de la v.a. X par rapport à $\mathbb{P}$.** Pour tout $A \in \mathcal{E}$, on a donc $\mathbb{P}_X(A) = \mathbb{P}(X \in A) := \mathbb{P}(\{\omega \in \Omega \;/\; X(\omega) \in A\})$.

Par le théorème du transfert il vient :

Proposition 1.2. *Soient X une v.a. à valeurs dans E, f une application $\mathcal{E}$-mesurable de E dans $\mathbb{R}^+$, alors,*

$$\int_\Omega f(X(\omega))d\mathbb{P}(\omega) = \int_E f(x)d\mathbb{P}_X(x).$$

L'existence des v.a. de loi donnée a priori est assurée par :

Proposition 1.3. *Soient $(E, \mathcal{E})$ un espace mesurable et μ une probabilité sur $(E, \mathcal{E})$. Alors il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans E et de loi de probabilité μ.*

On rappelle le théorème d'extension de Hahn-Carathéodory adapté au cas des probabilités (cf. [1], chapitre 2, rappels de cours) :

Proposition 1.4. Théorème de prolongement
Soient E un ensemble, $\mathcal{A}$ une famille de parties de E contenant E, stable par réunion finie et passage au complémentaire (i.e. $\mathcal{A}$ est une **algèbre sur** *E), μ une application de $\mathcal{A}$ dans $[0, +\infty]$ vérifiant $\mu(E) = 1$ et*

$$\mu\left(\bigcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mu(A_n)$$

pour toute suite $(A_n)_{\mathbb{N}}$ d'éléments deux à deux disjoints de $\mathcal{A}$ avec $\bigcup_n A_n \in \mathcal{A}$. Alors il existe une unique probabilité $\widehat{\mu}$ sur $(E, \sigma(\mathcal{A}))$ telle que, pour tout $A \in \mathcal{A}$, $\widehat{\mu}(A) = \mu(A)$ où $\sigma(\mathcal{A})$ désigne la tribu engendrée par $\mathcal{A}$.

Le résultat suivant est trés utile dans la recherche des lois de v.a.r. :

Proposition 1.5. *Soit $(E, \mathcal{E})$ un espace mesurable et μ une probabilité sur $(E, \mathcal{E})$. Une v.a. X à valeurs dans E sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ admet μ pour loi de probabilité si, et seulement si, pour toute application $\mathcal{E}$-mesurable et positive h de E dans $[0, +\infty]$,*

$$\int_\Omega h(X(\omega))d\mathbb{P}(\omega) = \int_E h(x)d\mu(x).$$

Une mesure μ sur $\mathbb{R}^d$ est dite **discrète** si elle peut s'écrire sous la forme $\mu = \sum_{n\in\mathbb{N}} p_n \delta_{e_n}$ où $(p_n)_{\mathbb{N}}$ est une suite de réels positifs ou nuls et $(e_n)_{\mathbb{N}}$ une suite d'éléments de $\mathbb{R}^d$.

Une v.a. X à valeurs dans $\mathbb{R}^d$ est dite **discrète** si sa loi est discrète. Dans ce cas, pour tout $n \in \mathbb{N}$, $p_n = \mathbb{P}(X = e_n)$.

Les v.a. de Poisson, de Bernoulli, binomiale, hypergéométrique sont des exemples de v.a.r. discrètes. De façon précise les v.a.r. discrètes sont les v.a.r. presque-sûrement à valeurs dans un ensemble dénombrable :

Proposition 1.6. (cf. exercice 1.1, page 16)
Une v.a. X à valeurs dans $\mathbb{R}^d$ est discrète si, et seulement si, il existe une partie D de $\mathbb{R}^d$ dénombrable telle que $\mathbb{P}(X \in D) = 1$. Dans ce cas la loi de la v.a.r. X s'écrit

$$\mathbb{P}_X = \sum_{x\in D} \mathbb{P}(X = x)\,\delta_x.$$

On dit aussi que la loi de X est **portée par l'ensemble** D pour exprimer l'égalité $\mathbb{P}(X \in D) = 1$.

Les v.a.r. discrètes constituent une famille de v.a.r importante dans les applications des probabilités, une autre classe de v.a.r. trés importante aussi est celle des v.a.r. absolument continues.

On dit qu'une mesure μ sur $\mathbb{R}^d$ est **absolument continue sur** $\mathbb{R}^d$ ou encore **à densité sur** $\mathbb{R}^d$ si elle admet une densité f par rapport à la mesure de Lebesgue $\lambda^{(d)}$ de $\mathbb{R}^d$, i.e. f est une application de $\mathbb{R}^d$ dans $[0, +\infty[$ intégrable au sens de Lebesgue sur $\mathbb{R}^d$ et telle que, pour tout borélien A de $\mathbb{R}^d$, $\mu(A) = \int_A f d\lambda^{(d)}$. On note alors $\mu = f \cdot \lambda^{(d)}$ ou $d\mu(x) = f(x)d\lambda^{(d)}(x)$ où, par abus, $d\mu(x) = f(x)dx$.

On dit qu'une v.a. à valeurs dans $\mathbb{R}^d$ est **absolument continue sur** $\mathbb{R}^d$ ou encore **à densité sur** $\mathbb{R}^d$ si sa loi est absolument continue sur $\mathbb{R}^d$.

Les v.a.r. de Gauss-Laplace, exponentielle, uniforme sur un intervalle de $\mathbb{R}$, de Cauchy sont des exemples de v.a.r. absolument continues sur $\mathbb{R}$. Pour un inventaire de lois classiques se reporter au formulaire du chapitre 9, page 225, de cet ouvrage.

Si μ est une probabilité sur $\mathbb{R}$, on appelle **fonction de répartition de** μ (en abrégé f.r.), et on note F_μ, l'application de $\mathbb{R}$ dans $[0, 1]$ définie, pour tout $x \in \mathbb{R}$, par $F_\mu(x) := \mu(]-\infty, x])$. Si X est une v.a.r., la f.r. de la loi de X est appelée **fonction de répartition de** X , on la note alors F_X.

La f.r. d'une probabilité a les propriétés suivantes :

Proposition 1.7. *La f.r. de* μ *est croissante, continue à droite sur* $\mathbb{R}$ *avec* $\lim_{x\to+\infty} F_\mu(x) = 1$ *et* $\lim_{x\to-\infty} F_\mu(x) = 0$*. Pour tout* $a \in \mathbb{R}$*,* F_μ *est continue en* a *si, et seulement si,* $\mu(\{a\}) = 0$*. De plus l'ensemble des points de discontinuité d'une f.r. est dénombrable.*

Réciproquement par application du théorème de prolongement on obtient (cf. [1], exercice 2.6) :

Proposition 1.8. *Soit* F *une application de* $\mathbb{R}$ *dans* $[0, 1]$*, croissante, continue à droite sur* $\mathbb{R}$ *avec* $\lim_{x\to+\infty} F(x) = 1$ *et* $\lim_{x\to-\infty} F(x) = 0$*. Alors il existe une probabilité et une seule* μ *définie sur* $\mathbb{R}$ *dont la fonction de répartition est égale à* F*.*

De plus, la probabilité μ *est déterminée, pour tous réels* a *et* b *tels que* $a < b$*, par*

$$\mu(]a, b]) = F(b) - F(a).$$

Proposition 1.9. (cf. exercice 1.16, page 21)
Soit F *une application de* $\mathbb{R}$ *dans* $[0, 1]$*, croissante, continue à droite sur* $\mathbb{R}$ *avec* $\lim_{x\to+\infty} F(x) = 1$ *et* $\lim_{x\to-\infty} F(x) = 0$*. Alors il existe un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ *et une v.a.r.* X *sur* $(\Omega, \mathcal{F}, \mathbb{P})$ *telle que* $F_X = F$*.*

On notera que dans ce cas la v.a.r. X n'est pas unique.

Le résultat suivant est souvent utilisé pour prouver l'unicité de certaines probabilités

Proposition 1.10. Lemme d'unicité des probabilités
Soit $(\Omega, \mathcal{F})$ *un espace mesurable,* $\mathcal{C}$ *une famille stable par intersection finie et engendrant la tribu* $\mathcal{F}$*. Si deux probabilités sur* $(\Omega, \mathcal{F})$ *coïncident sur* $\mathcal{C}$*, alors elles sont égales.*

1.1.2 Moments d'une variable aléatoire

Comme $\mathbb{P}$ est une mesure finie sur $(\Omega, \mathcal{F})$ on a, pour tout p et q dans $[1, +\infty[$ avec $p < q$, la chaîne d'inclusions :

$$L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \subseteq L^q(\Omega, \mathcal{F}, \mathbb{P}) \subseteq L^p(\Omega, \mathcal{F}, \mathbb{P}) \subseteq L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

Si X est une v.a.r. sur $(\Omega, \mathcal{F}, \mathbb{P})$ positive ou intégrable sur Ω, on appelle **espérance mathématique de** X , et on note $\mathbb{E}(X)$, la quantité (éventuellement infinie) $\mathbb{E}(X) := \int_\Omega X(\omega)d\mathbb{P}(\omega)$.

Plus généralement, si $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ et $p \in \mathbb{N}^*$, on appelle **moment d'ordre p de X** , resp. **moment centré d'ordre p de X,** le nombre réel

$$m_p := \int_\Omega X^p(\omega) d\mathbb{P}(\omega), \text{ resp. } m'_p := \int_\Omega (X(\omega) - m_1)^p d\mathbb{P}(\omega).$$

Le moment centré d'ordre 2 s'appelle **la variance de X,** et se note $\mathrm{Var}(X)$. Sa racine carrée positive s'appelle **l'écart type de X** et se note σ_X.

Si X et Y sont deux v.a.r. de carré intégrable, on appelle **covariance de X et Y,** et on note $\mathrm{Cov}(X,Y)$, le réel $\mathrm{Cov}(X,Y) := \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

Par application du théorème de transfert il vient :

Proposition 1.11. *Sous les conditions d'existence des différents moments,*

$$\begin{aligned}
m_1 &:= \mathbb{E}(X) = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \int_\mathbb{R} x\, d\mathbb{P}_X(x). \\
\sigma_X^2 &:= \mathrm{Var}(X) = \int_\Omega (X(\omega) - m_1)^2 d\mathbb{P}(\omega) = \int_\mathbb{R} (x - m_1)^2\, d\mathbb{P}_X(x). \\
\sigma_X^2 &= \mathrm{Cov}(X,X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2. \\
m_p &:= \int_\Omega X^p(\omega) d\mathbb{P}(\omega) = \int_\mathbb{R} x^p\, d\mathbb{P}_X(x). \\
m'_p &:= \int_\Omega (X(\omega) - m_1)^p\, d\mathbb{P}(\omega) = \int_\mathbb{R} (x - m_1)^p\, d\mathbb{P}_X(x).
\end{aligned}$$

Le résultat suivant établit un lien entre les moments et la fonction de répartition d'une v.a.r. positive. On notera qu'il n'est plus valable si la v.a.r. n'est plus supposée positive (cf. [1], exercice 6.8). On pourra également consulter l'exercice 1.7, page 18, de ce chapitre.

Proposition 1.12. *Si X est une v.a.r. positive de fonction de répartition F_X, alors, pour tout entier $n \geq 1$, l'application suivante*

$$t \in \mathbb{R} \mapsto n\, t^{n-1}\, \mathbb{P}(X > t)$$

est intégrable au sens de Riemann sur tout intervalle compact de $\mathbb{R}$ et vérifie la relation

$$\mathbb{E}[X^n] = \int_0^{+\infty} nt^{n-1} \mathbb{P}(X > t) dt = \int_0^{+\infty} nt^{n-1} (1 - F_X(t))\, dt.$$

Soit X une v.a. à valeurs dans $\mathbb{R}^d$ de composantes $X_1, X_2, \cdots, X_d$ intégrables sur Ω. On appelle **espérance de X,** et on note $\mathbb{E}(X)$, le vecteur de $\mathbb{R}^d$,

$$\mathbb{E}(X) := (\mathbb{E}(X_1), \mathbb{E}(X_2), \cdots, \mathbb{E}(X_d)).$$

Si X est une v.a. à valeurs matricielles, on appelle **espérance de** X, et on note $\mathbb{E}(X)$, la matrice dont les coefficients sont les espérances de ceux de X.

Soit X une v.a. à valeurs dans $\mathbb{R}^d$ de composantes $X_1, X_2, \cdots, X_d$ de carré intégrable sur Ω. On appelle **matrice de dispersion** de X l'espérance de la matrice carrée d'ordre d, $(X - \mathbb{E}(X))(X - \mathbb{E}(X))^*$, où * désigne l'opération de transposition des matrices. On dit aussi **matrice des covariances de** X et on la note D_X.

La terminologie est justifiée par la deuxième assertion de la proposition suivante (cf. exercice 1.12, page 20) :

Proposition 1.13. *Si X est une v.a. à valeurs dans $\mathbb{R}^d$ et M une matrice (déterministe) à coefficients réels à c lignes et d colonnes, alors*

1. *On a*

$$D_{[X-\mathbb{E}(X)]} = D_X, \quad \mathbb{E}(MX) = M\mathbb{E}(X), \quad D_{MX} = MD_XM^*.$$

2. *D_X est une matrice symétrique. D_X est* **de type positif** *i.e., pour tout $u \in \mathbb{R}^d$, $u^*D_Xu \geq 0$. Le coefficient d'indice (i,j) de D_X est la covariance* $\mathrm{Cov}(X_i, X_j)$ *des v.a.r. X_i et X_j. Les éléments diagonaux de D_X sont les variances des composantes de X.*
3. *Le déterminant de D_X est nul si, et seulement si, il existe des réels non tous nuls $a_0, a_1, \cdots, a_d$ tels que $a_1X_1 + a_2X_2 + \cdots + a_dX_d = a_0$ p.s.*

Donnons deux inégalités souvent utilisées en probabilité :

Proposition 1.14. Inégalité de Markov (cf. exercice 1.13, page 20)
Soit X une v.a. à valeurs dans $\mathbb{R}^d$ et φ une application borélienne de $\mathbb{R}^d$ dans $[0, +\infty[$, alors pour tout réel $a > 0$,

$$\mathbb{P}(\varphi(X) \geq a) \leq \frac{1}{a}\mathbb{E}(\varphi(X)).$$

Proposition 1.15. Inégalité de Bienaymé-Tchebycheff
Soit X une v.a.r. dans L^p où $p \in [1, +\infty[$, alors pour tout réel $a > 0$,

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq a\right) \leq \frac{1}{a^p}\,\mathbb{E}\left(|X - \mathbb{E}(X)|^p\right).$$

En particulier, si X est une v.a.r. de carré intégrable,

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq a\right) \leq \frac{1}{a^2}\,\mathrm{Var}(X).$$

1.1.3 Fonction caractéristique d'une variable aléatoire

La notion de fonction caractéristique d'une probabilité a été introduite et étudiée au chapitre 7 de [1]. On pourra s'y reporter pour plus de détails.

Rappelons cependant qu'on appelle **fonction caractéristique** d'une probabilité μ sur $\mathbb{R}^d$ (en abrégé f.c.), et on note Φ_μ, l'application de $\mathbb{R}^d$ dans $\mathbb{C}$ suivante

$$\Phi_\mu : u \in \mathbb{R}^d \mapsto \Phi_\mu(u) := \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} d\mu(x) \in \mathbb{C},$$

où $\langle u, x\rangle$ désigne le produit scalaire euclidien usuel des vecteurs x et u de $\mathbb{R}^d$.

Si X est une v.a. à valeurs dans $\mathbb{R}^d$ on appelle **fonction caractéristique de** X , la f.c. de sa loi $\mathbb{P}_X$. On la note Φ_X. On a donc la relation :

Proposition 1.16. *Pour toute v.a. X à valeurs dans $\mathbb{R}^d$ et pour tout vecteur $u \in \mathbb{R}^d$,*

$$\Phi_X(u) = \mathbb{E}\left(e^{i\langle u,x\rangle}\right).$$

La f.c. d'une v.a. à valeurs dans $\mathbb{R}^d$ est un outil trés utile dans l'identification des lois de v.a. en vertu du théorème suivant qui résulte des formules d'inversion de Perron-Sieltjès :

Proposition 1.17. Théorème d'injectivité
Deux v.a. à valeurs dans $\mathbb{R}^d$ ont la même loi si, et seulement si, elles ont la même fonction caractéristique.

La notion de f.c. est parfois utile dans les calculs de moments de v.a.r. grâce aux deux propositions suivantes :

Proposition 1.18. *Si est X une v.a.r. et n un entier naturel tels que $\mathbb{E}(|X^n|) < +\infty$, alors la f.c. Φ de X est de classe $\mathcal{C}^n$ sur $\mathbb{R}$ et sa dérivée d'ordre n est l'application*

$$\Phi^{(n)} : u \in \mathbb{R} \mapsto \Phi^{(n)}(u) = \int_{\mathbb{R}} (ix)^n e^{iux} d\mathbb{P}_X(x) = i^n \mathbb{E}\left(X^n e^{iuX}\right) \in \mathbb{C}.$$

En particulier, la dérivée d'ordre n de Φ en 0 est $\Phi^{(n)}(0) = i^n\mathbb{E}(X^n)$ et la f.c. Φ admet un développement limité en 0 à l'ordre n donné, pour tout réel u, par

$$\Phi(u) = \sum_{k=0}^{n} \frac{i^k\,\mathbb{E}\left(X^k\right)}{k!}\, u^k + o\left(u^n\right).$$

La proposition précédente montre qu'il y a un lien entre l'intégrabilité de la v.a.r. $|X|^n$ et la régularité de la f.c. de X : si la v.a.r. X est dans L^n, Φ est de classe $\mathcal{C}^n$. La réciproque est fausse. En effet, par exemple, il est possible de trouver une v.a.r. X non intégrable dont la f.c. est de classe $\mathcal{C}^1$ (cf. [8], exemple 8-7).

Cependant, on peut donner une réciproque partielle au résultat précédent sous la forme :

Proposition 1.19. *Soient X une v.a.r. et n un entier naturel tels que Φ soit dérivable en 0 à l'ordre n. Alors $\mathbb{E}\left(X^{2p}\right) < +\infty$ où $2p$ est le plus grand entier pair inférieur à n. En particulier, X admet des moments jusqu'à l'ordre $2p$.*

De plus, si Φ admet un développement limité en 0 à l'ordre n s'écrivant

$$\Phi(u) = 1 + \sum_{k=1}^{n} a_k u^k + o\left(u^n\right)$$

alors, pour tout entier k tel que $1 \leq k \leq 2p$, $\mathbb{E}\left(X^k\right) = (-i)^k a_k k!$.

1.2 Énoncés des exercices

Sauf mention contraire, les variables aléatoires considérées sont définies sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Exercice 1.1. Montrer qu'une v.a. X à valeurs dans $\mathbb{R}^d$ est discrète si, et seulement si, il existe une partie dénombrable D de $\mathbb{R}^d$ telle que $\mathbb{P}(X \in D) = 1$. Montrer que dans ce cas la loi de la v.a. X s'écrit

$$\mathbb{P}_X = \sum_{x \in D} \mathbb{P}(X = x)\, \delta_x.$$

Exercice 1.2. On suppose que X est une v.a.r. de loi uniforme sur un intervalle $[a, b]$, où a et b sont deux réels tels que $a < b$. Déterminer a et b sachant que $\mathbb{E}[X] = 10$ et $\mathrm{Var}(X) = 12$.

Exercice 1.3. Soit $(\Omega, \mathcal{F})$ un espace probabilisable et $(\mathbb{P}_i)_{i \in \mathbb{N}}$ une suite de probabilités définies sur $(\Omega, \mathcal{F})$. On considère une suite $(\alpha_i)_{i \in \mathbb{N}}$ de réels positifs telle que $\sum_{i=0}^{+\infty} \alpha_i = 1$ et on définit, pour tout $A \in \mathcal{F}$, $\mathbb{P}(A) := \sum_{i=0}^{+\infty} \alpha_i \mathbb{P}_i(A)$.

1. Montrer que $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{F})$.