

Cours de Probabilités

Jean-Yves DAUXOIS

Septembre 2013

Table des matières

1	Introduction au calcul des probabilités	7
1.1	Espace probabilisable et loi de variable aléatoire	8
1.1.1	Un exemple fondamental	8
1.1.2	Tribus	8
1.1.3	Mesures et probabilités	13
1.1.4	Variables aléatoires	18
1.1.5	Loi de probabilité d'une variable aléatoire	19
1.2	Conditionnement	19
1.2.1	Probabilité conditionnelle à un événement	20
1.2.2	Formule de Bayes	21
1.3	Indépendance en probabilité	22
1.3.1	Indépendance d'événements	22
1.3.2	Indépendance de tribus	25
1.3.3	Indépendance de variables aléatoires	25
1.3.4	Lien entre les différents types d'indépendance	26
1.4	Espace probabilisé produit	27
1.5	Loi conjointe d'un n -uplet de variables aléatoires indépendantes	29
2	Lois sur \mathbb{R} et lois sur \mathbb{R}^n	31
2.1	Variables aléatoires réelles	32
2.1.1	Fonction de répartition	32
2.1.2	Lois discrètes	35
2.1.3	Lois continues	39
2.1.4	Changement de variables	44
2.2	Vecteurs aléatoires	47
2.2.1	Fonction de répartition	47
2.2.2	Densité de probabilité	48
2.2.3	Loi conditionnelle et densité conditionnelle	50
2.2.4	Changement de variables	52

2.2.5	Indépendance	53
2.3	Extension de la notion de densité	57
2.3.1	Intégrale par rapport à une mesure	57
2.3.2	Absolue continuité d'une mesure par rapport à une autre. Densité	66
2.3.3	Mélange de lois	68
2.3.4	Densités conjointes, marginales et conditionnelles	69
3	Moments de variables aléatoires	71
3.1	Variables aléatoires réelles intégrables et espérance mathéma- tique	72
3.2	Moments de variables aléatoires réelles	75
3.2.1	Espace L^p	75
3.2.2	Espace L^2	77
3.3	Vecteurs aléatoires	80
3.3.1	Espérance mathématique	80
3.3.2	Covariance de deux v.a.r.	81
3.3.3	Matrice de covariance	83
3.3.4	Espérance conditionnelle	84
4	Caractérisation des lois : transformée de Laplace et fonction caractéristique	85
4.1	Transformée de Laplace	86
4.1.1	Variables aléatoires réelles	86
4.1.2	Vecteurs aléatoires	89
4.2	Fonction caractéristique	89
4.2.1	Intégrale d'une variable aléatoire complexe	89
4.2.2	Fonction caractéristique	90
5	Vecteurs gaussiens	93
5.1	Exemple fondamental	94
5.2	Définition	96
5.3	Propriétés des vecteurs aléatoires gaussiens	98
5.3.1	Transformation linéaire d'un vecteur gaussien	98
5.3.2	Vecteur gaussien et indépendance	98
6	Convergences	101
6.1	Convergence en loi	102
6.1.1	Définition	102
6.1.2	Caractérisation de la convergence en loi	102

6.1.3	Approximation de lois	104
6.2	Convergence en probabilité	107
6.2.1	Définition	107
6.2.2	Convergence en probabilité et convergence en loi . . .	111
6.3	Convergence presque sûre	112
6.3.1	Définition	112
6.3.2	Critères de convergence p.s.	112
6.3.3	Convergence presque sûre et convergence en probabilité	113
6.3.4	Loi forte des grands nombres	113
6.4	Convergence dans L^p	114
6.5	Résumé	115

Index	116
--------------	------------

Chapitre 1

Introduction au calcul des probabilités

1.1 Espace probabilisable et loi de variable aléatoire

1.1.1 Un exemple fondamental

Considérons le jeu du lancé d'un dé. Notons Ω l'ensemble de tous les résultats possibles (appelés aussi épreuves ou résultats élémentaires) de cette expérience aléatoire

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

On note $\omega = 3$ pour signifier que 3 est le résultat de l'épreuve.

Dans cette expérience aléatoire, on peut s'intéresser à des événements plus complexes qu'un simple résultat élémentaire. On peut, par exemple, considérer l'événement $A = \text{"le résultat est un nombre pair"}$ ou l'événement $B = \text{"le résultat est un nombre plus grand que 3"}$. On note \mathcal{A} l'ensemble de ces événements. Notons que l'on a toujours $\mathcal{A} \subset \mathcal{P}(\Omega)$, où $\mathcal{P}(\Omega)$ est l'ensemble des parties de Ω . Notons que l'inclusion précédente peut être stricte.

On dit que l'événement A s'est réalisé si le résultat de l'expérience ω est tel que $\omega \in A$.

Enfin, on peut donner à chaque événement une pondération ou encore une probabilité. Ainsi, si le dé n'est pas pipé, l'intuition nous dit que la probabilité d'avoir l'événement $A = \text{"le résultat est un nombre pair"}$ est $1/2$, i.e.

$$P(A) = \frac{1}{2}.$$

On peut bien sûr s'intéresser à la probabilité d'un événement $C = \text{"le résultat est un nombre pair plus grand ou égal à 4"}$. Remarquant que l'on a $C \subset A$, il sera alors naturel d'avoir

$$P(C) \leq P(A) = \frac{1}{2}.$$

Nous allons maintenant donner un formalisme plus mathématique à ce triplet fondamental (Ω, \mathcal{A}, P) que nous venons d'introduire.

1.1.2 Tribus

Tout phénomène aléatoire fait appel à deux ensembles de type différent.

- Un ensemble Ω , appelé espace fondamental ou univers, qui contient l'ensemble de tous les résultats possibles. Ces derniers sont également appelés épreuves.

- Une famille \mathcal{A} de parties (i.e. de sous ensembles) de Ω . Ces parties sont appelées des événements. On dit que l'événement A s'est réalisé si et seulement si le résultat ω de Ω qui s'est produit appartient à A .

En gardant en mémoire l'exemple fondamental, il est assez naturel de demander que l'ensemble \mathcal{A} vérifie un certain nombre de propriétés. En effet si A et B sont des événements de \mathcal{A} , on souhaite que les événements suivants le soient également.

- (i) $\bar{A} = \Omega \setminus A \in \mathcal{A}$. Si A s'est ou ne s'est pas réalisé, on doit pouvoir se prononcer sur l'événement complémentaire.
- (ii) $A \cup B \in \mathcal{A}$ et $A \cap B \in \mathcal{A}$. Si on peut dire que A s'est ou ne s'est pas réalisé, et de même pour B , on doit pouvoir dire si $A \cup B$ s'est ou ne s'est pas réalisé (et de même pour $A \cap B$).
- (iii) $A \setminus B \in \mathcal{A}$. On doit pouvoir dire si A s'est réalisé mais pas B .

Et plus généralement

- (iv) Si, pour tout n , on a $A_n \in \mathcal{A}$, alors on souhaite que

$$\bigcup_n A_n \in \mathcal{A} \text{ et } \bigcap_n A_n \in \mathcal{A}.$$

C'est pourquoi on demande à \mathcal{A} d'être une tribu.

Définition 1.1.1 *On dit qu'une famille \mathcal{A} de parties de Ω est une tribu si*

- (i) $\Omega \in \mathcal{A}$,
- (ii) \mathcal{A} est stable par passage au complémentaire, i.e.

$$A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A},$$

- (iii) \mathcal{A} est stable par réunion dénombrable, i.e.

$$(\forall n : A_n \in \mathcal{A}) \Rightarrow \left(\bigcup_n A_n \in \mathcal{A} \right).$$

Remarque. On montre facilement que ces conditions sont suffisantes pour que toutes celles précitées soient vérifiées. En effet:

$$\begin{aligned} A \cap B &= \overline{\bar{A} \cup \bar{B}} \in \mathcal{A} \\ A \setminus B &= A \cap \bar{B} \in \mathcal{A} \\ \emptyset &= \bar{\Omega} \in \mathcal{A} \end{aligned}$$

et si A_n appartient à \mathcal{A} , pour tout n , alors

$$\bigcap_n A_n = \overline{\left(\bigcup_n \bar{A}_n \right)} \in \mathcal{A}. \quad \diamond$$

Exemples de tribus.

- * $\mathcal{A} = \{\emptyset, \Omega\}$ est une tribu et est appelée tribu grossière. On ne peut en construire de plus petite.
- * $\mathcal{A} = \mathcal{P}(\Omega)$ est une tribu. C'est la tribu la plus fine, dans le sens où elle contient toutes les autres tribus sur Ω .
- * Soit A une partie de Ω . L'ensemble des parties

$$\mathcal{A} = \{\emptyset, A, \bar{A}, \Omega\}$$

est une tribu. \diamond

Définition 1.1.2 Lorsque \mathcal{A} est une tribu sur Ω , le couple (Ω, \mathcal{A}) est appelé espace probabilisable (ou mesurable).

Théorème 1.1.3 L'image réciproque d'une tribu par une application f est une tribu.

Preuve. Soit f une application de E vers F et \mathcal{F} une tribu sur F . Notons

$$\begin{aligned} \mathcal{E} = f^{-1}(\mathcal{F}) &= \{f^{-1}(B), \text{ pour } B \in \mathcal{F}\} \\ &= \{A \subset E \text{ tel que } f(A) \in \mathcal{F}\}. \end{aligned}$$

- * l'ensemble E est bien sûr élément de \mathcal{E} puisque $f(E) = F$.

* Soit A un élément de \mathcal{E} . Il existe donc un ensemble B dans \mathcal{F} tel que $A = f^{-1}(B)$. On peut alors écrire :

$$A = \{x \in E \text{ tel que } f(x) \in B\}.$$

D'où :

$$\begin{aligned} \bar{A} &= \{x \in E \text{ tel que } f(x) \notin B\} = \{x \in E \text{ tel que } f(x) \in \bar{B}\} \\ &= f^{-1}(\bar{B}). \end{aligned}$$

Or \bar{B} appartient à \mathcal{F} puisque \mathcal{F} est une tribu et \bar{A} est donc dans \mathcal{E} .

* Soient, pour tout n , A_n un élément de \mathcal{E} . Il existe donc pour tout n , un élément B_n de \mathcal{A} tel que $A_n = f^{-1}(B_n)$. D'où :

$$\begin{aligned} \bigcup_n A_n &= \{x \in E \text{ tel qu'il existe } n \text{ pour lequel } x \in A_n\} \\ &= \{x \in E \text{ tel qu'il existe } n \text{ pour lequel } f(x) \in B_n\} \\ &= \{x \in E \text{ tel que } f(x) \in \cup_n B_n\} = f^{-1}(\cup_n B_n), \end{aligned}$$

qui appartient à \mathcal{E} puisque $\cup_n B_n$ appartient à \mathcal{F} .

Ainsi \mathcal{E} est bien une tribu. \square

Théorème 1.1.4 Soit (Ω, \mathcal{A}) un espace probabilisable et Ω' une partie de Ω . L'ensemble

$$\{A \cap \Omega' : A \in \mathcal{A}\}$$

est une tribu sur Ω' et est appelée trace de la tribu \mathcal{A} sur Ω' .

Preuve. Notons

$$\mathcal{C} = \{C = A \cap \Omega' : A \in \mathcal{A}\}.$$

* On a $\Omega' = \Omega \cap \Omega'$ et donc $\Omega' \in \mathcal{C}$

* Soit C un élément de \mathcal{C} et notons C^* son complémentaire par rapport à Ω' . On a :

$$C^* = \bar{A} \cap \Omega' \in \mathcal{C}$$

* Supposons maintenant que, pour tout n , C_n soit dans \mathcal{C} . Il existe donc pour tout n ,

$$A_n \in \mathcal{A}, \quad \text{tel que } C_n = A_n \cap \Omega'.$$

D'où:

$$\bigcup_n C_n = \bigcup_n (A_n \cap \Omega') = \left(\bigcup_n A_n \right) \cap \Omega' \in \mathcal{C}.$$

Ainsi, \mathcal{C} est bien une tribu sur Ω' . \square

Théorème 1.1.5 *Soit I une partie de \mathbb{N} et $(\mathcal{A}_i)_{i \in I}$ une famille de tribus sur le même espace fondamental Ω . La famille de parties*

$$\mathcal{A} = \bigcap_{i \in I} \mathcal{A}_i$$

est une tribu.

Preuve.

* L'ensemble Ω est dans \mathcal{A}_i , pour tout i , il est donc un élément de \mathcal{A} .

* De plus, on a :

$$(A \in \mathcal{A}) \Rightarrow (\forall i : A \in \mathcal{A}_i) \Rightarrow (\forall i : \bar{A} \in \mathcal{A}_i) \Rightarrow (\bar{A} \in \mathcal{A}).$$

* Enfin, supposons que, pour tout n , on ait A_n dans \mathcal{A} . On a alors

$$(\forall n, \forall i, A_n \in \mathcal{A}_i) \Rightarrow (\forall i, \bigcup_n A_n \in \mathcal{A}_i) \Rightarrow (\bigcup_n A_n \in \mathcal{A}),$$

ce qui achève la démonstration. \square

Nous attirons l'attention du lecteur sur le point suivant. Si (\mathcal{A}_n) est une famille quelconque de parties d'un ensemble Ω et si un élément A est tel que

$$A \in \mathcal{A}_n,$$

pour tout n , alors on a :

$$A \in \bigcap_n \mathcal{A}_n.$$

En revanche, si (A_n) est une famille de parties d'un ensemble \mathcal{A} (i.e. $A_n \in \mathcal{A}$, pour tout n), on n'a pas nécessairement :

$$\bigcap_n A_n \in \mathcal{A},$$

sauf si \mathcal{A} est une tribu.

Théorème 1.1.6 *Soit \mathcal{F} une famille de parties de Ω . Il existe une plus petite tribu sur Ω qui contient \mathcal{F} . On l'appelle tribu engendrée par \mathcal{F} et on la note $\sigma(\mathcal{F})$.*

Preuve. Comme $\mathcal{P}(\Omega)$ est une tribu contenant \mathcal{F} , l'ensemble des tribus contenant \mathcal{F} n'est pas vide. L'intersection de ces tribus est d'après le théorème précédent encore une tribu. Elle contient \mathcal{F} et c'est forcément la plus petite tribu contenant \mathcal{F} . \square

Voyons un exemple particulier de tribu.

Définition 1.1.7 On appelle tribu borélienne sur \mathbb{R} , la tribu engendrée par les intervalles ouverts de la forme $] - \infty, x[$, pour tout x dans \mathbb{R} . On la note $\mathcal{B}_{\mathbb{R}}$.

On peut montrer que l'on a le résultat suivant.

Théorème 1.1.8 La tribu borélienne est également engendrée par les intervalles de la forme $] - \infty, x]$, $[x, +\infty[$, $]x, +\infty[$, $[x, y]$, $]x, y[$ etc...

1.1.3 Mesures et probabilités

Définition 1.1.9 On appelle mesure positive sur l'espace probabilisable (Ω, \mathcal{A}) toute application μ de \mathcal{A} dans

$$\overline{\mathbb{R}}^+ = [0, +\infty]$$

telle que d'une part l'on ait $\mu(\emptyset) = 0$ et que d'autre part pour toute suite (A_n) d'éléments de \mathcal{A} , deux à deux disjoints, on ait :

$$\mu\left(\bigcup_n A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

Le triplet $(\Omega, \mathcal{A}, \mu)$ est appelé espace mesuré.

Définition 1.1.10 Une mesure P sur (Ω, \mathcal{A}) telle que $P(\Omega) = 1$ est dite une probabilité. Le triplet (Ω, \mathcal{A}, P) est appelé espace probabilisé.

Proposition 1.1.11 Une probabilité vérifie les assertions suivantes :

- (i) $\forall A \in \mathcal{A}, P(\bar{A}) = 1 - P(A)$;
- (ii) $P(\emptyset) = 0$;
- (iii) $\forall (A, B) \in \mathcal{A}^2, P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
- (iv) Formule de Poincaré : soit A_1, \dots, A_n des événements de \mathcal{A} . On a :

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \\ &\quad + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\ &\quad + \dots + (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &\quad + \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n). \end{aligned}$$

(v) Si A et B , deux éléments de \mathcal{A} , sont tels que $A \subset B$, on a alors

$$P(A) \leq P(B).$$

(vi) Inégalité de Boole : si A_1, \dots, A_n sont des événements de \mathcal{A} , on a :

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

Preuve.

(i) $P(\Omega) = 1 = P(A) + P(\bar{A})$ puisque A et \bar{A} sont disjoints.

(ii) $P(\emptyset) = 1 - P(\Omega) = 0$.

(iii) On a

$$P(A \cup B) = P(A \cap \bar{B}) + P(B).$$

Or

$$P(A) = P(A \cap \{B \cup \bar{B}\}) = P(A \cap B) + P(A \cap \bar{B}).$$

Ainsi

$$P(A \cap \bar{B}) = P(A) - P(A \cap B).$$

D'où

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

(iv) Exercice.

(v) On a vu que

$$P(B \cap \bar{A}) = P(B) - P(A \cap B).$$

Or

$$A \subset B \Rightarrow A \cap B = A.$$

D'où

$$P(B) - P(A) = P(B \cap \bar{A}) \geq 0.$$

(vi) D'après le (iii) la formule est vraie pour $n = 2$. Supposons la vraie au rang $n - 1$. On a alors

$$P\left(\bigcup_{i=1}^n A_i\right) \leq P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) \leq \sum_{i=1}^{n-1} P(A_i) + P(A_n) = \sum_{i=1}^n P(A_i)$$

La formule est donc vraie pour tout n , par récurrence. \square

Remarque. Les propriétés (iii), (v) et (vi) restent vraies pour les mesures. La propriété (ii) reste vraie pour une mesure si celle-ci n'est pas dégénérée (i.e. de valeur tout le temps $+\infty$). \diamond

Définition 1.1.12 On dit qu'une suite (A_n) d'événements est croissante (resp. décroissante) si

$$\forall n : A_n \subset A_{n+1} \text{ (resp. } A_n \supset A_{n+1}\text{)}.$$

On admettra la proposition suivante :

Proposition 1.1.13 Si P est une probabilité sur (Ω, \mathcal{A}) alors

(i) Pour toute suite (A_n) d'événements croissante, on a :

$$P\left(\bigcup_n A_n\right) = \lim_{n \rightarrow +\infty} P(A_n) = \sup_n P(A_n).$$

(ii) Pour toute suite (A_n) d'événements décroissante, on a

$$P\left(\bigcap_n A_n\right) = \lim_{n \rightarrow +\infty} P(A_n) = \inf_n P(A_n).$$

Définition 1.1.14 Soit (Ω, \mathcal{A}) un espace mesurable tel que $\{\omega\} \in \mathcal{A}$ pour tout $\omega \in \Omega$.

- On dit qu'une mesure μ sur (Ω, \mathcal{A}) est discrète s'il existe une famille

$$D = \{\omega_n \in \Omega : n \in I\}$$

(où I est un ensemble d'indices fini ou dénombrable) d'éléments telle que

$$\mu(\Omega \setminus D) = 0$$

et

$$\forall A \in \mathcal{A} : \mu(A) = \mu(A \cap D) = \sum_{\omega_n \in A \cap D} \mu(\{\omega_n\}).$$

La mesure μ est dite concentrée sur D .

- On dit qu'une mesure μ sur (Ω, \mathcal{A}) est continue si elle ne possède pas d'atome, i.e. si, pour tout ω dans Ω , on a $\mu(\{\omega\}) = 0$.

Exemples fondamentaux.

1) Mesure de Dirac

On appelle mesure de Dirac au point ω_0 de Ω la probabilité discrète δ_{ω_0} définie, pour tout A dans \mathcal{A} par :

$$\begin{cases} \delta_{\omega_0}(A) = 1 & \text{si } \omega_0 \in A \\ \delta_{\omega_0}(A) = 0 & \text{sinon.} \end{cases}$$

On a donc

$$\delta_{\omega_0}(A) = \mathbb{I}_A(\omega_0).$$

2) Construction d'une probabilité discrète plus générale

Soit $D = (\omega_n)_{n \in I}$ une suite finie ou dénombrable d'éléments de Ω et $(p_n)_{n \in I}$ une famille de réels tels que

$$\forall n \in I, p_n \geq 0 \text{ et } \sum_{n \in I} p_n = 1.$$

Alors, l'application P définie par :

$$P = \sum_{n \in I} p_n \delta_{\omega_n}$$

est une probabilité discrète sur $(\Omega, \mathcal{P}(\Omega))$ ou (Ω, \mathcal{A}) pour toute tribu \mathcal{A} contenant tous les singletons.

On a

$$\forall n \in I : P(\{\omega_n\}) = p_n$$

et

$$\forall \omega \in \Omega \setminus (\omega_n)_{n \in I} : P(\{\omega\}) = 0.$$

On peut ainsi définir, par exemple, l'équiprobabilité sur $\{1, 2, \dots, N\}$. Reprenons en effet le jeu du lancé de dé. On a :

$$\Omega = \{1, \dots, 6\} \text{ et } \mathcal{A} = \mathcal{P}(\Omega).$$

Si on prend comme probabilité sur Ω , l'équiprobabilité, on aura nécessairement pour tout $n = 1, \dots, 6$:

$$p_n = P(\{\omega_n\}) = P(\text{le résultat est } n) = \frac{1}{\text{Card } \Omega} = \frac{1}{6}$$

et donc

$$P = \frac{1}{6} \sum_{n=1}^6 \delta_{\omega_n}.$$

Soit alors A une partie de Ω , on a alors :

$$P(A) = \sum_{n=1}^6 \delta_{\omega_n}(A) = \frac{\text{Card } A}{\text{Card } \Omega}$$

Bien sûr, si on a plus l'équiprobabilité, la formule classique précédente n'est plus vraie. Ainsi, si on prend un dé pipé ayant les probabilités suivantes des résultats élémentaires :

$$\begin{aligned} p_1 &= p_3 = p_5 = \frac{1}{12} \\ p_2 &= p_4 = p_6 = \frac{1}{4}. \end{aligned}$$

La probabilité de l'événement A = "le résultat est un nombre pair" est alors :

$$P(A) = \sum_{n=1}^6 \delta_{\omega_n}(A) = p_2 + p_4 + p_6 = \frac{3}{4} \neq \frac{\text{Card } A}{\text{Card } \Omega}.$$

3) Mesure de comptage

On définit sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ ou $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ la mesure :

$$\mu = \sum_{n=0}^{+\infty} \delta_n.$$

On vérifie aisément qu'il s'agit bien d'une mesure. Elle est discrète sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ et $(\mathbb{R}, B_{\mathbb{R}})$ puisque $D = \mathbb{N}$ est dénombrable et $\mu(\mathbb{R} \setminus \mathbb{N}) = 0$ dans le deuxième cas. Cette mesure est appelée mesure de comptage. Si on raisonne sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ et si $A \in \mathcal{P}(\mathbb{N})$,

$$\mu(A) = \sum_{n=0}^{+\infty} \delta_n(A) = \text{nombre d'éléments de } A.$$

Si on raisonne sur $(\mathbb{R}, B_{\mathbb{R}})$, la mesure $\mu(A)$ de l'événement A est le nombre d'entiers dans A .

4) Mesure de Lebesgue sur $(\mathbb{R}, B_{\mathbb{R}})$

On appelle mesure de Lebesgue sur $(\mathbb{R}, B_{\mathbb{R}})$ la mesure λ définie par :

$$\lambda([a, b]) = b - a,$$

où $a < b$ sont des réels. On vérifie l'existence et l'unicité d'une telle mesure et on a

$$\lambda([a, b]) = \lambda([a, b[) = \lambda([a, b]) = \lambda(]a, b]).$$

La mesure de Lebesgue est une mesure continue sur \mathbb{R} puisque, pour tout x dans \mathbb{R} , on a :

$$\lambda(\{x\}) = \lim_{n \rightarrow +\infty} \lambda\left(x - \frac{1}{n}, x + \frac{1}{n}\right) = \lim_{n \rightarrow +\infty} \left(x + \frac{1}{n} - x + \frac{1}{n}\right) = 0. \quad \diamond$$

1.1.4 Variables aléatoires

Définition 1.1.15 Soient (Ω, \mathcal{A}) et (E, \mathcal{B}) deux espaces probabilisables. Une application f de Ω vers E est dite mesurable (ou \mathcal{A} -mesurable si la confusion est possible) si

$$\forall B \in \mathcal{B}, \quad f^{-1}(B) \in \mathcal{A}.$$

On a déjà vu que $f^{-1}(\mathcal{B})$ est une tribu. Dire que la fonction f est mesurable revient donc à dire que $f^{-1}(\mathcal{B}) \subset \mathcal{A}$. Ainsi, pour tout événement B , l'ensemble :

$$f^{-1}(B) = \{\omega \in \Omega / f(\omega) \in B\}$$

est un événement de la tribu initiale. On utilise parfois la notation $f^{-1}(B) = [f \in B]$.

Notons en particulier que toute fonction continue est mesurable. De même, pour tout événement A de la tribu \mathcal{A} , la fonction $\mathbb{1}_A$ est mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Proposition 1.1.16

- Si f et g sont deux fonctions mesurables de (Ω, \mathcal{A}) vers $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ alors les fonctions $f + g$ et fg sont encore mesurables.
- Si f et g sont deux fonctions mesurables de (Ω, \mathcal{A}) vers (Ω', \mathcal{A}') et de (Ω', \mathcal{A}') vers $(\Omega'', \mathcal{A}'')$ respectivement, la fonction $g \circ f$ est mesurable de (Ω, \mathcal{A}) vers $(\Omega'', \mathcal{A}'')$.
- Si (f_n) est une suite de fonctions mesurables de (Ω, \mathcal{A}) vers $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, alors les fonctions

$$\sup_n f_n, \inf_n f_n, \limsup_n f_n \text{ et } \liminf_n f_n$$

sont mesurables, à condition qu'elles ne prennent pas de valeurs infinies.

Définition 1.1.17 Soit (Ω, \mathcal{A}, P) un espace probabilisé et (E, \mathcal{B}) un espace probabilisable. Une application mesurable X de (Ω, \mathcal{A}, P) vers (E, \mathcal{B}) est appelée *variable aléatoire*.

Tous les résultats sur les fonctions mesurables restent donc vrais pour les variables aléatoires. Ainsi, on pourra parler du *supremum* sur une famille infinie de variables aléatoires et de limite de variables aléatoires. On sera assuré qu'il s'agit encore de variables aléatoires.

Notations.

Si $(E, \mathcal{B}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, l'application X est dite **variable aléatoire réelle** (v.a.r.) ou unidimensionnelle ou univariée.

Si $(E, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$, l'application X est dite **vecteur aléatoire** ou variable aléatoire multidimensionnelle ou multivariée.

Si (E, \mathcal{B}) est tout, ou une partie, de $(\mathbb{Z}, \mathcal{B}_{\mathbb{Z}})$, l'application X est dite **v.a. discrète**.

1.1.5 Loi de probabilité d'une variable aléatoire

Soit X une variable aléatoire de (Ω, \mathcal{A}, P) vers (E, \mathcal{B}) . Définissons une application P_X de \mathcal{B} vers $[0, 1]$ par :

$$\forall B \in \mathcal{B}, \quad P_X(B) = P[X^{-1}(B)] = P[\{\omega : X(\omega) \in B\}].$$

La définition précédente a bien un sens puisque l'on a $X^{-1}(B) \in \mathcal{A}$, par mesurabilité de X . On peut donc prendre la probabilité de cet événement.

Définition 1.1.18 P_X est appelée *probabilité image de P par X ou encore loi de probabilité de la variable aléatoire X* . On note $P_X(B) = P(X \in B)$.

Ainsi, tout événement lié à X est connu dès que l'on connaît la loi P_X de X . On oubliera donc souvent dans la suite le détail de l'application $\omega \mapsto X(\omega)$ et on ne se préoccupera pas de ce qu'est exactement (Ω, \mathcal{A}, P) . On raisonnera uniquement sur (E, \mathcal{B}) et P_X .

Notons par ailleurs que tous les résultats obtenus pour X et P_X seront alors aussi valables pour toute variable aléatoire Y de même loi que X .

1.2 Conditionnement

Supposons que l'on joue au lancer de dé avec un dé dont les faces paires sont de couleur blanche et les faces impaires de couleur noire. Si de loin on

peut seulement distinguer la couleur blanche de la face obtenue, on modifiera naturellement les probabilités des événements. Ainsi on donnera la probabilité $1/3$ pour chaque face paire et la probabilité 0 pour chaque face impaire, plutôt que l'équirépartition initiale de probabilité $1/6$ pour chaque résultat élémentaire. On constate donc que la connaissance de la parité du résultat modifie les probabilités que l'on donne à chaque événement. On dit que l'on raisonne conditionnellement à l'événement "le résultat est pair".

1.2.1 Probabilité conditionnelle à un événement

Soit (Ω, \mathcal{A}, P) et B un événement de \mathcal{A} de probabilité non nulle. Si on sait que l'événement B s'est réalisé, donc que $\omega \in B$, pour tout événement A de \mathcal{A} on a :

$$\omega \in A \Leftrightarrow \omega \in A \cap B.$$

Cela nous conduit à considérer l'application :

$$\mu : \begin{cases} \mathcal{A} & \rightarrow \mathbb{R}^+ \\ A & \mapsto P(A \cap B). \end{cases}$$

L'application μ est une mesure sur \mathcal{A} mais n'est en général pas une probabilité car

$$\mu(\Omega) = P(\Omega \cap B) = P(B)$$

et n'est donc pas forcément égal à 1. On considère alors l'application

$$P^B = \frac{\mu}{P(B)}$$

qui, elle, est bien une probabilité sur (Ω, \mathcal{A}) .

Définition 1.2.1 *Pour tout événement B de probabilité non nulle, on appelle probabilité conditionnelle à B , la probabilité sur (Ω, \mathcal{A})*

$$\begin{aligned} P^B : \mathcal{A} & \rightarrow [0, 1] \\ A & \mapsto P^B(A) = \frac{P(A \cap B)}{P(B)}. \end{aligned}$$

$P^B(A)$ s'appelle probabilité conditionnelle à B de A (ou encore probabilité de A sachant B). On note aussi

$$P^B(A) = P(A/B).$$

Remarquons que l'on peut aussi voir cette probabilité comme une probabilité sur la tribu trace de \mathcal{A} sur B .

Proposition 1.2.2 *Par définition on a donc*

$$\begin{aligned}\forall A \in \mathcal{A}, \quad P(A \cap B) &= P(B)P(A/B) \\ &= P(A)P(B/A).\end{aligned}$$

Vérifions que P^B ainsi définie est bien une probabilité sur (Ω, \mathcal{A}) . Puisque, pour tout A dans \mathcal{A} , l'événement $A \cap B$ est inclus dans B , on a :

$$0 \leq P(A/B) \leq 1.$$

Trivialement on a également :

$$P(\Omega/B) = \frac{P(\Omega \cap B)}{P(B)} = 1.$$

Enfin, soit (A_n) une suite d'événements de \mathcal{A} deux à deux disjoints. On a :

$$\begin{aligned}P\left(\bigcup_n A_n / B\right) &= \frac{P((\bigcup_n A_n) \cap B)}{P(B)} = \frac{P(\bigcup_n (A_n \cap B))}{P(B)} = \frac{\sum_n P(A_n \cap B)}{P(B)} \\ &= \sum_n P(A_n/B)\end{aligned}$$

1.2.2 Formule de Bayes

Rappelons qu'une famille finie ou dénombrable d'ensembles (non vides) $(A_i)_{i \in I}$ est dite partition de Ω si on a : $A_i \cap A_j = \emptyset$ pour $i \neq j$ et $\Omega = \bigcup_{i \in I} A_i$. Supposons que la famille $(A_i)_{i \in I}$ soit une telle partition Ω . On suppose que, pour tout i dans I , on ait $P(A_i) \neq 0$. Pour tout événement B de probabilité non nulle on peut alors écrire :

$$\begin{aligned}P(B) &= P(\Omega \cap B) \\ &= P\left(\left(\bigcup_{i \in I} A_i\right) \cap B\right) = \sum_{i \in I} P(A_i \cap B) = \sum_{i \in I} P(A_i)P(B/A_i).\end{aligned}$$

Ainsi, pour tout j dans I , on a :

$$P(A_j/B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(A_j)P(B/A_j)}{\sum_{i \in I} P(A_i)P(B/A_i)}$$

Cette formule est appelée formule de Bayes. Voyons-en un exemple d'application.

Exemple du dépistage de la Syphilis.

On applique un test médical sur les patients pour déceler la maladie. On sait que si le patient est effectivement atteint, le test est positif dans 99% des cas. Mais on sait aussi qu'il y a 2% des cas où le résultat du test est positif alors que le consultant est en bonne santé. Sachant qu'un patient sur 1 000 est atteint de la Syphilis, calculer la probabilité qu'un patient soit atteint sachant que son test a été positif.

Soit M l'événement "le client est atteint" et T^+ l'événement "le test est positif".

Les données de l'énoncé peuvent être écrites de la manière suivante :

$$\begin{aligned} P(T^+/M) &= 0,99 \\ P(T^+/\bar{M}) &= 0,02 \\ \text{et } P(M) &= 1/1000. \end{aligned}$$

La probabilité $P(M/T^+)$, celle qu'un patient soit atteint sachant que son test est positif, est égale, d'après la formule de Bayes, à :

$$\begin{aligned} P(M/T^+) &= \frac{P(T^+/M) \times P(M)}{P(T^+/M)P(M) + P(T^+/\bar{M})P(\bar{M})} \\ &= \frac{0,99 \times 0,001}{0,99 \times 0,001 + 0,02 \times 0,999} = \frac{1}{21} \end{aligned}$$

Le test proposé ne semble pas être très "efficace"...

◇

1.3 Indépendance en probabilité

1.3.1 Indépendance d'événements

Définition 1.3.1 Deux événements A et B sur un même espace probabilisé (Ω, \mathcal{A}, P) sont dits indépendants en probabilité (noté $A \perp\!\!\!\perp B$) si

$$P(A \cap B) = P(A)P(B).$$

La notion d'indépendance est donc liée aux événements considérés et à la probabilité de référence.

Notons que si A et B sont deux événements de probabilité strictement positive, l'indépendance en probabilité est donc équivalente à chacune des deux assertions suivantes:

$$\begin{aligned} (i) \quad &P(B/A) = P(B) \\ (ii) \quad &P(A/B) = P(A). \end{aligned}$$

Il est important de ne pas confondre les notions d'événements indépendants, incompatibles et disjoints. Notons d'abord que les deux premières notions sont des notions ensemblistes et probabilistes et que la dernière est purement ensembliste. Rappelons que les événements A et B sont dits disjoints si $A \cap B = \emptyset$ et qu'ils sont dits incompatibles si $P(A \cap B) = 0$. Des événements disjoints sont donc forcément incompatibles, mais la réciproque est fausse.

Par ailleurs, des événements incompatibles sont rarement indépendants. Prenons l'exemple de deux événements A et B de \mathcal{A} , disjoints et tous les deux de probabilité strictement positive. Comme on vient de le dire, ils sont forcément incompatibles mais en revanche ils ne sont pas indépendants car si A est réalisé, B ne l'est pas.

Remarquons enfin que deux événements incompatibles ne peuvent en fait être indépendants que si l'un ou l'autre des deux événements est de probabilité nulle, i.e.

$$P(A \cap B) = 0 = P(A)P(B) \Leftrightarrow P(A) = 0 \text{ ou } P(B) = 0.$$

Proposition 1.3.2

- 1) Si A et B sont indépendants, alors $A \perp\!\!\!\perp \bar{B}$, $\bar{A} \perp\!\!\!\perp B$, $\bar{A} \perp\!\!\!\perp \bar{B}$.
- 2) Si l'événement A est tel que sa probabilité est soit nulle soit égale à 1, alors

$$\forall B \in \mathcal{A}, \quad A \perp\!\!\!\perp B.$$

Preuve.

- 1) On a

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A)P(B) + P(A \cap \bar{B}).$$

D'où

$$\begin{aligned} P(A \cap \bar{B}) &= P(A) - P(A)P(B) = P(A)(1 - P(B)) \\ &= P(A)P(\bar{B}), \end{aligned}$$

ce qui prouve l'indépendance entre A et \bar{B} . Les autres indépendances s'obtiennent ensuite facilement par symétrie.

- 2) Si l'événement A est tel que $P(A) = 0$, il vient

$$P(A \cap B) \leq P(A) = 0 = P(A)P(B)$$

et donc $A \perp\!\!\!\perp B$.

Si à l'opposé $P(A) = 1$, l'égalité $P(\bar{A}) = 0$ et ce qui précède entraîne alors

$$B \perp\!\!\!\perp \bar{A}, \quad \forall B \in \mathcal{A}$$

et donc

$$B \perp\!\!\!\perp A, \quad \forall B \in \mathcal{A}.$$

□

Définition 1.3.3 Soit $(A_i)_{i=1,\dots,n}$ une famille d'événements de \mathcal{A} . Ces événements sont dits (mutuellement) indépendants en probabilité si :

$$\forall I \subset \{1, \dots, n\} \quad P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Nous attirons l'attention du lecteur sur le point suivant : l'indépendance mutuelle entraîne clairement l'indépendance deux à deux mais la réciproque est fausse. En voici un contre-exemple.

Exemple. On lance deux dés équilibrés et de manière indépendante. Soit A l'événement “le premier dé amène un nombre pair”, B l'événement “le second dé amène un nombre impair” et C l'événement “les deux dés amènent des nombres de même parité”.

On calcule facilement les probabilités suivantes :

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

et

$$P(A \cap B) = P(B \cap C) = P(A \cap C) = \frac{1}{4}$$

Ainsi, les événements A , B et C sont indépendants deux à deux. En revanche ils ne sont pas mutuellement indépendants puisque l'on a :

$$P(A \cap B \cap C) = P(\emptyset) = 0 \neq P(A)P(B)P(C).$$

Cet exemple nous permet également de voir que l'événement A peut être indépendant de B et de C séparément, sans l'être de l'intersection $B \cap C$. ◇

Remarquons enfin que l'on peut généraliser cette notion d'indépendance mutuelle pour une famille non nécessairement finie d'éléments.

Définition 1.3.4 Une famille $(A_i)_{i \in I}$ d'événements est une famille d'événements mutuellement indépendants si, pour tout ensemble d'indices K fini et dans I , la famille $(A_i)_{i \in K}$ forme une famille d'événements mutuellement indépendants.

1.3.2 Indépendance de tribus

Soit $(\mathcal{A}_i)_{i=1,\dots,n}$ une famille de sous tribus de \mathcal{A} ,

Définition 1.3.5 *On dit que la famille $(\mathcal{A}_i)_{i=1,\dots,n}$ est une famille indépendante de sous tribus si pour toute famille d'événements $(A_i)_{i=1,\dots,n}$ où $A_i \in \mathcal{A}_i$, pour tout i , on a :*

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i).$$

En fait on peut à nouveau définir la notion d'indépendance pour toute famille (de cardinal quelconque) non vide de parties de \mathcal{A} . Une famille $(\mathcal{C}_i)_{i \in I}$ de parties de \mathcal{A} (i.e. $\mathcal{C}_i \subset \mathcal{A}$ pour tout i mais où \mathcal{C}_i n'est pas forcément une tribu) est dite indépendante si, quand I est fini, pour toute famille $(A_i)_{i \in I}$ où A_i est dans \mathcal{C}_i pour tout i , on a

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

ou si, quand I est infini, toute sous-famille finie est indépendante.

1.3.3 Indépendance de variables aléatoires

Soit une famille $(X_i)_{i=1,\dots,n}$ de variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{A}, P) et à valeurs respectivement dans l'espace probabilisable (E_i, \mathcal{B}_i) .

Définition 1.3.6 *Une famille $(X_i)_{i=1,\dots,n}$ de variables aléatoires est dite indépendante en probabilité si :*

$$\forall (B_i)_{i=1,\dots,n} \text{ où } B_i \in \mathcal{B}_i, \text{ pour tout } i,$$

on a :

$$P\left(\bigcap_{i=1}^n \{X_i \in B_i\}\right) = \prod_{i=1}^n P(\{X_i \in B_i\}).$$

Théorème 1.3.7 *Si, pour tout i , les fonctions φ_i sont des fonctions mesurables de (E_i, \mathcal{B}_i) vers (E'_i, \mathcal{B}'_i) , alors l'indépendance des variables aléatoires $(X_i)_{i=1,\dots,n}$ entraîne celle des $(\varphi_i(X_i))_{i=1,\dots,n}$.*

Preuve. Pour toute famille $(B'_i)_{i=1,\dots,n}$ où $B'_i \in \mathcal{B}'_i$, pour tout i , on a

$$\varphi_i^{-1}(B'_i) = B_i \in \mathcal{B}_i$$

par mesurabilité des φ_i . Il vient alors :

$$(\varphi_i(X_i))^{-1}(B'_i) = X_i^{-1}(\varphi_i^{-1}(B'_i)) = X_i^{-1}(B_i).$$

D'où :

$$\begin{aligned} P\left(\bigcap_{i=1}^n \{\varphi_i \circ X_i \in B'_i\}\right) &= P\left(\bigcap_{i=1}^n (\varphi_i(X_i))^{-1}(B'_i)\right) = P\left(\bigcap_{i=1}^n \{X_i^{-1}(B_i)\}\right) \\ &= P\left(\bigcap_{i=1}^n \{X_i \in B_i\}\right) = \prod_{i=1}^n P(X_i \in B_i) \\ &= \prod_{i=1}^n P(\varphi_i(X_i) \in B'_i) \end{aligned}$$

et les $(\varphi_i(X_i))_{i=1,\dots,n}$ sont bien des variables aléatoires indépendantes. \square

Exemples.

* Si X et Y sont des v.a.r. indépendantes, X^2 et $\log Y$ le sont encore.

* Si X, Y, Z, T et V sont des variables aléatoires indépendantes et si f est mesurable de \mathbb{R}^3 vers \mathbb{R} , alors X et $U = f(Y, Z, T)$ sont indépendantes. De même $X, g(Y, Z)$ et $h(T, V)$ sont indépendantes pour des fonctions g et h mesurables. \diamond

1.3.4 Lien entre les différents types d'indépendance

Soit (Ω, \mathcal{A}, P) un espace probabilisé et $(\mathcal{A}_i)_{i=1,\dots,n}$ une famille de sous tribus de \mathcal{A} .

Proposition 1.3.8 *On a l'équivalence entre les assertions suivantes :*

- i) $\perp\!\!\!\perp_{i=1}^n \mathcal{A}_i$
- ii) $\perp\!\!\!\perp_{i=1}^n X_i$, pour toute famille $(X_i)_{i=1,\dots,n}$ de v.a. où X_i est \mathcal{A}_i mesurable
- iii) $\perp\!\!\!\perp_{i=1}^n \mathbb{I}_{A_i}$, pour toute famille $(A_i)_{i=1,\dots,n}$ où $A_i \in \mathcal{A}_i$ pour tout i
- iv) $\perp\!\!\!\perp_{i=1}^n A_i$, pour toute famille $(A_i)_{i=1,\dots,n}$ où $A_i \in \mathcal{A}_i$ pour tout i

Proposition 1.3.9 Soit $(A_i)_{i=1,\dots,n}$ une famille d'événements sur un même espace probabilisé (Ω, \mathcal{A}, P) . On a les équivalences suivantes :

$$\coprod_{i=1}^n A_i \Leftrightarrow \coprod_{i=1}^n \sigma(A_i) \Leftrightarrow \coprod_{i=1}^n \mathbb{I}_{A_i}.$$

Proposition 1.3.10 Soit $(X_i)_{i=1,\dots,n}$ une famille de variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{A}, P) et à valeurs respectivement dans $(E_i, \mathcal{B}_i)_{i=1,\dots,n}$. On a les équivalences suivantes :

$$\coprod_{i=1}^n X_i \Leftrightarrow \coprod_{i=1}^n \sigma(X_i) \Leftrightarrow \coprod_{i=1}^n \{X_i \in B_i\}, \text{ pour tout } B_i \in \mathcal{B}_i.$$

Remarque. La famille de parties $\sigma(X_i)$ est la tribu engendrée par X_i . C'est la plus petite tribu rendant X_i mesurable. On a :

$$\sigma(X_i) = X_i^{-1}(\mathcal{B}).$$

◇

1.4 Espace probabilisé produit

Jusqu'à présent on a essentiellement parlé de l'observation d'un phénomène unique. On peut pourtant s'intéresser à un phénomène φ qui est la juxtaposition de n phénomènes φ_i où chaque phénomène φ_i est modélisé par $(\Omega_i, \mathcal{A}_i, P_i)$. Pour modéliser $\varphi = (\varphi_1, \dots, \varphi_n)$ il nous faut déterminer l'espace probabilisé (Ω, \mathcal{A}, P) associé.

De façon naturelle, si ω_i est l'observation du phénomène φ_i , le n -uplet $\omega = (\omega_1, \dots, \omega_n)$ est une observation du phénomène φ . On prendra donc comme espace fondamental :

$$\Omega = \Omega_1 \times \dots \times \Omega_n = \prod_{i=1}^n \Omega_i.$$

Intéressons nous maintenant à la construction de la tribu \mathcal{A} sur Ω . Si pour tout i , l'ensemble A_i est un événement de \mathcal{A}_i (i.e. $A_i \in \mathcal{A}_i$), il est naturel d'attendre que $A_1 \times \dots \times A_n$ soit un événement de \mathcal{A} . C'est pourquoi on pose la définition suivante.

Définition 1.4.1 On appelle tribu produit des $(\mathcal{A}_i)_{i=1,\dots,n}$, et on note

$$\otimes_{i=1}^n \mathcal{A}_i,$$

la tribu engendrée par les pavés mesurables $A_1 \times \dots \times A_n$ où A_i appartient à \mathcal{A}_i , pour tout i , i.e. :

$$\otimes_{i=1}^n \mathcal{A}_i = \sigma \{A_1 \times \dots \times A_n : A_i \in \mathcal{A}_i \text{ pour tout } i = 1, \dots, n\}.$$

On choisit alors \mathcal{A} comme tribu pour modéliser le phénomène aléatoire φ .

Il nous faut maintenant définir une probabilité P sur cet espace produit (Ω, \mathcal{A}) , à partir de l'ensemble des P_i . Sans information complémentaire, on ne peut pas...

En revanche, cela devient possible si on suppose, par exemple, que les phénomènes φ_i sont indépendants (au sens intuitif et non mathématique). Cela revient à dire que les événements $(A_i)_{i=1,\dots,n}$, où chaque A_i est dans \mathcal{A}_i , sont indépendants (toujours au sens intuitif puisque cela n'a pas de sens mathématique, les événements A_i étant associés à des probabilités P_i différentes).

Notons, à ce propos, que dans \mathcal{A} l'événement A_i s'écrit

$$B_i = \Omega \times \cdots \times \Omega \times A_i \times \Omega \times \cdots \times \Omega.$$

Une fois que l'on aura déterminé P , l'indépendance intuitive de $(A_i)_{i=1,\dots,n}$ se traduira par l'indépendance en probabilité des B_i , i.e. :

$$P\left(\bigcap_{i=1}^n B_i\right) = \prod_{i=1}^n P(B_i).$$

Notons que

$$\bigcap_{i=1}^n B_i = A_1 \times \cdots \times A_n$$

et que naturellement, pour des questions de cohérence, on doit avoir

$$P(B_i) = P_i(A_i).$$

En résumé, on souhaitera donc que la probabilité P mise sur l'espace probabilisable produit (Ω, \mathcal{A}) vérifie

$$P(A_1 \times \cdots \times A_n) = \prod_{i=1}^n P_i(A_i).$$

Le théorème suivant montre (bien que sa preuve ne soit pas donnée ici !) qu'une telle probabilité existe et que, de plus, elle est unique.

Théorème 1.4.2 *Il existe une probabilité unique P sur*

$$(\prod_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i)$$

telle que pour tout A_i dans \mathcal{A}_i , pour $i = 1, \dots, n$, on ait :

$$P(A_1 \times \dots \times A_n) = \prod_{i=1}^n P_i(A_i).$$

Cette probabilité P est appelée probabilité produit des P_i et est notée

$$P = \otimes_{i=1}^n P_i.$$

Définition 1.4.3 *L'espace*

$$(\Pi_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i, \otimes_{i=1}^n P_i)$$

est appelé espace probabilisé produit des espaces $(\Omega_i, \mathcal{A}_i, P_i)$.

Remarque. Si, pour tout i , on a $\Omega_i = \Omega$, $\mathcal{A}_i = \mathcal{A}$ et $P_i = P_0$, l'espace produit est noté

$$(\Omega, \mathcal{A}, P_0)^{\otimes n}.$$

◇

Exemple. Considérons le jeu du lancé de deux dés. Pour chaque dé, on a déjà défini l'espace probabilisé associé au phénomène.

$$\begin{aligned} \Omega &= \{1, 2, \dots, 6\} \\ \mathcal{A} &= \mathcal{P}(\Omega) \\ P &= \text{équirépartition sur } \Omega. \end{aligned}$$

On note $\varphi = (\omega_1, \omega_2)$ le résultat du lancé des deux dés. D'après ce qui précède, on modélise ce phénomène par $(\Omega, \mathcal{A}, P)^{\otimes 2}$. ◇

On pourra alors, grâce à cette structure, parler de la somme des deux résultats, du maximum, etc...

1.5 Loi conjointe d'un n -uplet de variables aléatoires indépendantes

Soit X_1, \dots, X_n des variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{A}, P) et à valeurs vers respectivement $(\Omega_i, \mathcal{A}_i)_{i=1, \dots, n}$. On admet que le vecteur (X_1, \dots, X_n) est encore une variable aléatoire de (Ω, \mathcal{A}, P) vers

$$(\Pi_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i).$$

On peut en effet montrer qu'une fonction h à valeurs dans l'espace mesurable $(\Pi_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i)$ est mesurable si, et seulement si, $\pi_i \circ h$ est $(\Omega_i, \mathcal{A}_i)$ -mesurable, où π_i est la projection sur la i -ième coordonnée.

Définition 1.5.1 On appelle loi conjointe du vecteur $X = (X_1, \dots, X_n)$ la loi P_X de X sur

$$(\Pi_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i).$$

La loi P_{X_i} de chacune des variables aléatoires X_i est alors appelée loi marginale.

Proposition 1.5.2 Les variables aléatoires $(X_i)_{i=1, \dots, n}$ sont indépendantes si et seulement si on a :

$$P_X = \otimes_{i=1}^n P_{X_i}.$$

Preuve. D'après la définition de variables aléatoires indépendantes on a les équivalences suivantes :

$$\begin{aligned} & \prod_{i=1}^n X_i \\ \Leftrightarrow & \forall A_i \in \mathcal{A}_i, \text{ pour } i = 1, \dots, n : P \left[\bigcap_{i=1}^n X_i^{-1}(A_i) \right] = \prod_{i=1}^n P(X_i^{-1}(A_i)) \\ \Leftrightarrow & \forall A_i \in \mathcal{A}_i, \text{ pour } i = 1, \dots, n : P [X^{-1}(A_1 \times \dots \times A_n)] = \prod_{i=1}^n P_{X_i}(A_i) \\ \Leftrightarrow & \forall A_i \in \mathcal{A}_i, \text{ pour } i = 1, \dots, n : P_X(A_1 \times \dots \times A_n) = \prod_{i=1}^n P_{X_i}(A_i) \\ \Leftrightarrow & P_X = \otimes_{i=1}^n P_{X_i}. \quad \square \end{aligned}$$

Chapitre 2

Lois sur \mathbb{R} et lois sur \mathbb{R}^n

2.1 Variables aléatoires réelles

Soit (Ω, \mathcal{A}, P) un espace probabilisé et X une v.a. de (Ω, \mathcal{A}, P) vers $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

On a vu dans le chapitre précédent que si X et Y sont des v.a.r., alors pour tout couple (λ, μ) dans \mathbb{R}^2 , l'application $\lambda X + \mu Y$ est encore une variable aléatoire (i.e. l'ensemble de v.a.r. forme donc un espace vectoriel). Il a été également vu que XY est aussi une variable aléatoire et que si (X_n) est une suite de v.a.r. les applications

$$\inf_n X_n, \text{ et } \sup_n X_n$$

sont, par exemple, encore des variables aléatoires.

2.1.1 Fonction de répartition

Définition 2.1.1 On appelle fonction de répartition (f.d.r.) de la v.a.r. X , la fonction F_X définie sur \mathbb{R} par

$$F_X(x) = P_X(]-\infty, x]) = P(X \in]-\infty, x]) = P(X \leq x).$$

Proposition 2.1.2 La fonction de répartition F_X d'une v.a.r. X satisfait les propriétés suivantes :

- i) $0 \leq F_X(x) \leq 1$ pour tout x dans \mathbb{R} ;
- ii) La fonction F_X est croissante (au sens large) ;
- iii) la fonction F_X est continue à droite ;
- iv) On a

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \text{ et } \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

Preuve. La propriété i) est évidente puisque la probabilité de n'importe quel événement est toujours positive et inférieure à 1.

Pour établir le ii) considérons x et x' deux réels tels que $x \leq x'$. On a bien sûr l'inclusion :

$$]-\infty, x] \subset]-\infty, x']$$

et donc

$$P_X(]-\infty, x]) \leq P_X(]-\infty, x']).$$

Pour le *iii*) considérons une suite (h_n) de réels décroissante vers 0. Pour tout x dans \mathbb{R} , on a :

$$P_X([x, x + h_n]) = F_X(x + h_n) - F_X(x).$$

Or la suite d'intervalles $([x, x + h_n])_n$ est décroissante avec n . Ainsi il vient :

$$\lim_{n \rightarrow +\infty} P_X([x, x + h_n]) = P_X\left(\bigcap_n [x, x + h_n]\right) = P_X(\emptyset) = 0.$$

On en déduit que

$$\lim_{n \rightarrow +\infty} F_X(x + h_n) = F_X(x)$$

et la fonction F_X est donc bien continue à droite.

Pour établir le *iv*), considérons la suite d'intervalles $(]-\infty, -n])_n$ décroissante vers \emptyset quand n tend vers $+\infty$. On a :

$$\begin{aligned} \lim_{x \rightarrow -\infty} F_X(x) &= \lim_{n \rightarrow +\infty} F_X(-n) = \lim_{n \rightarrow +\infty} P_X(]-\infty, -n]) \\ &= P_X\left(\bigcap_{n=1}^{+\infty}]-\infty, -n]\right) = P_X(\emptyset) = 0. \end{aligned}$$

L'égalité

$$\lim_{x \rightarrow +\infty} F_X(x) = 1$$

s'obtient par un raisonnement similaire en considérant la suite d'intervalles $(]-\infty, n])_n$ croissante vers \mathbb{R} . \square

Remarquons que cette définition de la fonction de répartition est communément appelée version anglo-saxonne. On peut, dans certains ouvrages, trouver la version française obtenue en définissant

$$\tilde{F}_X(x) = P(X < x) = P_X(]-\infty, x[).$$

Elle est continue à gauche (et non plus à droite) et vérifie sinon les mêmes propriétés que précédemment. Notons que, dans le cas d'une f.d.r. discontinue (par exemple celui des v.a.r. discrètes), ces définitions donnent des résultats différents.

En effet, si x est un atome pour P_X , i.e.

$$P_X(\{x\}) = P(X = x) > 0,$$

alors on a :

$$F_X(x) = P(X \leq x) = P(X < x) + P(X = x) = \tilde{F}_X(x) + P(X = x).$$

Bien sûr, si P_X n'a pas d'atome, i.e. si P_X est une mesure continue,

$$F_X(x) = \tilde{F}_X(x).$$

Il est donc important de bien savoir quelle version de la fonction de répartition on utilise ! Par la suite, en ce qui nous concerne, nous utiliserons uniquement la version anglo-saxonne.

On admettra le théorème suivant :

Théorème 2.1.3 *Toute fonction F définie sur \mathbb{R} et vérifiant les propriétés (i), (ii), (iii) et (iv) est une fonction de répartition d'une v.a.r.*

Proposition 2.1.4 *Le saut $p_0 = F_X(x_0) - F_X(x_0^-)$ de la fonction de répartition F_X au point x_0 est égal à $P(X = x_0)$.*

Preuve. Soit (h_n) une suite de réels strictement positifs, décroissante vers 0. On a, pour tout n ,

$$P(X \in]x_0 - h_n, x_0]) = F_X(x_0) - F_X(x_0 - h_n).$$

Comme $(]x_0 - h_n, x_0])_n$ est une suite décroissante vers $\{x_0\}$, on a :

$$\begin{aligned} P_X(\{x_0\}) &= P_X\left(\bigcap_{n=1}^{+\infty}]x_0 - h_n, x_0]\right) \\ &= \lim_{n \rightarrow +\infty} P(]x_0 - h_n, x_0]) = F_X(x_0) - F_X(x_0^-). \quad \square \end{aligned}$$

Définition 2.1.5 *On appelle quantile d'ordre α (pour $0 < \alpha < 1$) de la loi de X , tout réel x_α tel que*

$$P(X \leq x_\alpha) \geq \alpha \text{ et } P(X \geq x_\alpha) \geq 1 - \alpha.$$

Terminologie particulière.

* Tout quantile d'ordre $1/2$ est appelé valeur médiane de la loi de X . S'il est unique il est appelé médiane de X et est noté $méd(X)$. Dans l'autre cas, l'ensemble des valeurs médianes constitue un intervalle médian.

- * Les quartiles Q_1 , Q_2 et Q_3 sont les quantiles (uniques) de la loi de X d'ordres respectivement 0.25, 0.50 et 0.75. L'intervalle inter-quartile est l'intervalle $[Q_1, Q_3]$.
- * Les 9 déciles sont obtenus en prenant $\alpha_k = k/10$ pour $k = 1, \dots, 9$.
- * Les 99 centiles sont obtenus en prenant $\alpha_k = k/100$ pour $k = 1, \dots, 99$.

Notons que la médiane sépare en deux intervalles de probabilité égale le support de X . De même, on a $P(X \in [Q_1, Q_3]) = 0.5$.

2.1.2 Lois discrètes

Définition 2.1.6 Une v.a.r. X est dite *discrète* si la loi P_X est une mesure de probabilité discrète.

D'après la définition d'une mesure discrète vue au chapitre 1, il existe donc une famille D (finie ou dénombrable) de réels telle que

$$P_X(\mathbb{R} \setminus D) = 0$$

et

$$\forall A \in \mathcal{B}_{\mathbb{R}}, \quad P_X(A) = \sum_{x \in A \cap D} P_X(\{x\}).$$

Les éléments de D sont donc les atomes, i.e. tels que :

$$P_X(\{x\}) > 0, \quad \forall x \in D.$$

Proposition 2.1.7 La fonction de répartition F_X d'une v.a.r. discrète est une fonction en escalier, dont les sauts sont situés sur les atomes, i.e. sur les éléments de D .

Preuve. Immédiate. □

Principales lois de v.a.r. discrètes.

a) Loi de Dirac

Soit $x_0 \in \mathbb{R}$. Une v.a.r. X est dite de loi de Dirac δ_{x_0} si elle est à valeurs dans \mathbb{R} et telle que $P_X = \delta_{x_0}$. On a donc, pour tout borélien A :

$$P_X(A) = \delta_{x_0}(A) = \begin{cases} 1 & \text{si } x_0 \in A \\ 0 & \text{sinon} \end{cases}$$

De plus on a :

$$\begin{aligned} P_X(\{x_0\}) &= P(X = x_0) = 1 \\ \text{et } P_X(\{x\}) &= P(X = x) = 0, \text{ pour tout } x \neq x_0. \end{aligned}$$

On dit que la v.a.r. X est presque sûrement (*p.s.*) égale à x_0 .

b) **Loi uniforme discrète**

Une v.a.r. X est dite variable aléatoire de loi uniforme discrète sur un ensemble D fini si on a, pour tout d dans D :

$$P_X(\{d\}) = P(X = d) = \frac{1}{\text{Card } D}.$$

b) **Loi de Bernoulli**

Une v.a.r. X est dite variable aléatoire de Bernoulli de paramètre p , (pour $p \in [0, 1]$) si elle est à valeurs dans $D = \{0, 1\}$ et si

$$\begin{aligned} P_X(\{1\}) &= P(X = 1) = p ; \\ P_X(\{0\}) &= P(X = 0) = 1 - p. \end{aligned}$$

C'est bien une v.a.r. discrète puisque $D = \{0, 1\}$ et $P(\mathbb{R} \setminus D) = 0$. On a en particulier : $\forall x \notin \{0, 1\}, P_X(\{x\}) = 0$.

Cette loi est, par exemple, utilisée pour modéliser le tirage au hasard dans une urne contenant des boules bleues en proportion p et des boules rouges en proportion $1 - p$. Si on note $\{X = 1\}$ l'événement "le résultat du tirage a donné une boule bleue" et $\{X = 0\}$ l'événement "le résultat est une boule rouge", alors la v.a. X suit une loi de Bernoulli de paramètre p .

c) **Loi binomiale**

Une v.a.r. X est dite de loi Binomiale de paramètres n et p (pour $n \in \mathbb{N}^*$ et $p \in [0, 1]$) si elle est à valeurs dans $D = \{0, 1, \dots, n\}$ et si

$$P(X = k) = P_X(\{k\}) = C_n^k p^k (1 - p)^{n-k},$$

pour $k = 0, 1, \dots, n$. On écrit $X \sim B(n, p)$.

Remarquons que, par exemple, si $n \geq 5$ et si $A = [1, 5]$, on a alors

$$P_X(A) = \sum_{x \in A \cap D} P_X(\{x\}) = \sum_{k=1}^5 P_X(\{k\}) = \sum_{k=1}^5 C_n^k p^k (1 - p)^{n-k}.$$

Cette loi apparaît, entre autres, lors de tirages avec remise dans une urne contenant des boules bleues en proportion p et des boules rouges en proportion $q = 1 - p$. Sur n tirages, si X est le nombre de boules bleues obtenues, la loi de X est une binomiale $B(n, p)$.

On montre facilement que la loi binomiale $B(n, p)$ est la loi de la somme de n v.a.r. indépendantes et de même loi de Bernoulli de paramètre p .

d) Loi géométrique

Une v.a.r. X est dite de loi géométrique de paramètre p , pour p compris entre 0 et 1, si elle est à valeurs dans $D = \mathbb{N}^*$ et si

$$P(X = k) = (1 - p)^{k-1} p.$$

On note $X \sim G(p)$.

Cette loi apparaît, par exemple, lors de tirages successifs, indépendants et avec remise dans une urne contenant des boules bleues en proportion p et des boules rouges en proportion $1 - p$.

Si X est le nombre de tirages effectués lors de l'apparition de la 1^{ère} boule bleue alors la loi de X est une géométrique $G(p)$. La v.a.r. X est donc le rang d'arrivée de la 1^{ère} boule bleue.

On peut aussi trouver dans la littérature la loi géométrique à valeurs dans \mathbb{N} et elle a pour probabilité élémentaire $P(X = k) = (1 - p)^k p$. Dans notre exemple, cette dernière donne la loi du nombre de boules rouges obtenues avant l'apparition de la 1^{ère} boule bleue.

e) Loi binomiale négative

Une v.a.r. X est dite de loi binomiale négative de paramètres n et p si elle est à valeurs dans $D = \mathbb{N}$ et si

$$P(X = k) = C_{n+k-1}^{n-1} p^n q^k, \text{ où } q = 1 - p.$$

Reprenons l'exemple de tirages successifs au hasard, indépendants et avec remise dans une urne contenant des boules bleues et rouges en proportion respectivement p et $q = 1 - p$.

Soit Y le nombre de tirages que l'on doit faire pour obtenir n boules bleues. Alors la v.a.r. $X = Y - n$ représentant donc le nombre de boules rouges obtenues avant d'avoir n boules bleues suit une loi binomiale négative.

On retrouve facilement que

$$P(X = k) = C_{n+k-1}^{n-1} p^n q^k$$

puisque l'événement $\{X = k\}$ signifie que sur les $k+n$ tirages on a eu k boules rouges et n boules bleues, dont l'une est la dernière tirée. La probabilité de chaque résultat élémentaire permettant à l'événement $\{X = k\}$ d'être vérifié est donc $p^n q^k$. Or l'événement $\{X = k\}$ est la réunion de C_{n+k-1}^{n-1} résultats élémentaires différents : une boule bleue étant tirée en dernier, il reste C_{n+k-1}^{n-1} façons différentes de placer les autres boules bleues.

Remarquons que, bien sûr, pour $n = 1$ la loi de Y est une loi géométrique $G(p)$.

f) Loi de Poisson

Une v.a.r. X est dite de loi de Poisson de paramètre λ , si elle est à valeurs dans $D = \mathbb{N}$ et si

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

On note $X \sim \mathcal{P}(\lambda)$.

Pour donner un exemple, anticipons un peu sur les lois continues. Considérons un lot de machines identiques, aux comportements indépendant et dont le temps d'attente avant l'arrivée d'une panne est une exponentielle de paramètre λ . On met en marche la première machine et quand survient la panne sur celle-ci, on la remplace immédiatement par une autre machine et ainsi de suite sur un intervalle de temps $[0, t]$. Le nombre de pannes observées durant cette période suit alors une loi de Poisson $\mathcal{P}(\lambda t)$.

g) Loi hypergéométrique

Une v.a.r. X est dite de loi hypergéométrique de paramètre (n, N, M) où n, N et M sont des entiers tels que $M < N$ et $n \leq N$, si elle est à valeurs dans $D = \mathbb{N} \cap [\max(0, n - (N - M)), \min(n, M)]$ et si

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}.$$

pour $\max(0, n - (N - M)) \leq k \leq \min(n, M)$

Cette loi apparaît lorsque l'on tire au hasard et sans remise dans une urne contenant M boules blanches et $N - M$ boules noires ($M < N$) (et donc en tout N boules). Si on tire au hasard et sans remise n boules successivement ($n \leq N$), le nombre X de boules blanches obtenues suit une loi hypergéométrique (n, N, M) . L'expression de la probabilité $P(X = k)$ se comprend alors toute seule.

2.1.3 Lois continues

Définition 2.1.8 *On dit qu'une v.a.r. X est de loi continue si sa loi P_X est une mesure de probabilité continue.*

Une v.a.r. (i.e. à valeurs dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$) continue est donc telle que :

$$\forall x \in \mathbb{R}, \quad P_X(\{x\}) = 0.$$

Proposition 2.1.9 *Une v.a.r. est continue si et seulement si sa fonction de répartition est continue.*

Preuve. Immédiate d'après la proposition au début de ce chapitre donnant l'expression des sauts d'une fonction de répartition. \square

Définition 2.1.10 *On dit que la loi P_X d'une v.a.r. X admet f_X comme densité s'il existe une telle fonction f_X positive et telle que pour tout $x \in \mathbb{R}$ on ait*

$$F_X(x) = \int_{-\infty}^x f_X(u) du.$$

Une v.a.r. (ou sa loi) qui admet une densité est dite absolument continue.

On verra à la fin de ce chapitre, dans la partie “extension de la notion de densité”, que cette définition est équivalente à l'existence d'une fonction f_X positive et telle que

$$\forall B \in \mathcal{B}_{\mathbb{R}}, \quad P_X(B) = P(X \in B) = \int_B f_X(x) dx$$

où l'intégrale est prise au sens de Lebesgue.

Remarquons ensuite que si X est une variable aléatoire absolument continue de densité f_X et si \tilde{f} est une fonction égale à f_X sauf sur un ensemble fini ou dénombrable de points de \mathbb{R} , alors pour tout x de \mathbb{R} , on a

$$\int_{-\infty}^x f_X(u) du = \int_{-\infty}^x \tilde{f}(u) du.$$

Ainsi, la fonction \tilde{f} est encore une densité pour X . Toutes les densités sont donc équivalentes pour cette relation et on appelle densité n'importe quel élément de cette classe d'équivalence.

Théorème 2.1.11 *Une fonction f sur \mathbb{R} est une densité de probabilité si et seulement si elle vérifie les trois assertions suivantes :*

- i) f est positive
- ii) f est mesurable.
- iii) f est intégrable et

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

Proposition 2.1.12 *Si f_X est continue sur un intervalle $[a, b]$, alors F_X est dérivable sur $[a, b]$ et on a $f_X = F'_X$.*

Preuve. On a, pour tout x dans l'intervalle $[a, b]$:

$$F_X(x) = \int_{-\infty}^x f(u)du = \int_{-\infty}^a f(u)du + \int_a^x f(u)du$$

et la proposition découle du résultat classique sur la dérivation de la fonction

$$x \mapsto \int_a^x f(u)du. \quad \square$$

Proposition 2.1.13 *Une variable aléatoire réelle absolument continue est continue mais la réciproque est fausse.*

Preuve. Si X est absolument continue, on a alors

$$P_X(\{x\}) = \int_x^x f_X(x)dx = 0,$$

pour tout x dans \mathbb{R} et la variable aléatoire X est bien continue.

On peut trouver des variables aléatoires continues sur \mathbb{R} mais qui ne sont pas absolument continues. Cependant, dans le cadre de ce cours, on se trouvera rarement dans cette situation. En revanche dans \mathbb{R}^2 , il est plus facile de trouver des variables aléatoires continues mais qui ne sont pas absolument continues. \square

Remarquons enfin qu'une variable aléatoire (ou sa loi) n'est pas soit discrète, soit continue, soit absolument continue. Le théorème suivant, dû à Lebesgue, précise ceci.

Théorème 2.1.14 Soit F une fonction de répartition. Alors il existe trois fonctions de répartition F_1 discrète, F_2 absolument continue et F_3 singulière (i.e. continue mais non absolument continue) et trois nombres réels α_1 , α_2 et α_3 positifs et de somme 1 tel que F puisse s'écrire sous la forme

$$F = \alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3.$$

Principales lois de probabilité sur \mathbb{R} absolument continues.

a) **Loi uniforme sur $[a, b]$**

Une v.a.r. X à valeurs dans $[a, b]$ est dite de loi uniforme sur cet intervalle si elle est absolument continue et admet pour densité

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

On note $X \sim U_{[a, b]}$.

Sa fonction de répartition est

$$F(x) = \begin{cases} 0 & \text{pour } x \leq a \\ \frac{x-a}{b-a} & \text{pour } a \leq x \leq b \\ 1 & \text{pour } x \geq b. \end{cases}$$

La loi uniforme la plus célèbre est celle dont le support est l'intervalle $[0, 1]$.

b) **Loi normale $N(\mu, \sigma^2)$**

Une v.a.r. X à valeurs dans \mathbb{R} est dite de loi normale de moyenne μ et de variance σ^2 si elle est absolument continue et admet pour densité

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

pour $x \in \mathbb{R}$. La loi $N(0, 1)$ est appelée loi normale centrée réduite.

Notons le résultat suivant

$$\begin{array}{ll} \text{si } X \sim N(\mu, \sigma^2) & \text{alors } \frac{X-\mu}{\sigma} \sim N(0, 1) \\ \text{si } X \sim N(0, 1) & \text{alors } \mu + \sigma X \sim N(\mu, \sigma^2). \end{array}$$

La fonction de répartition de la loi normale n'a pas d'expression explicite mais on l'exprime souvent en fonction de celle de la loi $N(0, 1)$, que l'on note souvent φ . On a

$$\varphi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

Ainsi, si $X \sim N(\mu, \sigma^2)$, alors

$$F_X(x) = \varphi\left(\frac{x - \mu}{\sigma}\right).$$

c) **Loi exponentielle**

Soit λ un réel strictement positif. Une v.a.r. X à valeurs dans \mathbb{R}_*^+ est dite de loi exponentielle de paramètre λ si elle est absolument continue et admet pour densité

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0, +\infty[}(x).$$

On note $X \sim \mathcal{E}(\lambda)$.

Sa fonction de répartition est :

$$F(x) = 1 - e^{-\lambda x},$$

pour tout x positif.

d) **Loi Gamma**

Rappelons en premier lieu l'expression de la fonction Gamma (ou seconde fonction d'Euler) pour tout α positif

$$\Gamma(\alpha) = \int_0^{+\infty} e^{-u} u^{\alpha-1} du.$$

On a les relations suivantes : $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ et si n est un entier $\Gamma(n) = (n-1)!$. On a enfin

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Une v.a.r. X à valeurs dans \mathbb{R}_*^+ est dite de loi Gamma $\gamma(\alpha, \beta)$, où α et β sont des réels strictement positifs, si elle est absolument continue et admet pour densité

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1} \mathbb{1}_{\mathbb{R}_*^+}(x).$$

Les paramètres α et β sont appelés respectivement paramètres de forme et d'échelle. On note $X \sim \gamma(\alpha, \beta)$.

Notons que si le paramètre de forme vaut 1, on retrouve une loi exponentielle $\mathcal{E}(\beta)$.

De plus une loi $\gamma(\frac{n}{2}, \frac{1}{2})$ pour n dans \mathbb{N}^* est aussi appelée loi du χ^2 à n degrés de liberté. On note $X \sim \chi^2(n)$.

e) **Loi Bêta** (de première espèce)

Comme précédemment rappelons en premier lieu l'expression de la première fonction d'Euler appelée aussi fonction Bêta. Elle est définie pour tout a et b positifs par

$$\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx.$$

Notons que l'on a :

$$\beta(a, b) = \beta(b, a) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Une v.a.r. X à valeurs dans $[0, 1]$ est dite de loi Bêta de paramètres a et b si elle est absolument continue et admet pour densité

$$f(x) = \frac{1}{\beta(a, b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{[0,1]}(x).$$

On note $X \sim \text{Bêta}(a, b)$.

f) **Loi de Student**

Une v.a.r. X à valeurs dans \mathbb{R} est dite de loi de Student à n degrés de liberté si elle est absolument continue de densité :

$$f(x) = \frac{1}{\sqrt{n}\beta(\frac{1}{2}, \frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

On note $X \sim T(n)$.

g) **Loi de Fisher**

Une v.a.r. X à valeurs dans \mathbb{R}^+ est dite de loi de Fisher à n et m degrés de liberté, si elle est absolument continue de densité :

$$f(x) = \frac{1}{\beta(\frac{n}{2}, \frac{m}{2})} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{x^{\frac{n}{2}-1}}{(m + nx)^{\frac{n+m}{2}}} \mathbb{1}_{\mathbb{R}^+}(x)$$

On note $X \sim F(n, m)$.

h) **Loi log-normale**

Une v.a.r. X à valeurs dans $]0, +\infty[$ est dite de loi log-normale de paramètre μ et σ^2 si la v.a.r. $Y = \log X$ est de loi normale $N(\mu, \sigma^2)$. On note $X \sim LN(\mu, \sigma^2)$.

Sa fonction de répartition est alors

$$F(x) = \varphi\left(\frac{\log x - \mu}{\sigma}\right) \mathbb{1}_{\mathbb{R}_*^+}(x),$$

où φ est toujours la fonction de répartition de la loi $N(0, 1)$.

Sa densité est :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left(-\frac{1}{2} \left(\frac{\log x - \mu}{\sigma}\right)^2\right) \mathbb{1}_{0,+\infty[}(x).$$

i) Loi de Cauchy

Une v.a.r. à valeurs dans \mathbb{R} est dite de loi de Cauchy $C(0, 1)$ si elle est absolument continue et admet pour densité

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2},$$

pour $x \in \mathbb{R}$.

2.1.4 Changement de variables

Le problème que l'on se propose d'étudier dans cette partie est la détermination de la loi de fonctions d'une v.a.r. dont on connaît la loi.

Soit donc X une v.a.r. de loi P_X et de fonction de répartition F_X . Soit φ une application mesurable de \mathbb{R} vers \mathbb{R} . La v.a.r. $Y = \varphi \circ X$ est donc encore une v.a.r. et on cherche à déterminer sa loi.

Une première méthode, convenant autant aux variables discrètes que continues, consiste à déterminer la fonction de répartition F_Y de Y .

On a, pour tout y dans \mathbb{R}

$$\begin{aligned} F_Y(y) &= P(Y \in]-\infty, y]) = P(\varphi \circ X \in]-\infty, y]) \\ &= P(X \in \varphi^{-1}(]-\infty, y])) = P_X(\varphi^{-1}(]-\infty, y])). \end{aligned}$$

Voyons deux exemples d'application de cette méthode.

Exemple 1. Supposons que la v.a.r. X suive une loi $N(0, 1)$ et posons $Y = X^2$. On a :

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y).$$

On constate déjà que l'on a : $F_Y(y) = 0$ si $y \leq 0$. Par ailleurs,

$$\begin{aligned} F_Y(y) &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P(X \leq \sqrt{y}) - P(X < -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \end{aligned}$$

car la v.a.r. X est continue. De plus, comme cette dernière est absolument continue sur \mathbb{R} et de densité f_X continue sur \mathbb{R} , sa f.d.r. F_X est dérivable

(de dérivée f_X). Par composition de fonctions dérivables, la f.d.r. F_Y est dérivable sur \mathbb{R}^+ (la v.a.r. Y est donc absolument continue) et la densité de Y est donc, pour $y > 0$:

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \\ &= \frac{1}{\sqrt{y}} \times \frac{1}{\sqrt{2\pi}} e^{-y/2} = \frac{\left(\frac{1}{2}\right)^{1/2}}{\sqrt{\pi}} y^{\frac{1}{2}-1} e^{-y/2} \mathbb{1}_{\mathbb{R}^+}(y). \end{aligned}$$

Ainsi, la loi de la v.a.r. Y est une $\gamma(\frac{1}{2}, \frac{1}{2})$ et le carré d'une loi normale centrée réduite suit une loi du $\chi^2(1)$. Ce résultat nous permet également de prouver l'égalité $\Gamma(1/2) = \sqrt{\pi}$ annoncée précédemment. \diamond

Exemple 2. Soit X une v.a.r. de fonction de répartition F_X continue et strictement croissante. Prenons $\varphi = F_X$ et cherchons la loi de la v.a.r.

$$Y = \varphi \circ X = F_X(X).$$

On a, pour tout y dans $[0, 1]$ (la f.d.r. F_Y étant nulle pour $y \leq 0$ et égale à 1 pour $y \geq 1$),

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(F_X(X) \leq y) = P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y. \end{aligned}$$

Par caractérisation de la fonction de répartition, on en déduit que la v.a.r. $F_X(X)$ est de la loi $U_{[0,1]}$. Ce résultat est souvent utilisé pour faire des simulations. \diamond

Une deuxième méthode pour calculer la loi de $\varphi(X) = Y$ est donnée par le théorème suivant et ne convient que pour des variables aléatoires absolument continues.

Théorème 2.1.15 *Soient S et T deux ouverts de \mathbb{R} et X une v.a.r. absolument continue à valeurs dans S et de densité f_X . Soit φ une bijection de S vers $T = \text{Im } \varphi$, continûment différentiable ainsi que son inverse (φ est dite \mathcal{C}^1 -difféomorphisme). Alors, la v.a.r. $Y = \varphi(X)$ est absolument continue, à valeurs dans T et de densité :*

$$f_Y(y) = f_X(\varphi^{-1}(y)) |(\varphi^{-1})'(y)| \mathbb{1}_{\text{Im } \varphi}(y).$$

Preuve. On a :

$$F_Y(y) = P_X(\varphi^{-1}([-\infty, y])) = \int_{\{x: \varphi(x) \leq y\}} f_X(x) \, dx.$$

Puisque φ est inversible et φ^{-1} continûment différentiable, le changement de variable $x = \varphi^{-1}(u)$ dans l'intégrale donne

$$F_Y(y) = \int_{-\infty}^y f_X(\varphi^{-1}(u)) |(\varphi^{-1})'(u)| \, du.$$

Donnons une justification de l'apparition de la valeur absolue dans l'expression précédente. La fonction $(\varphi^{-1})'$ étant continue, on peut séparer son domaine en intervalles où elle est positive et en intervalle où elle est négative. Sur ces intervalles φ^{-1} est donc respectivement croissante et décroissante.

Dans les intervalles correspondant au premier cas (i.e. $(\varphi^{-1})' \geq 0$) on a bien la valeur absolue. Dans le second, où $(\varphi^{-1})' \leq 0$, comme l'ordre entre les bornes de l'intervalle est interverti dans le changement de variable, on retrouve bien la valeur absolue quand on intègre sur un intervalle croissant pour les u . \square

Exemple. Appliquons cette formule pour le calcul de la densité de la loi log-normale. On a vu qu'une v.a.r. X est de loi $LN(\mu, \sigma^2)$ si la v.a.r. $Y = \log X$ est de loi $N(\mu, \sigma^2)$. La fonction $\varphi = \exp$ est clairement un \mathcal{C}^1 -difféomorphisme de \mathbb{R} dans \mathbb{R}_*^+ d'inverse $\varphi^{-1} = \ln$ et telle que :

$$(\varphi^{-1})'(x) = \frac{1}{x}.$$

Ainsi, d'après la formule du changement de variable, on a :

$$\begin{aligned} f_X(x) &= f_Y(\varphi^{-1}(x)) |(\varphi^{-1})'(x)| \mathbb{1}_{\text{Im}\varphi}(x) \\ &= \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{\ln x - \mu}{\sigma} \right)^2} \times \frac{1}{x} \mathbb{1}_{\mathbb{R}_*^+}(x) \end{aligned}$$

et on retrouve bien la densité de la loi log-normale donnée précédemment. \diamond

Notons, avant d'aborder la partie suivante, que la fonction caractéristique et la transformée de Laplace introduits au chapitre 4, nous permettront de disposer de nouveaux outils pour répondre à cette question du changement de variables.

2.2 Vecteurs aléatoires

On a déjà vu que l'on appelle vecteur aléatoire toute variable aléatoire à valeurs dans $(\mathbb{R}^n, B_{\mathbb{R}^n}) = (\mathbb{R}, B_{\mathbb{R}})^{\otimes n}$. On notera X_i , la i -ième coordonnée du vecteur X et rappelons que celle-ci est encore une variable aléatoire.

2.2.1 Fonction de répartition

Définition 2.2.1 On appelle fonction de répartition (conjointe) du vecteur aléatoire $X = (X_1, \dots, X_n)$ l'application F_X définie sur \mathbb{R}^n et à valeurs dans $[0, 1]$ par :

$$F_X(x) = P_X([-\infty, x_1] \times \dots \times [-\infty, x_n]) = P_X\left(\bigcap_{i=1}^n \{X \leq x_i\}\right),$$

où $x = (x_1, \dots, x_n)$ est un vecteur de \mathbb{R}^n .

Proposition 2.2.2 On a :

$$\lim_{\forall i, x_i \rightarrow -\infty} F_X(x) = 0 \quad \text{et} \quad \lim_{\forall i, x_i \rightarrow +\infty} F_X(x) = 1.$$

Définition 2.2.3 Tout sous vecteur aléatoire de dimension strictement inférieure à n et extrait du vecteur X est appelé variable aléatoire marginale.

Ainsi les variables aléatoires X_1, \dots, X_{n-1} et X_n sont chacune des marginales de X mais (X_1, X_2) et (X_1, X_{n-1}) sont aussi des variables aléatoires marginales, etc...

Proposition 2.2.4 La fonction de répartition conjointe d'un vecteur aléatoire $X = (X_1, \dots, X_n)$ permet de déterminer les fonctions de répartition de toutes les marges.

Preuve. On a, pour tout $i = 1, \dots, n$,

$$F_{X_i}(x_i) = \lim_{\forall j \neq i, x_j \rightarrow +\infty} F_X(x) = F_X(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty).$$

De même :

$$F_{X_1, X_2}(x_1, x_2) = \lim_{\forall i > 2, x_i \rightarrow +\infty} F_X(x) = F_X(x_1, x_2, +\infty, \dots, +\infty)$$

et ainsi de suite pour les autres marges. □

Nous attirons l'attention du lecteur sur le fait que la réciproque de cette proposition est fautive. Se donner toutes les fonctions de répartitions marginales ne suffit pas pour définir la fonction de répartition conjointe.

Exemple de vecteur aléatoire discret : la loi multinomiale. Une variable aléatoire $X = (N_1, \dots, N_m)$ suit une loi multinomiale de paramètres (n, p_1, \dots, p_m) , où n est un entier et p_1, \dots, p_m sont des réels strictement positifs tels que $\sum_{i=1}^m p_i = 1$, si

$$P_X(n_1, \dots, n_m) = \frac{n!}{n_1! \dots n_m!} p_1^{n_1} \dots p_m^{n_m}$$

pour (n_1, \dots, n_m) tels que :

$$\sum_{i=1}^m n_i = n.$$

Cette loi est, par exemple, utilisée pour modéliser le tirage avec remise dans une urne contenant des boules de m couleurs différentes en proportion respective p_1, \dots, p_m . Si on effectue n tirages dans cette urne, alors la loi conjointe du vecteur $X = (N_1, \dots, N_m)$, où les nombres N_1, \dots, N_m sont les nombres de boules obtenues pour chaque couleur, est de loi multinomiale de paramètres (n, p_1, \dots, p_m) . \diamond

2.2.2 Densité de probabilité

Définition 2.2.5 On dit que le vecteur aléatoire $X = (X_1, \dots, X_n)$ (ou sa loi) est absolument continu(e) si il existe une fonction mesurable

$$f : (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}) \rightarrow (\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$$

telle que, pour tout $x = (x_1, \dots, x_n)$ dans \mathbb{R}^n , on ait :

$$P_X([-\infty, x_1] \times \dots \times [-\infty, x_n]) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(u_1, \dots, u_n) du_1 \dots du_n.$$

La fonction f_X est appelée densité de probabilité conjointe du vecteur X .

Proposition 2.2.6 Toute densité de probabilité conjointe f_X de \mathbb{R}^n vérifie les trois assertions suivantes :

- i) f_X est positive ;
- ii) f_X est mesurable ;

iii) f_X est intégrable et

$$\int_{\mathbb{R}^n} f_X(x_1, \dots, x_n) dx_1 \cdots dx_n = 1.$$

Réciproquement toute fonction f_X dans \mathbb{R}^n vérifiant i), ii) et iii) est une densité de probabilité.

On verra à la fin de ce chapitre, dans la partie “extension de la notion de densité”, que cette définition est équivalente à l’existence d’une fonction f_X définie sur \mathbb{R}^n , positive, mesurable et telle que, pour tout borélien B de \mathbb{R}^n , on ait

$$P_X(B) = P(X \in B) = \int_B f_X(x_1, \dots, x_n) dx_1 \cdots dx_n$$

Notons par ailleurs que si la densité f_X est continue au point (x_1^0, \dots, x_n^0) , on a :

$$f_X(x_1^0, \dots, x_n^0) = \frac{\partial^n F}{\partial x_1 \cdots \partial x_n}(x_1^0, \dots, x_n^0).$$

En fait, on peut montrer que cette propriété est toujours vraie sauf sur un ensemble de mesure de Lebesgue sur \mathbb{R}^n nul.

De même que précédemment nous avons déterminé les fonctions de répartitions marginales d’un vecteur aléatoire, nous allons maintenant voir comment exprimer les densités marginales en fonction de la densité conjointe.

On a, pour tout x_i dans \mathbb{R} ,

$$\begin{aligned} P_{X_i}([-\infty, x_i]) &= P(X \in \mathbb{R} \times \cdots \times \mathbb{R} \times [-\infty, x_i] \times \mathbb{R} \times \cdots \times \mathbb{R}) \\ &= \int_{\mathbb{R} \times \cdots \times [-\infty, x_i] \times \cdots \times \mathbb{R}} f_X(u_1, \dots, u_n) du_1 \cdots du_n \\ &= \int_{[-\infty, x_i]} g_i(u_i) du_i, \end{aligned}$$

où

$$g_i(u_i) = \int_{\mathbb{R}^{n-1}} f_X(u_1, \dots, u_n) du_1 \cdots du_{i-1} du_{i+1} \cdots du_n.$$

La v.a.r. X_i est donc absolument continue et sa densité est g_i .

De même soit $Z = (X_1, \dots, X_k)$, pour $k < n$, un vecteur extrait de X . Pour tout réel (x_1, \dots, x_k) de \mathbb{R}^k , on a

$$\begin{aligned} &P_Z([-\infty, x_1] \times \cdots \times [-\infty, x_k]) \\ &= P_X([-\infty, x_1] \times \cdots \times [-\infty, x_k] \times \mathbb{R} \times \cdots \times \mathbb{R}) \\ &= \int_{[-\infty, x_1] \times \cdots \times [-\infty, x_k]} g(u_1, \dots, u_k) du_1 \cdots du_k, \end{aligned}$$

où

$$g(u_1, \dots, u_k) = \int_{\mathbb{R}^{n-k}} f_X(u_1, \dots, u_n) du_{k+1} \cdots du_n.$$

Comme, par permutation, on peut mettre toutes les marges sous cette forme, on a montré le résultat suivant :

Proposition 2.2.7 *Si X est un vecteur aléatoire absolument continu, tout vecteur aléatoire marginal est également absolument continu et sa densité est obtenue en intégrant la densité conjointe de X par rapport aux coordonnées restantes.*

2.2.3 Loi conditionnelle et densité conditionnelle

Variables aléatoires discrètes

Supposons que $Z = (X, Y)$ soit un vecteur dans $\mathbb{R}^n \times \mathbb{R}^m$ de variables aléatoires discrètes, prenant ses valeurs sur un ensemble fini ou dénombrable D tel que : $\forall z \in D, P_Z(\{z\}) > 0$.

Notons respectivement I et J , parties de \mathbb{R}^n et \mathbb{R}^m , les ensembles des atomes des loïs P_X et P_Y , i.e.

$$\begin{aligned} & \forall x \in I : P_X(\{x\}) > 0, \quad \text{où } x = (x_1, \dots, x_n) \\ \text{et} \quad & \forall y \in J : P_Y(\{y\}) > 0, \quad \text{où } y = (y_1, \dots, y_m). \end{aligned}$$

Pour tout x dans I , la mesure discrète définie sur J par :

$$P_Y^{X=x}(\{y\}) = P(Y = y / X = x) = \frac{P(Y = y \cap X = x)}{P(X = x)}$$

est une probabilité discrète sur \mathbb{R}^m , de support J .

Définition 2.2.8 *Pour tout x dans I , la fonction $P_Y^{X=x}$ définie sur J et à valeurs dans $[0, 1]$ est appelée loi de probabilité de Y conditionnelle à $X = x$.*

On peut, bien sûr, de manière symétrique définir la loi de probabilité de X conditionnelle à $Y = y$.

Variables aléatoires continues

Soit $Z = (X, Y)$ une variable aléatoire à valeurs dans $\mathbb{R}^n \times \mathbb{R}^m$ de loi absolument continue et de densité f_Z . On a vu que les v.a.r. X et Y sont également absolument continues et possèdent donc des densités f_X et f_Y . Posons

$$A = [x \in \mathbb{R}^n : f_X(x) > 0].$$

Ayant trivialement

$$P(X \in \mathbb{R}^n) = P(X \in A) + P(X \in \bar{A}).$$

et

$$P(X \in \bar{A}) = \int_{\bar{A}} f_X(x_1, \dots, x_n) dx_1 \cdots dx_n = 0,$$

puisque la densité f_X est identiquement nulle sur \bar{A} , on a

$$P(X \in A) = P(X \in \mathbb{R}^n) = 1.$$

Ainsi, pour tout x dans A , donc pour P_X -presque-tout x dans \mathbb{R}^n , on peut considérer l'application

$$\begin{aligned} \mathbb{R}^m &\rightarrow \mathbb{R}^+ \\ y &\mapsto \frac{f_Z(x, y)}{f_X(x)}. \end{aligned}$$

Cette fonction est positive, mesurable et intégrale sur \mathbb{R}^m telle que :

$$\int_{\mathbb{R}^m} \frac{f_Z(x, y)}{f_X(x)} dy = \frac{1}{f_X(x)} \int_{\mathbb{R}^m} f_Z(x, y) dy = 1.$$

C'est donc une densité de probabilité sur $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$.

Définition 2.2.9 *La fonction*

$$\begin{aligned} f_Y^{X=x} : \quad \mathbb{R}^m &\rightarrow \mathbb{R}^+ \\ y &\mapsto \frac{f_Z(x, y)}{f_X(x)} \end{aligned}$$

est une densité de probabilité sur \mathbb{R}^m et est appelée densité conditionnelle de Y sachant que $X = x$. On note $P_Y^{X=x}$ la loi de probabilité associée, appelée loi de Y sachant que $X = x$.

Bien sûr, on définit de manière tout à fait similaire la loi de X sachant que $Y = y$.

Il résulte de cette définition que, pour P_X -presque-tout x de \mathbb{R}^n , on a :

$$f_Z(x, y) = f_Y^{X=x}(y) f_X(x).$$

Ainsi, si on connaît la densité marginale f_X et la densité conditionnelle $f_Y^{X=x}$, on a immédiatement l'expression de la densité conjointe $f_Z(x, y)$.

Nous attirons l'attention du lecteur sur le fait que bien que l'on dise densité de Y sachant que $X = x$ ou loi de Y sachant que $X = x$, il ne s'agit pas d'une probabilité conditionnelle à l'événement $\{X = x\}$ car celui-ci est de probabilité nulle. Cela n'aurait pas de sens.

2.2.4 Changement de variables

La question que l'on se pose dans ce paragraphe est la même que celle vue précédemment en unidimensionnel. Soit X un vecteur aléatoire dans \mathbb{R}^n et φ une application mesurable de \mathbb{R}^n vers \mathbb{R}^n . On veut déterminer la loi du vecteur aléatoire $\varphi(X)$. Rappelons en premier lieu que le jacobien d'une fonction

$$\begin{aligned} H : \quad \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ x = (x_1, \dots, x_n) &\mapsto H(x) = (h_1(x), h_2(x), \dots, h_n(x)) \end{aligned}$$

est le déterminant de la matrice des dérivées premières, i.e.

$$J_H = \begin{vmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \dots & \frac{\partial h_1}{\partial x_n} \\ \frac{\partial h_2}{\partial x_1} & & & \\ \vdots & & & \\ \frac{\partial h_n}{\partial x_1} & & & \frac{\partial h_n}{\partial x_n} \end{vmatrix}.$$

Théorème 2.2.10 *Soit X un vecteur aléatoire à valeurs dans un ouvert S de \mathbb{R}^n , de loi absolument continue et de densité f_X . Soit φ un \mathcal{C}^1 -difféomorphisme de S vers un ouvert T (i.e. une bijection continûment différentiable et de réciproque également continûment différentiable). Alors le vecteur aléatoire $Y = \varphi(X)$ est absolument continu et de densité, pour $y = (y_1, \dots, y_n) \in \mathbb{R}^n$,*

$$f_Y(y) = f_X(\varphi^{-1}(y)) |J_{\varphi^{-1}}(y)| \mathbb{1}_{\text{Im}\varphi}(y),$$

où $J_{\varphi^{-1}}$ est le jacobien de la fonction φ^{-1} .

Notons que, parfois, pour des raisons de calcul, il est plus simple de déterminer J_φ et on utilise alors l'égalité $J_{\varphi^{-1}} = J_\varphi^{-1}$.

Nous avons également une méthode assez proche pour déterminer la loi d'une transformée d'un vecteur aléatoire, basée sur la caractérisation suivante :

Théorème 2.2.11 *Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^n . Pour que sa loi P_X , absolument continue, soit de densité f_X il faut et il suffit que pour toute fonction borélienne bornée $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que $\varphi(X)$ soit intégrable ou positive, on ait :*

$$\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}^n} \varphi(x) f_X(x) dx.$$

2.2.5 Indépendance

Soit une famille $(X_i)_{i=1,\dots,n}$ de variables aléatoires à valeurs respectivement dans $(E_i, \mathcal{B}_i)_{i=1,\dots,n}$. Rappelons que la famille $(X_i)_{i=1,\dots,n}$ est indépendante si, pour toute famille d'événements $(B_i)_{i=1,\dots,n}$ où B_i appartient à \mathcal{B}_i pour tout i , on a :

$$P\left(\bigcap_{i=1}^n \{X_i \in B_i\}\right) = \prod_{i=1}^n P(\{X_i \in B_i\}),$$

ce qui, en notant $X = (X_1, \dots, X_n)$, est équivalent à

$$P_X = \otimes_{i=1}^n P_{X_i}.$$

Théorème 2.2.12 *Il y a équivalence entre les assertions suivantes :*

- i) *la famille $(X_i)_{i=1,\dots,n}$ est une famille de v.a.r. indépendantes ;*
- ii) *la fonction de répartition conjointe F_X est le produit des fonctions de répartitions marginales, i.e. :*

$$F_X = \prod_{i=1}^n F_{X_i}.$$

Si de plus la v.a.r. X est absolument continue sur \mathbb{R}^n de densité f_X (continue), les assertions précédentes sont encore équivalentes à

- iii) *la densité conjointe est le produit des densités marginales, i.e. :*

$$f_X(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i).$$

Enfin, si $n = 2$, celles-ci sont encore équivalentes à chacune des deux suivantes

- iv) $f_{X_2}^{X_1=x_1} = f_{X_2}$;
- v) $f_{X_1}^{X_2=x_2} = f_{X_1}$.

Remarque. Bien sûr on obtient le même style d'équivalence que les points *iii*), *iv*) et *v*) pour des variables aléatoires discrètes, en remplaçant les densités par rapport à Lebesgue par les densités par rapport à la mesure de dénombrement, i.e. en utilisant les probabilités discrètes. \diamond

Preuve.

i) \Rightarrow *ii*) : On a

$$F_X(x) = P\left(\bigcap_{i=1}^n \{X_i \leq x_i\}\right) = \prod_{i=1}^n P(\{X_i \leq x_i\}) = \prod_{i=1}^n F_{X_i}(x_i).$$

ii) \Rightarrow *i*) : Pour tout $x = (x_1, \dots, x_n)$ de \mathbb{R}^n on a :

$$P_X([-\infty, x_1] \times \dots \times [-\infty, x_n]) = \prod_{i=1}^n P_{X_i}([-\infty, x_i])$$

et on peut montrer que cela est suffisant pour avoir

$$P_X = \otimes_{i=1}^n P_{X_i}.$$

ii) \Rightarrow *iii*) : Par hypothèse, on a

$$F_X = \prod_{i=1}^n F_{X_i}.$$

Si les densités $(f_{X_i})_{i=1, \dots, n}$ sont continues, on a :

$$\frac{\partial^n F_X}{\partial x_1 \dots \partial x_n} = f_{X_1} \dots f_{X_n}.$$

iii) \Rightarrow *ii*) : Soit $x = (x_1, \dots, x_n)$ un réel de \mathbb{R}^n , on a :

$$\begin{aligned} & P\left(\bigcap_{i=1}^n \{X_i \in]-\infty, x_i]\}\right) = P_X([-\infty, x_1] \times \dots \times [-\infty, x_n]) \\ &= \int_{]-\infty, x_1] \times \dots \times [-\infty, x_n]} f_{X_1}(u_1) \dots f_{X_n}(u_n) du_1 \dots du_n \\ &= \int_{]-\infty, x_1]} f_{X_1}(u_1) du_1 \dots \int_{]-\infty, x_n]} f_{X_n}(u_n) du_n \\ &= P(X_1 \in]-\infty, x_1]) \dots P(X_n \in]-\infty, x_n]) \end{aligned}$$

et, par définition, la famille $(X_i)_{i=1, \dots, n}$ est donc bien indépendante.

Enfin pour $n = 2$, on a les équivalences suivantes :

$$\begin{aligned} X_1 \perp\!\!\!\perp X_2 &\Leftrightarrow f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) \\ &\Leftrightarrow f_{X_1}(x_1) = f_{X_1}^{X_2=x_2}(x_1) \\ &\Leftrightarrow f_{X_2}(x_2) = f_{X_2}^{X_1=x_1}(x_2). \end{aligned} \quad \square$$

Exemple. Soit (X, Y) un couple de variables aléatoires indépendantes dont chacune suit une loi $N(0, 1)$. Posons

$$U = \frac{X+Y}{\sqrt{2}} \quad \text{et} \quad V = \frac{X-Y}{\sqrt{2}}.$$

Le couple (U, V) est alors formé de variables aléatoires réelles indépendantes et dont chacune suit une loi $N(0, 1)$.

En effet, puisque les v.a.r. X et Y sont indépendantes et de même loi $N(0, 1)$, la densité du couple (X, Y) est

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

La fonction φ de \mathbb{R}^2 dans \mathbb{R}^2 définie par

$$\varphi(x, y) = \left(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}} \right)$$

est un \mathcal{C}^1 -difféomorphisme. Déterminons l'expression de son inverse. On a :

$$\begin{cases} u = \frac{x+y}{\sqrt{2}} \\ v = \frac{x-y}{\sqrt{2}} \end{cases} \Leftrightarrow \begin{cases} \frac{2x}{\sqrt{2}} = u+v \\ \frac{2y}{\sqrt{2}} = u-v \end{cases} \Leftrightarrow \begin{cases} x = \frac{u+v}{\sqrt{2}} \\ y = \frac{u-v}{\sqrt{2}}. \end{cases}$$

Ainsi le Jacobien de φ^{-1} est :

$$J_{\varphi^{-1}}(u, v) = \begin{vmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{vmatrix} = -\frac{1}{2} - \frac{1}{2} = -1.$$

On applique alors le théorème du changement de variable et la densité du couple (u, v) est :

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}(\varphi^{-1}(u, v)) |J_{\varphi^{-1}}(u, v)| \mathbb{1}_{\text{Im}\varphi}(u, v) \\ &= \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left[\left(\frac{u+v}{2}\right)^2 + \left(\frac{u-v}{2}\right)^2\right]\right\} |-1| \mathbb{1}_{\mathbb{R}^2}(u, v) \\ &= \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(u^2 + v^2)\right\} \mathbb{1}_{\mathbb{R}^2}(u, v) \\ &= \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \mathbb{1}_{\mathbb{R}}(u) \cdot \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \mathbb{1}_{\mathbb{R}}(v). \end{aligned}$$

Les variables aléatoires U et V sont donc indépendantes puisque la densité conjointe se factorise et elles sont toutes les deux de loi $N(0, 1)$. \diamond

Ce théorème peut également nous permettre de déterminer la loi d'une fonction de deux variables aléatoires. Donnons en un exemple en déterminant la loi de la somme de deux variables aléatoires.

Soit (X, Y) un couple aléatoire sur \mathbb{R}^2 , absolument continu et de densité $f_{X,Y}$. La fonction φ définie sur \mathbb{R}^2 et à valeurs dans \mathbb{R}^2 par

$$\varphi(x, y) = (x + y, y)$$

est \mathcal{C}^1 -un difféomorphisme. L'équivalence :

$$\begin{cases} u = x + y \\ v = y \end{cases} \Leftrightarrow \begin{cases} x = u - v \\ y = v. \end{cases}$$

permet de déterminer l'expression de φ^{-1} et d'en déduire son Jacobien :

$$J_{\varphi^{-1}}(u, v) = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1.$$

Ainsi, d'après le théorème du changement de variable, la densité du couple $(U, V) =_v \text{arphi}(X, Y)$ est :

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}(\varphi^{-1}(u, v)) |J_{\varphi^{-1}}(u, v)| \mathbb{1}_{\text{Im}\varphi}(u, v) \\ &= f_{X,Y}(u - v, v) \mathbb{1}_{\mathbb{R}^2}(u, v). \end{aligned}$$

On peut alors déterminer la densité de la loi marginale de U

$$f_U(u) = \int_{\mathbb{R}} f_{X,Y}(u - v, v) dv,$$

pour tout u dans \mathbb{R} . De plus, si les variables aléatoires X et Y sont indépendantes de densité respective f_X et f_Y , on a :

$$f_{X+Y}(u) = \int_{\mathbb{R}} f_X(u - v) f_Y(v) dv,$$

pour tout u dans \mathbb{R} . On vient ainsi d'établir le théorème :

Théorème 2.2.13 *La loi de la somme de deux variables aléatoires indépendantes, absolument continues et de densité f_X et f_Y respectivement est de densité*

$$f_{X+Y}(u) = \int_{\mathbb{R}} f_X(u - v) f_Y(v) dv,$$

*pour tout u dans \mathbb{R} . Cette densité est appelée la convoluée de f_X et f_Y . On note $f_X * f_Y$ le produit de convolution.*

2.3 Extension de la notion de densité

2.3.1 Intégrale par rapport à une mesure

Définition 2.3.1 *On appelle fonction étagée une fonction mesurable f , définie sur un espace probabilisable (Ω, \mathcal{A}) à valeurs dans $(\overline{\mathbb{R}}, \mathcal{B}_{\mathbb{R}})$, ne prenant qu'un nombre fini de valeurs $(y_i)_{i=1, \dots, n}$*

On note $A_i = f^{-1}(\{y_i\})$. Puisque la fonction f est mesurable, les ensembles A_i sont tous dans \mathcal{A} et on écrit :

$$f = \sum_{i=1}^n y_i \mathbb{1}_{A_i}.$$

Notons \mathcal{E}^+ l'ensemble des fonctions étagées positives.

Théorème 2.3.2 *Toute fonction réelle positive est mesurable si elle est limite croissante d'une suite de fonctions étagées positives.*

Définition 2.3.3 *Soit f un élément de \mathcal{E}^+ d'expression*

$$f = \sum_{i=1}^n y_i \mathbb{1}_{A_i}.$$

Le réel $\int f d\mu$ (éventuellement infini) défini par

$$\int f d\mu = \sum_{i=1}^n y_i \mu(A_i)$$

est appelé intégrale de f par rapport à μ .

Remarquons que pour toute fonction f de \mathcal{E}^+ son intégrale $\int f d\mu$ par rapport à μ est positive. De plus si A est un événement de \mathcal{A} , on a

$$\mu(A) = \int \mathbb{1}_A d\mu.$$

Par ailleurs on peut montrer que si l'on peut écrire :

$$f = \sum_{i=1}^n y_i \mathbb{1}_{A_i} = \sum_{j=1}^m y_j \mathbb{1}_{B_j}$$

alors

$$\sum_{i=1}^n y_i \mu(A_i) = \sum_{j=1}^m y_j \mu(B_j)$$

et l'intégrale $\int f d\mu$ ne dépend donc pas de l'écriture choisie pour f .

On note enfin

$$\begin{aligned} \int_A f d\mu &= \int \mathbb{1}_A f d\mu = \int \left(\sum_{i=1}^n y_i \mathbb{1}_{A \cap A_i} \right) d\mu \\ &= \sum_{i=1}^n y_i \mu(A \cap A_i). \end{aligned}$$

Proposition 2.3.4 *L'intégrale vérifie les propriétés suivantes :*

- 1) *Pour toutes fonctions f et g de \mathcal{E}^+ et pour tout couple (λ, γ) dans $\mathbb{R}^+ \times \mathbb{R}^+$, la fonction $\lambda f + \gamma g$ est dans \mathcal{E}^+ et on a la propriété de linéarité suivante :*

$$\int (\lambda f + \gamma g) d\mu = \lambda \int f d\mu + \gamma \int g d\mu.$$

- 2) *Si f et g sont des fonctions de \mathcal{E}^+ telles que $f \leq g$ alors*

$$\int f d\mu \leq \int g d\mu.$$

Preuve. 1) Soient f et g deux fonctions étagées quelconques d'expressions

$$f = \sum_{i=1}^n x_i \mathbb{1}_{A_i} \quad \text{et} \quad g = \sum_{j=1}^m y_j \mathbb{1}_{B_j}.$$

Les familles $(A_i)_{i=1,\dots,n}$ et $(B_j)_{j=1,\dots,m}$ forment chacune des familles d'ensembles disjoints. On peut donc, sans perte de généralités, supposer qu'elles forment chacune une partition de Ω , i.e. qu'elles vérifient de plus

$$\bigcup_{i=1}^n A_i = \Omega = \bigcup_{j=1}^m B_j.$$

On peut alors écrire

$$A_i = \bigcup_{j=1}^m (A_i \cap B_j) \quad \text{et} \quad B_j = \bigcup_{i=1}^n (B_j \cap A_i).$$

D'où l'expression de la fonction étagée $f + g$

$$f + g = \sum_{i=1}^n \sum_{j=1}^m (x_i + y_j) \mathbb{1}_{A_i \cap B_j}$$

et, par définition de l'intégrale,

$$\begin{aligned} \int (f + g) d\mu &= \sum_{i=1}^n \sum_{j=1}^m (x_i + y_j) \mu(A_i \cap B_j) \\ &= \sum_{i=1}^n x_i \left(\sum_{j=1}^m \mu(A_i \cap B_j) \right) + \sum_{j=1}^m y_j \left(\sum_{i=1}^n \mu(B_j \cap A_i) \right) \\ &= \int f d\mu + \int g d\mu. \end{aligned}$$

La démonstration de l'égalité

$$\int \lambda f d\mu = \lambda \int f d\mu$$

est évidente.

2) On peut écrire $g = f + g - f$ où f et $g - f$ sont des fonctions de \mathcal{E}^+ . Par linéarité de l'intégrale, on a :

$$\int g d\mu = \int f d\mu + \int (g - f) d\mu,$$

d'où on tire :

$$\int g d\mu \geq \int f d\mu. \quad \square$$

Étendons dans un premier temps la notion d'intégrale aux fonctions mesurables positives mais non nécessairement étagées.

Définition 2.3.5 *Soit f une fonction réelle, mesurable et positive. On a vu dans un théorème précédent qu'il existe une suite (f_n) de fonctions étagées croissantes vers f . L'intégrale de f par rapport à la mesure μ est alors*

$$\int f d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu.$$

Soit maintenant une fonction réelle f mesurable de signe quelconque. Définissons les parties positives et négatives de f par :

$$\begin{aligned} f^+ &= f\mathbb{1}_{\{f \geq 0\}} = \max(f, 0) \\ f^- &= -f\mathbb{1}_{\{f \leq 0\}} = -\inf(f, 0). \end{aligned}$$

Comme produit de fonction mesurables f^+ et f^- sont mesurables et elles sont toutes les deux positives. On peut donc définir, d'après ce qui précède, les intégrales :

$$\int f^+ d\mu \quad \text{et} \quad \int f^- d\mu.$$

Remarquons que l'on peut exprimer la fonction f en fonction de f^+ et f^- par l'égalité :

$$f = f^+ - f^-.$$

Définition 2.3.6 Une fonction réelle f mesurable est dite μ -intégrable si on a :

$$\int f^+ d\mu < +\infty \quad \text{et} \quad \int f^- d\mu < +\infty.$$

Le réel

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

est alors appelé intégrale de f par rapport à μ .

Proposition 2.3.7 Une fonction réelle f est μ -intégrable si, et seulement si,

$$\int |f| d\mu < +\infty.$$

Preuve. Grâce à l'égalité $|f| = f^+ + f^-$, on a les équivalences suivantes :

$$\begin{aligned} & f \text{ est } \mu\text{-intégrable} \\ \Leftrightarrow & \int f^+ d\mu < +\infty \quad \text{et} \quad \int f^- d\mu < +\infty \\ \Leftrightarrow & \int |f| d\mu < +\infty. \quad \square \end{aligned}$$

Exemples.

a) **Intégrale par rapport à une mesure de Dirac**

Rappelons que la mesure de Dirac définie au point ω_0 est la probabilité discrète δ_{ω_0} telle que, pour tout A dans \mathcal{A} , on ait :

$$\delta_{\omega_0}(A) = \mathbb{1}_A(\omega_0)$$

Pour tout événement A dans la tribu \mathcal{A} , la fonction $\mathbb{1}_A$ étant trivialement dans \mathcal{E}^+ , on a :

$$\int \mathbb{1}_A d\delta_{\omega_0} = \delta_{\omega_0}(A) = \mathbb{1}_A(\omega_0).$$

Ainsi, pour toute fonction étagée f dans \mathcal{E}^+ d'expression

$$f = \sum_{i=1}^n y_i \mathbb{1}_{A_i},$$

on a :

$$\int f d\delta_{\omega_0} = \sum_{i=1}^n y_i \delta_{\omega_0}(A_i) = \sum_{i=1}^n y_i \mathbb{1}_{A_i}(\omega_0) = f(\omega_0).$$

Supposons maintenant que la fonction f ne soit plus étagée mais seulement positive et mesurable. On a vu qu'il existe une suite croissante (f_n) dans \mathcal{E}^+ convergeant vers f . On a alors, par définition,

$$\int f d\delta_{\omega_0} = \lim_{n \rightarrow +\infty} \int f_n d\delta_{\omega_0} = \lim_{n \rightarrow +\infty} f_n(\omega_0) = f(\omega_0).$$

Supposons enfin que f soit une fonction réelle mesurable (de signe quelconque) telle que $|f(\omega_0)| < +\infty$. On a alors

$$\int f^+ d\delta_{\omega_0} = f^+(\omega_0) < +\infty$$

et

$$\int f^- d\delta_{\omega_0} = f^-(\omega_0) < +\infty.$$

La fonction f est donc intégrable par rapport à δ_{ω_0} et d'intégrale

$$\begin{aligned} \int f d\delta_{\omega_0} &= \int f^+ d\delta_{\omega_0} - \int f^- d\delta_{\omega_0} \\ &= f^+(\omega_0) - f^-(\omega_0) = f(\omega_0). \end{aligned}$$

Ainsi, pour toute fonction mesurable f telle que $|f(\omega_0)| < +\infty$, l'intégrale de f par rapport à la mesure de Dirac en ω_0 est égale à la valeur prise par la fonction f en ω_0 .

b) Intégrale par rapport à une mesure discrète

C'est une généralisation directe de l'intégrale par rapport à la mesure de Dirac.

Soit μ une mesure discrète (éventuellement de masse totale supérieure à 1), d'ensemble des atomes $D = \{\omega_n \text{ où } n \in I\}$, où I est un ensemble fini ou dénombrable d'indices. Notons $p_n = \mu(\{\omega_n\})$. On a vu dans le chapitre 1, que l'on peut écrire :

$$\mu = \sum_{n \in I} p_n \delta_{\omega_n}.$$

Pour tout événement A dans la tribu \mathcal{A} , on a :

$$\int \mathbb{1}_A d\mu = \mu(A) = \sum_{n \in I} p_n \delta_{\omega_n}(A) = \sum_{n \in I} p_n \mathbb{1}_A(\omega_n).$$

Pour une fonction étagée $f \in \mathcal{E}^+$ d'expression

$$f = \sum_{i=1}^k y_i \mathbb{1}_{A_i},$$

on a alors

$$\begin{aligned} \int f d\mu &= \sum_{i=1}^k y_i \mu(A_i) = \sum_{i=1}^k y_i \sum_{n \in I} p_n \delta_{\omega_n}(A_i) \\ &= \sum_{i=1}^k y_i \sum_{n \in I} p_n \mathbb{1}_{A_i}(\omega_n) \\ &= \sum_{n \in I} p_n \sum_{i=1}^k y_i \mathbb{1}_{A_i}(\omega_n) = \sum_{n \in I} p_n f(\omega_n). \end{aligned}$$

On peut montrer, par le même raisonnement que celui fait précédemment pour la mesure de Dirac, que pour toute fonction mesurable f telle que

$$\sum_{n \in I} p_n |f(\omega_n)| < +\infty,$$

on a :

$$\int f d\mu = \sum_{n \in I} p_n f(\omega_n).$$

Appliquons ceci à la mesure de dénombrement. Comme on l'a déjà vu, la mesure de dénombrement ou de comptage est un cas particulier de la mesure discrète précédente, où $D = \mathbb{N}$ et $p_n = 1$, pour tout n dans \mathbb{N} . On écrit donc

$$\mu = \sum_{n \in \mathbb{N}} \delta_n.$$

D'après ce que l'on vient de voir, toute fonction réelle mesurable f telle que

$$\sum_n |f(n)| < +\infty,$$

est μ -intégrable et

$$\int f d\mu = \sum_{n \in \mathbb{N}} f(n).$$

c) **Intégrale par rapport à la mesure de Lebesgue sur \mathbb{R}**

On a vu que, pour tout intervalle $]a, b]$, la mesure de Lebesgue de cet intervalle est $\lambda([a, b]) = b - a$. Ainsi

$$\int \mathbb{I}_{]a, b]} d\lambda = \lambda([a, b]) = b - a = \int_a^b dx.$$

L'intégrale de $\mathbb{I}_{]a, b]}$ par rapport à la mesure de Lebesgue est alors, dans ce cas, égale à son intégrale au sens de Riemann. On peut montrer que dans la majorité des cas l'intégrale de Lebesgue (intégrale par rapport à la mesure de Lebesgue) et de Riemann sont confondues.

En effet si f est intégrable au sens de Riemann sur un intervalle donné $[a, b]$, alors la fonction f est intégrable au sens de Lebesgue et les deux intégrales sont égales i.e.

$$\int_{[a, b]} f d\lambda = \int_a^b f(x) dx.$$

De même sur l'intervalle $]a, b[$ où a et b peuvent prendre respectivement les valeurs $-\infty$ et $+\infty$, si $|f|$ est intégrable au sens de Riemann, alors f est intégrable au sens de Lebesgue et les deux intégrales sont confondues :

$$\int_{]a, b[} f d\lambda = \int_a^b f(x) dx.$$

Donnons maintenant quelques propriétés de l'intégrale ainsi définie.

Proposition 2.3.8 1) L'ensemble des fonctions μ -intégrables forme un \mathbb{R} -espace vectoriel et l'intégrale par rapport à μ est une forme linéaire i.e. pour toutes fonctions f et g μ -intégrables et pour tout couple (λ, γ) dans \mathbb{R}^2 , la fonction $\lambda f + \gamma g$ est μ -intégrable et on a l'égalité

$$\int (\lambda f + \gamma g) d\mu = \lambda \int f d\mu + \gamma \int g d\mu.$$

De même pour tous événements A et B disjoints dans \mathcal{A} , on a :

$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$

2) Monotonie

Si f et g sont deux fonctions telles que $f \geq g$, on a alors l'inégalité :

$$\int f d\mu \geq \int g d\mu.$$

En particulier si f est positive alors l'intégrale $\int f d\mu$ l'est aussi.

3) Si f et g sont deux fonctions telles que $|f| \leq g$ (donc g est positive) et g est μ -intégrable alors f est μ -intégrable.

4) Si la fonction f est μ -intégrable, on a l'inégalité :

$$\left| \int f d\mu \right| \leq \int |f| d\mu.$$

Preuve. La plupart de ces résultats s'obtiennent facilement. Démontrons la dernière inégalité. On a

$$\left| \int f d\mu \right| = \left| \int f^+ d\mu - \int f^- d\mu \right| \leq \int f^+ d\mu + \int f^- d\mu = \int |f| d\mu. \quad \square$$

Introduisons maintenant la notion de négligeabilité.

Définition 2.3.9 Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et \mathcal{P} une propriété définie sur Ω . On dira que \mathcal{P} est vérifiée μ -presque-partout (μ -p.p.) s'il existe un ensemble N de \mathcal{A} tel que $\mu(N) = 0$ et que pour tout ω dans \bar{N} la propriété \mathcal{P} soit vérifiée.

Exemples.

- * $f = 0$ μ -p.p. signifie qu'il existe N dans \mathcal{A} avec $\mu(N) = 0$ tel que pour tout ω dans \bar{N} on ait : $f(\omega) = 0$.
- * f est finie μ -p.p. si il existe N dans \mathcal{A} avec $\mu(N) = 0$ tel que pour tout ω dans \bar{N} on ait : $f(\omega) < +\infty$. \diamond

Proposition 2.3.10 1) Si f est μ -intégrable alors f est finie μ -p.p.
 2) Pour une fonction f mesurable on a l'équivalence

$$f = 0 \quad \mu\text{-p.p.} \quad \Leftrightarrow \quad \int |f| d\mu = 0.$$

Chacune de ces deux assertions implique que

$$\int f d\mu = 0.$$

3) Soient f et g deux fonctions mesurables telles que f soit μ -intégrable et $f = g$ μ -p.p. Alors g est μ -intégrable et

$$\int f d\mu = \int g d\mu.$$

4) Si f est une fonction réelle mesurable, alors pour tout événement A dans \mathcal{A} tel que $\mu(A) = 0$, on a

$$\int_A f d\mu = 0.$$

Donnons enfin, sans leurs démonstrations, deux théorèmes fondamentaux de la théorie de l'intégration.

Théorème 2.3.11 (théorème de la convergence monotone de Beppo-Lévi).
 Soit (f_n) une suite croissante de fonctions mesurables positives. Alors

$$\int \lim_{n \rightarrow +\infty} f_n d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu.$$

Théorème 2.3.12 (théorème de la convergence dominée ou de Lebesgue).

Soit (f_n) une suite de fonctions réelles, mesurables et μ -intégrables. Si la suite (f_n) converge vers f , μ -p.p., et s'il existe une fonction g mesurable positive telle que, pour tout n , on ait :

$$|f_n| \leq g,$$

alors f est μ -intégrable et

$$\lim_{n \rightarrow +\infty} \int f_n d\mu = \int f d\mu.$$

2.3.2 Absolue continuité d'une mesure par rapport à une autre. Densité

Soit (Ω, \mathcal{A}) un espace mesurable et μ et ν deux mesures positives sur cet espace.

Définition 2.3.13

- On dit que ν est absolument continue par rapport à μ si, pour tout A dans \mathcal{A} tel que $\mu(A) = 0$, on a $\nu(A) = 0$. On note $\nu \ll \mu$.
- On dit que ν admet une densité par rapport à μ s'il existe une fonction f mesurable positive telle que :

$$\forall A \in \mathcal{A} : \quad \nu(A) = \int_A f d\mu.$$

Ces notions généralisent bien les résultats vus au début de ce chapitre. En effet, on a dit que la loi d'une variable aléatoire X est absolument continue s'il existe f mesurable positive telle que

$$\forall A \in \mathcal{A}, \quad P_X(A) = \int_A f(x) dx.$$

Or, on a vu que dans les “bons cas”

$$\int_A f(x) dx = \int_A f d\lambda$$

où λ est la mesure de Lebesgue. Ainsi, dire que P_X est absolument continue veut en fait dire qu'elle admet une densité par rapport à Lebesgue. Cela veut également dire que P_X est absolument continue par rapport à λ . En effet, si A est un événement tel que $\lambda(A) = 0$, on a vu précédemment que cela implique

$$\int_A f d\lambda = 0$$

et on a donc bien que : $P_X(A) = 0$.

On pourrait montrer que ces deux définitions “absolue continuité” et “existence d'une densité” sont équivalentes pour les mesures que nous avons considérées (i.e. mesures positives qui sont par définition σ -additives).

Théorème 2.3.14 Une mesure ν est absolument continue par rapport à une autre mesure μ si, et seulement si, ν admet une densité par rapport à μ .

Intéressons nous maintenant à l'existence d'une densité par rapport à une mesure discrète.

Remarquons par exemple que toute probabilité discrète (ce serait également vrai pour une mesure) sur une partie de \mathbb{N} (i.e. sur $D \subset \mathbb{N}$) est absolument continue par rapport à la mesure de dénombrement. En effet, on peut écrire

$$P = \sum_{n \in D} p_n \delta_n$$

où les p_n sont positifs et tels que $\sum_{n \in D} p_n = 1$. Si μ est la mesure de dénombrement, i.e.

$$\mu = \sum_{n \in \mathbb{N}} \delta_n$$

et soit un événement A dans \mathcal{A} tel que $\mu(A) = 0$, alors par définition de μ , on a

$$\sum_{n \in \mathbb{N}} \delta_n(A) = 0.$$

D'où, pour tout n dans \mathbb{N} , on a : $\delta_n(A) = 0$. Il vient donc

$$P(A) = \sum_{n \in D} p_n \delta_n(A) = 0$$

et P est bien absolument continue par rapport à μ .

Par ailleurs, soit une fonction f de \mathbb{R} vers \mathbb{R} définie par

$$\forall n \in D : f(n) = p_n \quad \text{et} \quad \forall x \in \mathbb{R} \setminus D : f(x) = 0.$$

On a, pour tout événement A dans \mathcal{A} ,

$$\begin{aligned} P(A) = \sum_{n \in D} p_n \delta_n(A) &= \sum_{n \in \mathbb{N}} f(n) \mathbb{1}_A(n) \\ &= \int f \mathbb{1}_A d\mu = \int_A f d\mu \end{aligned}$$

Donc la densité de P par rapport à μ est la fonction f nulle sur tout \mathbb{R} , sauf sur D où sa valeur pour tout n est p_n .

Exemple :

On a vu que la loi binomiale $B(n, p)$ d'une v.a.r. X peut s'écrire :

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

Ainsi, P_X est absolument continue par rapport à la mesure de dénombrement $\mu = \sum_n \delta_n$ ou $\mu = \sum_{k=0}^n \delta_k$ et est de densité :

$$\begin{aligned} f(k) &= C_n^k p^k (1-p)^{n-k}, \text{ pour } k = 0, \dots, n \\ \text{et } f(x) &= 0, \forall x \neq 0, \dots, n. \end{aligned}$$

Notons que l'on a alors :

$$P_X(A) = \int_A f d\mu = \sum_{k \in A} f(k) = \sum_{k \in A} C_n^k p^k (1-p)^{n-k}$$

2.3.3 Mélange de loix

Soient π_1 et π_2 deux probabilités sur \mathbb{R} . Supposons que π_1 soit absolument continue par rapport à la mesure de Lebesgue (de densité f), i.e. absolument continue au sens vu au début du chapitre, et supposons que π_2 soit discrète sur D partie de \mathbb{N} avec, pour tout n dans D

$$\pi_2(\{n\}) = p_n,$$

i.e. absolument continue par rapport à la mesure de dénombrement.

Alors, pour tout α dans $]0, 1[$, la mesure

$$P = \alpha \pi_1 + (1 - \alpha) \pi_2$$

est encore une probabilité et, pour tout A dans $\mathcal{B}_{\mathbb{R}}$:

$$\begin{aligned} P(A) &= \int \mathbb{1}_A dP = \alpha \int \mathbb{1}_A d\pi_1 + (1 - \alpha) \int \mathbb{1}_A d\pi_2 \\ &= \alpha \int_A f(x) dx + (1 - \alpha) \sum_{n \in A \cap D} p_n. \end{aligned}$$

Exemple. Mélange d'une Dirac en 0 et d'une exponentielle.

Considérons un matériel qui, quand il est mis en marche, a une probabilité $1 - \alpha$ de tomber en panne dès l'instant initial et qui, ensuite, est tel que le temps d'attente avant l'arrivée d'une panne suit une loi exponentielle de paramètre λ .

Notons π_1 la loi absolument continue de densité

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{]0, +\infty[}(x)$$

et $\pi_2 = \delta_0$ (loi absolument continue par rapport à la mesure de dénombrement).

Ainsi pour tout x strictement positif

$$P([-\infty, x]) = \alpha \int_{-\infty}^x f(u) du + (1 - \alpha)$$

et $P(\{0\}) = 1 - \alpha =$ probabilité de tomber en panne dès l'instant initial. \diamond

2.3.4 Densités conjointes, marginales et conditionnelles

Les résultats vus précédemment sur les densités conjointes, marginales et conditionnelles restent vrais si les variables aléatoires ne sont pas absolument continues.

Considérons deux v.a. X et Y définies sur un même espace (Ω, \mathcal{A}, P) et à valeurs respectivement dans $(\Omega_1, \mathcal{A}_1)$ et $(\Omega_2, \mathcal{A}_2)$. Le couple (X, Y) est alors défini sur (Ω, \mathcal{A}, P) et est à valeurs dans $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$.

On dira que (X, Y) admet une densité par rapport à une mesure quelconque $\mu_1 \otimes \mu_2$ si, et seulement si, sa loi $P_{X,Y}$ admet une densité par rapport à $\mu_1 \otimes \mu_2$, i.e. s'il existe une fonction $f_{X,Y}$ mesurable, positive sur $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, telle que, pour tout A_1 et A_2 respectivement dans \mathcal{A}_1 et \mathcal{A}_2 , on ait :

$$P_{(X,Y)}(A_1 \times A_2) = \int_{A_1 \times A_2} f_{X,Y}(x, y) d\mu_1(x) d\mu_2(y).$$

Alors P_X admet pour densité f_X par rapport à μ_1 , donnée par :

$$f_X(x) = \int_{\Omega_2} f_{X,Y}(x, y) d\mu_2(y)$$

et la loi conditionnelle à $X = x$ de Y a pour densité par rapport à μ_2

$$f_Y^{X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

C'est donc en particulier le cas pour $\Omega_1 = \Omega_2 = \mathbb{R}$ et par exemple :

$$\begin{cases} \mu_1 = \text{mesure de Lebesgue} \\ \mu_2 = \text{mesure de dénombrement.} \end{cases}$$

Chapitre 3

Moments de variables aléatoires

3.1 Variables aléatoires réelles intégrables et espérance mathématique

Les résultats sur l'intégration de fonctions mesurables par rapport à une mesure positive, et vus dans le chapitre précédent, restent bien sûr vrais pour une v.a.r. qui est, par définition, une variable aléatoire de (Ω, \mathcal{A}, P) vers $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

On peut donc considérer l'intégrale de X par rapport à sa mesure de probabilité P . Cette intégrale, quand elle existe, est appelée espérance de X .

Définition 3.1.1 *Soit X une v.a.r. positive ou P -intégrable, i.e. telle que*

$$\int |X| dP < +\infty.$$

L'intégrale

$$\int X dP = \int_{\Omega} X(\omega) dP(\omega)$$

est appelée espérance mathématique de X et est notée $\mathbb{E}X$.

Toutes les propriétés de linéarité, de monotonie, de convergence pour les intégrales restent donc vraies pour l'espérance mathématique.

De même, la notion de négligeabilité est conservée mais, dans le langage des probabilités, on ne dit plus “ μ -p.p.” mais “ P -presque sûrement” ou simplement “presque sûrement”, noté *p.s.*, quand il n'y a pas de risque de confusion.

On a vu précédemment que bien souvent on ignore ce qu'est l'espace (Ω, \mathcal{A}, P) et on ne connaît bien que l'espace probabilisé $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P_X)$. C'est pourquoi, en particulier dans le calcul de l'espérance mathématique d'une v.a.r., on utilise le théorème suivant qui permet de transformer le calcul de l'intégrale sur (Ω, \mathcal{A}, P) en un calcul d'une intégrale sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P_X)$.

Théorème 3.1.2 (Théorème du transport)

Soit X une v.a.r. de (Ω, \mathcal{A}, P) vers $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ de loi P_X et h une fonction mesurable de $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ vers lui même, positive ou P_X -intégrable. On a alors :

$$\int_{\Omega} h(X) dP = \int_{\Omega} h \circ X(\omega) dP(\omega) = \int_{\mathbb{R}} h(x) dP_X(x).$$

On note ce réel $\mathbb{E}(h(X))$.

Ce théorème revient en quelque sorte à faire un changement de variable $x = X(\omega)$ dans l'intégrale.

Ainsi, si la fonction $h(x) = x$ est P_X -intégrable, i.e. si

$$\int_{\mathbb{R}} |x| dP_X(x) < +\infty,$$

alors on peut parler de l'espérance de X et elle est égale à

$$\mathbb{E}X = \int_{\mathbb{R}} x dP_X(x).$$

Si la loi P_X de X est absolument continue par rapport à une mesure μ et de densité f par rapport à cette mesure, alors d'après ce que l'on a vu à la fin du chapitre précédent, on peut écrire :

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

Deux cas particuliers sont fondamentaux.

a) Si la variable aléatoire X est absolument continue et intégrable (i.e. si sa loi est absolument continue par rapport à Lebesgue) on a :

$$\mathbb{E}X = \int_{\mathbb{R}} x f(x) d\lambda(x) = \int_{-\infty}^{+\infty} x f(x) dx.$$

b) Si la variable aléatoire X est discrète à valeurs dans

$$D = \{x_1, \dots, x_n, \dots\},$$

sa loi est absolument continue par rapport à la mesure de dénombrement

$$\mu = \sum_{n=1}^{+\infty} \delta_{x_n}.$$

Sous l'hypothèse que l'on ait

$$\sum_{n=1}^{+\infty} |x_n| P(X = x_n) < +\infty$$

(i.e. X est intégrable par rapport à μ), on a :

$$\begin{aligned} \mathbb{E}X &= \int_{\mathbb{R}} x f(x) d\mu(x) = \sum_{n=1}^{+\infty} x_n f(x_n) \\ &= \sum_{n=1}^{+\infty} x_n P(X = x_n). \end{aligned}$$

Exemples.

* Calculons l'espérance d'une v.a.r. absolument continue de loi exponentielle. On a :

$$\begin{aligned}\mathbb{E}X &= \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbb{R}} x dP_X(x) = \int_0^{+\infty} \lambda x e^{-\lambda x} dx \\ &= \int_0^{+\infty} u e^{-u} \frac{du}{\lambda} = \frac{\Gamma(2)}{\lambda} = \frac{1}{\lambda}\end{aligned}$$

* Calculons l'espérance d'une v.a.r. X de loi de Bernoulli $B(p)$. On a vu que, dans ce cas, $D = \{0, 1\}$ et la loi de Bernoulli est donc dominée par la mesure de comptage ou de dénombrement sur \mathbb{N} , mais aussi, plus simplement, par la mesure $\delta_0 + \delta_1$. On a

$$\begin{aligned}\mathbb{E}X &= \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbb{R}} x dP_X(x) = \int_{\mathbb{R}} x f(x) d(\delta_0 + \delta_1)(x) \\ &= 0 \cdot f(0) + 1 \cdot f(1) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = p. \quad \diamond\end{aligned}$$

Notons que le théorème du transport n'est pas seulement utile pour nous aider à calculer l'espérance d'une variable aléatoire X dont on connaît la loi. Ainsi, si φ est une fonction mesurable et X une variable aléatoire de loi P_X , on peut alors calculer l'espérance de $Y = \varphi(X)$, si celle-ci existe, et cela sans déterminer la loi de Y . En effet, si φ est P_X -intégrable, on a :

$$\mathbb{E}(Y) = \mathbb{E}(\varphi(X)) = \int \varphi(x) dP_X(x).$$

Proposition 3.1.3 (Inégalité de Jensen)

Soit φ une fonction convexe de \mathbb{R} vers lui même et X une v.a.r. telles que X et $\varphi(X)$ soient intégrables. On a alors :

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

Rappel. Une fonction f de \mathbb{R} vers lui même est dite convexe si, pour tout couple (x, y) de \mathbb{R}^2 et pour tout λ de $[0, 1]$, on a :

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Notons, en particulier, qu'une fonction deux fois dérivable dont la dérivée seconde est positive est une fonction convexe.

Exemple. On peut vérifier que la fonction valeur absolue est une fonction convexe. D'où

$$|\mathbb{E}X| \leq \mathbb{E}|X|$$

et on retrouve bien le résultat classique. \diamond

Définition 3.1.4 Soit (Ω, \mathcal{A}, P) un espace probabilisé. On note $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ (ou simplement \mathcal{L}^1 quand il n'y a pas de risque de confusion), l'espace des v.a.r. intégrables définies sur (Ω, \mathcal{A}) .

En utilisant les résultats du chapitre 2, on peut montrer que la notion d'égalité P -presque sûre est une relation d'équivalence sur l'espace \mathcal{L}^1 . Ainsi on peut raisonner uniquement sur les classes d'équivalences.

Définition 3.1.5 On appelle $L^1(\Omega, \mathcal{A}, P)$ l'ensemble des classes d'équivalences sur \mathcal{L}^1 .

Par abus de langage, on confondra dans la suite la variable aléatoire et sa classe d'équivalence.

Théorème 3.1.6 L'espace $L^1(\Omega, \mathcal{A}, P)$ est un espace vectoriel sur \mathbb{R} , normé par

$$\|X\|_{L^1} = \mathbb{E}(|X|) = \int |X| dP.$$

Preuve. On a déjà montré que l'espace $L^1(\Omega, \mathcal{A}, P)$ est un \mathbb{R} -espace vectoriel. D'autre part, on a :

$$\begin{aligned} \|\lambda X\|_{L^1} &= \mathbb{E}(|\lambda X|) = |\lambda| \mathbb{E}(|X|) = |\lambda| \|X\|_{L^1} \\ \|X + Y\|_{L^1} &= \mathbb{E}(|X + Y|) \leq \mathbb{E}(|X| + |Y|) \leq \|X\|_{L^1} + \|Y\|_{L^1} \\ \text{et} \quad \|X\|_{L^1} &= 0 \Leftrightarrow \mathbb{E}|X| = 0 \Leftrightarrow X = 0 \text{ p.s.} \Leftrightarrow X = 0 \text{ dans } L^1. \quad \square \end{aligned}$$

3.2 Moments de variables aléatoires réelles

3.2.1 Espace L^p

Définition 3.2.1 Soit p un réel tel que $p \geq 1$. On dit qu'une variable aléatoire X appartient à l'espace $L^p(\Omega, \mathcal{A}, P)$ si elle est de puissance p -intégrable, i.e. si

$$\mathbb{E}|X|^p < +\infty.$$

La norme sur $L^p(\Omega, \mathcal{A}, P)$ est :

$$\|X\|_p = (\mathbb{E}|X|^p)^{1/p} = \left(\int_{\Omega} |X|^p dP \right)^{1/p}.$$

Définition 3.2.2 Pour une variable aléatoire X dans $L^p(\Omega, \mathcal{A}, P)$, on appelle

- moment d'ordre p le réel : $\mathbb{E}X^p$.
- moment absolu d'ordre p le réel : $\mathbb{E}|X|^p$.
- moment centré d'ordre p le réel : $\mathbb{E}((X - \mathbb{E}X)^p)$.

Proposition 3.2.3 Soit p et q deux réels tels que : $1 \leq p < q$. On a $\|X\|_p \leq \|X\|_q$ et donc $L^q \subset L^p$.

Preuve. Elle repose sur l'utilisation de l'inégalité de Jensen avec la fonction $\varphi(x) = x^{q/p}$, qui est bien une fonction convexe sur \mathbb{R}^+ , et avec la v.a.r. $Y = |X|^p$.

D'après l'inégalité de Jensen on peut écrire :

$$\varphi(\mathbb{E}Y) \leq \mathbb{E}\varphi(Y),$$

d'où, on tire :

$$\begin{aligned} (\mathbb{E}|X|^p)^{q/p} &\leq \mathbb{E}|X|^q \\ \Leftrightarrow \|X\|_p &\leq \|X\|_q. \end{aligned}$$

Ainsi, on a les relations suivantes :

$$X \in L^q \Leftrightarrow \|X\|_q < +\infty \Rightarrow \|X\|_p < +\infty \Rightarrow X \in L^p. \quad \square$$

On admet les deux théorèmes suivants :

Théorème 3.2.4 (Inégalité de Hölder)

Pour tous réels p et q tels que $p > 1$, $q > 1$ et $1/p + 1/q = 1$ et toutes v.a. X et Y respectivement dans $L^p(\Omega, \mathcal{A}, P)$ et $L^q(\Omega, \mathcal{A}, P)$, on a :

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

Théorème 3.2.5 (Inégalité de Minkowski)

Soit un réel $p \geq 1$. Si X et Y sont deux éléments de $L^p(\Omega, \mathcal{A}, P)$ alors

$$X + Y \in L^p(\Omega, \mathcal{A}, P) \quad \text{et} \quad \|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Ce dernier théorème nous permet de montrer que, pour tout réel $p \geq 1$, l'espace $L^p(\Omega, \mathcal{A}, P)$ est un espace vectoriel normé. On peut en fait montrer qu'ils est également complet et donc un espace de Banach.

Étudions plus en détail le cas, particulièrement intéressant, de l'espace L^2 .

3.2.2 Espace L^2

L'inégalité de Hölder, vue précédemment, appliquée pour $p = q = 2$ est souvent appelée inégalité de Schwartz.

Proposition 3.2.6 Soit X et Y deux v.a.r. dans L^2 . On a :

$$\|XY\|_1 \leq \|X\|_2 \|Y\|_2.$$

Ainsi, si les v.a.r. X et Y sont dans L^2 , la v.a.r. XY est dans L^1 . En revanche, il n'est pas suffisant que X et Y soient dans L^1 pour que XY le soit également.

Proposition 3.2.7 L'application de $L^2 \times L^2$ vers \mathbb{R} définie par :

$$(X, Y) \mapsto \langle X, Y \rangle = \mathbb{E}XY = \int XY dP$$

est un produit scalaire.

Preuve. Il est facile de montrer qu'il s'agit d'une forme bilinéaire symétrique et strictement positive. \square

Théorème 3.2.8 L'espace L^2 est un espace de Hilbert, i.e. un espace préhilbertien (espace vectoriel muni d'un produit scalaire) complet pour la norme associée au produit scalaire.

La norme dans L^2 est :

$$\|X\|_2 = \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{E}X^2}.$$

Définition 3.2.9 On appelle variance d'une v.a.r. X dans L^2 , son moment centré d'ordre 2, i.e. :

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}X)^2) = \langle X - \mathbb{E}X, X - \mathbb{E}X \rangle.$$

L'écart-type de X est la racine carrée de sa variance

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

Proposition 3.2.10 Pour toute v.a.r. X dans L^2 on a les propriétés suivantes :

- i) $\text{Var}(X) = \mathbb{E}X^2 - \mathbb{E}^2X$;
- ii) $X = a$ p.s. $\Leftrightarrow \text{Var}(X) = 0$;
- iii) pour tout couple (a, b) dans \mathbb{R}^2 , on a : $\text{Var}(aX + b) = a^2 \text{Var}(X)$;
- iv)

$$\text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \inf_{a \in \mathbb{R}} \mathbb{E}(X - a)^2.$$

Remarque. La propriété iv) montre que l'espérance est la projection orthogonale au sens de la norme L^2 sur les sous-espaces des variables aléatoires constantes.

Preuve. Soit X une v.a.r. quelconque dans L^2 .

i) On a :

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2 - 2X\mathbb{E}X + \mathbb{E}^2X) \\ &= \mathbb{E}X^2 - 2\mathbb{E}X\mathbb{E}X + \mathbb{E}^2X = \mathbb{E}X^2 - \mathbb{E}^2X. \end{aligned}$$

ii) Supposons que X soit telle que $\text{Var}(X) = 0$. On a alors les équivalences suivantes :

$$\text{Var}(X) = 0 \Leftrightarrow \mathbb{E}(X - \mathbb{E}X)^2 = 0 \Leftrightarrow X - \mathbb{E}X = 0 \text{ p.s. } \Leftrightarrow X = \mathbb{E}X \text{ p.s.}$$

iii) Soit (a, b) un couple dans \mathbb{R}^2 . On a :

$$\begin{aligned} \text{Var}(aX + b) &= \mathbb{E}(aX + b - \mathbb{E}(aX + b))^2 \\ &= \mathbb{E}(aX + b - a\mathbb{E}X - b)^2 \\ &= \mathbb{E}(a^2(X - \mathbb{E}X)^2) = a^2\mathbb{E}(X - \mathbb{E}X)^2 = a^2\text{Var}(X). \end{aligned}$$

iv) Soit a un réel quelconque. On peut écrire :

$$\begin{aligned}\mathbb{E}(X - a)^2 &= \mathbb{E}(X - \mathbb{E}X + \mathbb{E}X - a)^2 \\ &= \mathbb{E}((X - \mathbb{E}X)^2 + 2(X - \mathbb{E}X)(\mathbb{E}X - a) + (\mathbb{E}X - a)^2) \\ &= \mathbb{E}(X - \mathbb{E}X)^2 + 2(\mathbb{E}X - a)\mathbb{E}(X - \mathbb{E}X) + (\mathbb{E}X - a)^2 \\ &= \mathbb{E}(X - \mathbb{E}X)^2 + (\mathbb{E}X - a)^2.\end{aligned}$$

D'où, pour tout a dans \mathbb{R} , on a :

$$\mathbb{E}(X - \mathbb{E}X)^2 \leq \mathbb{E}(X - a)^2$$

et comme $\mathbb{E}X$ est une constante, on a bien le résultat. \square

Terminologie particulière. Soit X une v.a.r. dans L^2 . Les v.a.r.

$$X - \mathbb{E}X \quad \text{et} \quad \frac{X - \mathbb{E}X}{\sigma_X}$$

sont respectivement appelées v.a.r. centrée et v.a.r. centrée réduite associées à X .

Proposition 3.2.11 (*Inégalités de Markov et de Bienaymé-Tchebychev*)

i) *Markov.* Soit X une v.a.r. intégrable (i.e. dans $L^1(\Omega, \mathcal{A}, P)$). On a :

$$P(|X| \geq c) \leq \frac{\mathbb{E}|X|}{c},$$

pour tout réel c strictement positif.

ii) *Bienaymé-Tchebychev.* Soit X une v.a.r. dans $L^2(\Omega, \mathcal{A}, P)$. On a :

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

Preuve.

i) Pour une v.a.r. X quelconque dans L^1 , on a :

$$\begin{aligned}\mathbb{E}|X| &= \int_{\Omega} |X| dP = \int_{|X| < c} |X| dP + \int_{|X| \geq c} |X| dP \\ &\geq \int_{|X| \geq c} |X| dP \geq c \int_{|X| \geq c} dP = cP(|X| \geq c).\end{aligned}$$

ii) De même, en utilisant le i) et sous l'hypothèse supplémentaire que X soit de carré intégrable, on a :

$$P(|X - \mathbb{E}X| \geq c) = P((X - \mathbb{E}X)^2 \geq c^2) \leq \frac{1}{c^2} \mathbb{E}(X - \mathbb{E}X)^2 = \frac{\text{Var}(X)}{c^2}. \square$$

Avant d'aborder le cas des vecteurs aléatoires, revenons sur les conditions suffisantes pour que le produit XY de deux v.a.r. X et Y appartienne à L^1 .

On a vu que si X et Y sont dans L^2 alors la v.a.r. XY est dans L^1 , d'après Minkowski (ou Schwartz). On a également déjà signalé qu'il n'est pas suffisant que X et Y soient dans L^1 pour que XY le soit également. En revanche le résultat est vrai si de plus X et Y sont indépendantes.

Proposition 3.2.12 *Si X et Y sont deux v.a.r. intégrables et indépendantes, alors*

i) *la v.a.r. XY est intégrable (i.e. $XY \in L^1$).*

ii) $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$.

Preuve. Démontrons ce résultat uniquement dans le cas de deux v.a.r. absolument continues. On a :

$$\begin{aligned} \mathbb{E}|XY| &= \int_{\mathbb{R}^2} |xy| f_{X,Y}(x,y) dx dy = \int_{\mathbb{R}^2} |x| |y| f_X(x) f_Y(y) dx dy \\ &= \left(\int_{\mathbb{R}} |x| f_X(x) dx \right) \left(\int_{\mathbb{R}} |y| f_Y(y) dy \right) = \mathbb{E}|X| \mathbb{E}|Y|, \end{aligned}$$

qui est fini par hypothèse. Ainsi, on a bien le i). En supprimant les valeurs absolues on obtient le ii). \square

3.3 Vecteurs aléatoires

Soient $X = (X_1, \dots, X_n)$ et $Y = (Y_1, \dots, Y_n)$ des vecteurs aléatoires à valeurs dans \mathbb{R}^n tels que, pour tout i , les v.a.r. X_i et Y_i soient intégrables.

3.3.1 Espérance mathématique

Définition 3.3.1 *On appelle espérance mathématique de X , le vecteur $\mathbb{E}X$ de \mathbb{R}^n de composantes $\mathbb{E}X_1, \dots, \mathbb{E}X_n$.*

On note souvent

$$\mathbb{E}X = \begin{pmatrix} \mathbb{E}X_1 \\ \vdots \\ \mathbb{E}X_n \end{pmatrix}$$

L'opérateur espérance est à nouveau linéaire, i.e. pour tout couple (α, β) dans \mathbb{R}^2 et tous vecteurs X et Y de même dimension et intégrables, on a :

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}X + \beta \mathbb{E}Y.$$

3.3.2 Covariance de deux v.a.r.

Définition 3.3.2 Soient X et Y deux v.a.r. intégrables telles que leur produit XY soit intégrable. On appelle covariance de X et Y le réel

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

Remarque. Si X et Y sont deux v.a.r. dans L^2 , la covariance de X et Y existe et

$$\text{Cov}(X, Y) = \langle X - \mathbb{E}X, Y - \mathbb{E}Y \rangle. \quad \diamond$$

Proposition 3.3.3 L'opérateur covariance vérifie les propriétés suivantes :

- i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- ii) $\text{Cov}(aX + X', Y) = a \text{Cov}(X, Y) + \text{Cov}(X', Y)$
- iii) $\text{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y.$

La covariance est donc un opérateur bilinéaire symétrique.

Preuve. Triviale. \square

Proposition 3.3.4 Soit $(X_i)_{i=1, \dots, n}$ une famille de v.a.r. dans L^2 . On a :

$$i) \text{Var}(X_1 + X_2) = \text{Var}(X_1) + 2 \text{Cov}(X_1, X_2) + \text{Var}(X_2).$$

ii)

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j).$$

Preuve. i) On peut écrire :

$$\begin{aligned}
 \text{Var}(X_1 + X_2) &= \mathbb{E} (X_1 + X_2 - (\mathbb{E}X_1 + \mathbb{E}X_2))^2 \\
 &= \mathbb{E} ((X_1 - \mathbb{E}X_1) + (X_2 - \mathbb{E}X_2))^2 \\
 &= \mathbb{E} ((X_1 - \mathbb{E}X_1)^2 + 2(X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2) + (X_2 - \mathbb{E}X_2)^2) \\
 &= \text{Var}(X_1) + 2 \text{Cov}(X_1, X_2) + \text{Var}(X_2).
 \end{aligned}$$

ii) Se montre aisément par récurrence. \square

Proposition 3.3.5 *Si X et Y sont deux v.a.r. intégrables et indépendantes, on a :*

$$\text{Cov}(X, Y) = 0.$$

Preuve. On a :

$$\text{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = \mathbb{E}X\mathbb{E}Y - \mathbb{E}X\mathbb{E}Y = 0. \quad \square$$

Nous attirons l'attention du lecteur sur le fait que la réciproque est fausse, comme le montre le contre-exemple suivant.

Contre-exemple. Soit X une v.a.r. discrète telle que :

$$P(X = 1) = P(X = -1) = \frac{1}{4} \quad \text{et} \quad P(X = 0) = \frac{1}{2}.$$

Soit Y la v.a.r. définie par $Y = X^2$. Notons que l'on a :

$$\mathbb{E}X = (-1) \times \frac{1}{4} + 0 \times \frac{1}{2} + 1 \times \frac{1}{4} = 0$$

et $XY = X^3 = X$. Ainsi, il vient :

$$\text{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = \mathbb{E}X - \mathbb{E}X\mathbb{E}Y = 0.$$

Mais X et Y ne sont pas indépendantes, puisque l'on a :

$$P(X = 1 \cap Y = 0) = 0 \neq P(X = 1)P(Y = 0) = \frac{1}{4} \times \frac{1}{2}. \quad \diamond$$

Corollaire 3.3.6 *Si $(X_i)_{i=1, \dots, n}$ est une famille de v.a.r. dans L^2 et indépendantes*

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i).$$

Preuve. Immédiate en utilisant la proposition de la variance d'une somme de variables aléatoires quelconques. \square

Définition 3.3.7 Si X et Y sont deux v.a.r. non constantes de L^2 , on appelle coefficient de corrélation linéaire de X et de Y , le réel :

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Proposition 3.3.8 Pour toutes v.a.r. X et Y dans L^2 non constantes, on a : $\rho_{X,Y} \in [-1, 1]$.

3.3.3 Matrice de covariance

Définition 3.3.9 Soit $X = (X_1, \dots, X_n)$ un vecteur de v.a.r. tel que chacune de ses composantes est dans L^2 (i.e. $X_i \in L^2$ pour tout $i = 1, \dots, n$). On appelle matrice de covariance de X , la matrice Γ_X carrée d'ordre n et de terme général $\text{Cov}(X_i, X_j)$.

La matrice Γ_X est forcément symétrique et ses termes diagonaux sont les variances des composantes des vecteurs.

Définissons maintenant l'espérance d'une matrice aléatoire. Soit $M = (Y_{ij})_{i,j}$ une matrice $n \times p$ aléatoire intégrable i.e. chaque terme Y_{ij} est une variable aléatoire réelle intégrable. On note $\mathbb{E}M$, la matrice $(\mathbb{E}(Y_{ij}))_{i,j}$.

Ainsi, si M^1 et M^2 sont des matrices de même dimension, on a :

$$\mathbb{E}(M^1 + M^2) = \mathbb{E}M^1 + \mathbb{E}M^2.$$

Si $A = (a_{ij})_{i,j}$ est une matrice $k \times n$ de réels, on a :

$$\mathbb{E}(AM) = A\mathbb{E}M,$$

où le produit est pris au sens des matrices. Si $B = (b_{ij})_{i,j}$ est une matrice $p \times q$ de réels, on a :

$$\mathbb{E}MB = \mathbb{E}M \cdot B.$$

Notons enfin que la matrice Γ_X peut s'écrire sous la forme :

$$\Gamma_X = \mathbb{E}((X - \mathbb{E}X)(X - \mathbb{E}X)^T),$$

le vecteur X étant noté sous forme d'un vecteur colonne.

Proposition 3.3.10 Soit $X = (X_1, \dots, X_n)$ un vecteur constitué de n variables aléatoires dans L^2 . Si A est une matrice $p \times n$ de réels, alors les composantes de $Y = AX$ sont des v.a.r. dans L^2 et on a :

$$\begin{aligned}\mathbb{E}Y &= A\mathbb{E}X \\ \Gamma_Y &= A\Gamma_X A^T.\end{aligned}$$

Preuve. Immédiate. □

3.3.4 Espérance conditionnelle

Soit (X, Y) un couple de v.a.r. On a vu dans le chapitre précédent comment déterminer les lois conditionnelles $P_Y^{X=x}$ et $P_X^{Y=y}$.

Définition 3.3.11 Si l'intégrale

$$\int_{\mathbb{R}} y dP_Y^{X=x}(y)$$

existe (i.e. si $h(y) = y$ est $P_Y^{X=x}$ -intégrable), on l'appelle espérance conditionnelle de Y sachant que $X = x$. On la note $\mathbb{E}(Y/X = x)$.

La fonction $e : x \rightarrow \mathbb{E}(Y/X = x)$ est une fonction réelle d'une variable réelle. On peut montrer qu'elle est mesurable et on peut considérer sa composition avec la variable aléatoire X , i.e. considérer $e \circ X$. Celle-ci définit une variable aléatoire réelle que l'on appelle espérance conditionnelle de Y sachant X , notée $\mathbb{E}(Y/X)$.

Exemples.

* Supposons que les lois de X et Y soient discrètes. Soit I et J les ensembles des atomes des lois P_X et P_Y . Pour tout x_i dans I , on a :

$$\begin{aligned}\mathbb{E}(Y/X = x_i) &= \sum_{y_j \in J} y_j P_Y^{X=x_i}(y_j) \\ &= \sum_{y_j \in J} y_j P(Y = y_j / X = x_i).\end{aligned}$$

* Lorsque la loi $P_Y^{X=x}$ est absolument continue par rapport à Lebesgue, i.e. conditionnellement à $X = x$ la v.a.r. Y est une v.a.r. absolument continue de densité $f_Y^{X=x}$, on a :

$$\mathbb{E}(Y/X = x) = \int_{\mathbb{R}} y f_Y^{X=x}(y) dy. \quad \diamond$$

Chapitre 4

Caractérisation des lois : transformée de Laplace et fonction caractéristique

L'objet de ce chapitre est de caractériser la loi d'une variable aléatoire à l'aide d'une fonction comme on l'a déjà fait avec la fonction de répartition ou les densités dans le cas des v.a.r. absolument continues. On cherche donc une fonction qui contienne toute l'information sur cette variable aléatoire, sur les moments, sur sa loi.

4.1 Transformée de Laplace

4.1.1 Variables aléatoires réelles

Définition 4.1.1 Soit X une v.a.r. définie sur un espace probabilisé (Ω, \mathcal{A}, P) . Soit I un intervalle contenant 0, tel que, pour tout réel s dans I , la v.a.r. e^{sX} soit intégrable. On appelle transformée de Laplace de la v.a.r. X la fonction définie sur I et à valeurs dans \mathbb{R}^+ par :

$$\mathcal{L}_X(s) = \mathbb{E}(e^{sX}).$$

La transformée de Laplace est aussi souvent appelée fonction génératrice des moments. On trouve également dans la littérature la transformée de Laplace définie par :

$$\mathcal{L}_X(s) = \mathbb{E}(e^{-sX}).$$

Énonçons les propriétés suivantes que nous ne démontrerons pas.

Proposition 4.1.2

1) La transformée de Laplace est toujours définie en 0 et

$$\mathcal{L}_X(0) = \mathbb{E}1 = 1.$$

Il existe des variables aléatoires, telles celles de loi de Cauchy, dont la transformée de Laplace n'est définie que pour $s = 0$.

2) Si X est bornée, alors \mathcal{L}_X est définie et continue sur tout \mathbb{R} .

3) Si X est positive, alors \mathcal{L}_X est continue et bornée sur $] -\infty, 0]$.

Exemples de transformée de Laplace.

* Supposons que la v.a.r. X suive une loi de Poisson $\mathcal{P}(\lambda)$. On a :

$$\begin{aligned} \mathcal{L}_X(s) &= \mathbb{E}(e^{sX}) = \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} e^{-\lambda} e^{sn} = e^{-\lambda} \sum_{n=0}^{+\infty} \frac{(\lambda e^s)^n}{n!} \\ &= e^{-\lambda} e^{\lambda e^s} = e^{\lambda(e^s - 1)} < +\infty, \end{aligned}$$

pour tout s dans \mathbb{R} . La transformée de Laplace d'une loi de Poisson est donc définie sur (tout) \mathbb{R} .

* Supposons que la v.a.r. X suive une loi binomiale $B(n, p)$. On a :

$$\mathcal{L}_X(s) = \mathbb{E}(e^{sX}) = \sum_{k=0}^n e^{sk} C_n^k p^k (1-p)^{n-k} = (pe^s + 1-p)^n < +\infty,$$

pour tout s dans \mathbb{R} .

* Supposons que la v.a.r. X suive une loi $\gamma(\alpha, \beta)$. On a :

$$\mathcal{L}_X(s) = \mathbb{E}(e^{sX}) = \int_{\mathbb{R}^+} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\beta-s)x} dx.$$

Cette intégrale est convergente si, et seulement si,

$$\beta - s > 0 \Leftrightarrow s < \beta.$$

On effectue le changement de variables $u = (\beta - s)x$ dans l'intégrale pour obtenir :

$$\begin{aligned} \mathcal{L}_X(s) &= \mathbb{E}(e^{sX}) = \int_{\mathbb{R}^+} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{u^{\alpha-1}}{(\beta-s)^{\alpha-1}} e^{-u} \frac{du}{\beta-s} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)(\beta-s)^\alpha} \Gamma(\alpha) = \frac{1}{\left(1 - \frac{s}{\beta}\right)^\alpha}, \end{aligned}$$

pour tout s dans l'intervalle $] -\infty, \beta[$. ◇

L'intérêt de la transformée de Laplace réside dans le théorème suivant.

Théorème 4.1.3 *La transformée de Laplace d'une variable aléatoire caractérise la loi de cette variable aléatoire. Autrement dit, si deux v.a.r. ont la même transformée de Laplace, alors elles ont la même loi.*

Ainsi déterminer la transformée de Laplace d'une variable aléatoire est un moyen supplémentaire de déterminer sa loi.

Théorème 4.1.4 *Soit (X, Y) un couple de variables aléatoires indépendantes dont chacune admet une transformée de Laplace sur, respectivement, les intervalles I et J . Alors la somme de ces deux v.a.r. admet une transformée de Laplace sur $I \cap J$ et on a :*

$$\mathcal{L}_{X+Y}(s) = \mathcal{L}_X(s) \mathcal{L}_Y(s).$$

Preuve. Par définition, on a :

$$\mathcal{L}_{X+Y}(s) = \mathbb{E}(e^{s(X+Y)}) = \mathbb{E}(e^{sX} e^{sY}) = \mathbb{E}(e^{sX}) \mathbb{E}(e^{sY}),$$

puisque l'indépendance des v.a.r. X et Y entraîne celle des v.a.r. e^{sX} et e^{sY} . Tout réel s dans $I \cap J$ assurant que les deux termes de la dernière égalité sont finis, on a bien l'existence de \mathcal{L}_{X+Y} sur $I \cap J$ et

$$\mathcal{L}_{X+Y}(s) = \mathcal{L}_X(s) \mathcal{L}_Y(s). \quad \square$$

Nous attirons l'attention du lecteur sur le fait que la réciproque est fausse. Avoir $\mathcal{L}_{X+Y}(s) = \mathcal{L}_X(s) \mathcal{L}_Y(s)$ n'implique pas l'indépendance entre X et Y .

Exemple. Supposons que les v.a.r. X et Y soient indépendantes et de loi respectivement de poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. D'après le théorème on a donc :

$$\mathcal{L}_{X+Y}(s) = \mathcal{L}_X(s) \mathcal{L}_Y(s) = e^{\lambda(e^s-1)} e^{\mu(e^s-1)} = e^{(\lambda+\mu)(e^s-1)} = \mathcal{L}_{\mathcal{P}(\lambda+\mu)}(s).$$

Par caractérisation, la loi de la v.a.r. $X + Y$ est donc une loi de Poisson $\mathcal{P}(\lambda + \mu)$. \diamond

Nous admettons enfin le théorème suivant.

Théorème 4.1.5 *Soit X une variable aléatoire admettant une transformée de Laplace sur un intervalle ouvert $I =]-u_1, u_2[$ (différent du vide) de \mathbb{R} , où u_1 et u_2 sont des réels strictement positifs. On a alors :*

i) *la variable aléatoire admet tous les moments d'ordre entier, i.e.*

$$\forall k \in \mathbb{N}^* : \quad \mathbb{E}(|X|^k) < +\infty;$$

ii) *si $u_0 = \min(u_1, u_2)$ et $t < u_0$, alors:*

$$\mathcal{L}_X(s) = \sum_{n=1}^{+\infty} \frac{s^n}{n!} \mathbb{E}X^n,$$

pour tout s dans $] -t, t[$, ce qui justifie son appellation comme fonction génératrice des moments ;

iii) *on a, pour tout entier k positif,*

$$\mathbb{E}X^k = \left. \frac{\partial^k \mathcal{L}_X(s)}{\partial s^k} \right|_{s=0} = \mathcal{L}_X^{(k)}(0).$$

4.1.2 Vecteurs aléatoires

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire dans \mathbb{R}^n .

Définition 4.1.6 On appelle *transformée de Laplace du vecteur X* (si elle existe) la fonction de \mathbb{R}^n vers \mathbb{R} définie pour $s = (s_1, \dots, s_n)$ par :

$$\mathcal{L}_X(s) = \mathbb{E} \left(e^{\langle s, X \rangle} \right) = \mathbb{E} \left(e^{\sum_{i=1}^n s_i X_i} \right).$$

Les propriétés restent les mêmes que dans le cas unidimensionnel. Mais son inconvénient majeur de ne pas toujours exister reste également le même.

Proposition 4.1.7 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de \mathbb{R}^n admettant une transformée de Laplace sur un ouvert \mathcal{O} de \mathbb{R}^n . Les v.a.r. $(X_j)_{j=1, \dots, n}$ sont indépendantes si, et seulement si, pour tout $s = (s_1, \dots, s_n)$ dans \mathcal{O} on a :

$$\mathcal{L}_{X_1, \dots, X_n}(s_1, \dots, s_n) = \prod_{i=1}^n \mathcal{L}_{X_i}(s_i).$$

4.2 Fonction caractéristique

4.2.1 Intégrale d'une variable aléatoire complexe

Définition 4.2.1 On appelle *variable aléatoire complexe* toute application Z de (Ω, \mathcal{A}, P) vers \mathbb{C} qui à tout ω dans Ω associe

$$Z(\omega) = X(\omega) + iY(\omega),$$

où X et Y sont des variables aléatoires réelles.

Définition 4.2.2 On dit qu'une variable aléatoire complexe Z est *P -intégrable* si les v.a.r. X et Y sont P -intégrables. L'intégrale de Z par rapport à P est alors définie par :

$$\mathbb{E}Z = \int Z dP = \int X dP + i \int Y dP = \mathbb{E}X + i\mathbb{E}Y.$$

Notons que les principales propriétés de l'intégrale sont conservées.

4.2.2 Fonction caractéristique

Définition 4.2.3 On appelle fonction caractéristique d'un vecteur aléatoire $X = (X_1, \dots, X_n)$ de \mathbb{R}^n , la fonction définie pour tout $t = (t_1, \dots, t_n)$ de \mathbb{R}^n et à valeurs dans \mathbb{C} par :

$$\varphi_X(t) = \varphi_X(t_1, \dots, t_n) = \mathbb{E} \left(e^{i \langle t, X \rangle} \right) \quad (4.1)$$

$$= \mathbb{E} \left(e^{i \sum_{j=1}^n t_j X_j} \right). \quad (4.2)$$

La fonction caractéristique φ_X d'une v.a.r. X est donc :

$$\varphi_X(t) = \mathbb{E}(e^{itX}).$$

L'analogie avec la transformée de Laplace est grande et ses propriétés sont similaires. De plus, la fonction caractéristique existe toujours ce qui n'était pas le cas pour la transformée de Laplace. Elle a cependant l'inconvénient de faire appel à la théorie des fonctions d'une variable aléatoire complexe.

Exemples.

* Supposons que la v.a.r. X suive une loi de Poisson $\mathcal{P}(\lambda)$. On a :

$$\begin{aligned} \varphi_X(t) &= \mathbb{E}(e^{itX}) = \sum_{k=0}^{+\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda e^{it})^k}{k!} \\ &= e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}. \end{aligned}$$

* On peut montrer que si une v.a.r. X suit une loi normale $N(0, 1)$, alors sa fonction caractéristique est définie pour tout t dans \mathbb{R} par :

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

Soit maintenant une variable aléatoire Y de loi $N(\mu, \sigma^2)$. En notant $X = (Y - \mu)/\sigma$, on peut écrire :

$$\begin{aligned} \varphi_Y(t) &= \mathbb{E}(e^{itY}) = \mathbb{E}(e^{it(\sigma X + \mu)}) = \mathbb{E}(e^{it\mu} e^{it\sigma X}) \\ &= e^{it\mu} \mathbb{E}(e^{it\sigma X}) \\ &= e^{it\mu} \varphi_X(t\sigma) = e^{it\mu} e^{-\frac{t^2 \sigma^2}{2}} = \exp \left\{ it\mu - \frac{t^2 \sigma^2}{2} \right\}. \quad \diamond \end{aligned}$$

Pour $X = (X_1, \dots, X_n)$ un vecteur de \mathbb{R}^n , notons φ_X et φ_{X_i} , respectivement les fonctions caractéristiques de X et de X_i . On a alors les résultats suivants.

Théorème 4.2.4 *Pour $X = (X_1, \dots, X_n)$ un vecteur de \mathbb{R}^n , notons φ_X et φ_{X_i} , respectivement les fonctions caractéristiques de X et de X_i . On a alors les résultats suivants :*

- i) $\forall t \in \mathbb{R}^n : |\varphi_X(t)| \leq \varphi_X(0) = 1$;
- ii) $\forall j \in \{1, \dots, n\}$, et $\forall t_j \in \mathbb{R} : \varphi_{X_j}(t_j) = \varphi_X(0, \dots, 0, t_j, 0, \dots, 0)$;
- iii) $\forall t \in \mathbb{R}^n : \overline{\varphi_X(t)} = \varphi_X(-t)$;
- iv) la fonction φ est uniformément continue sur \mathbb{R}^n ;
- v) Si on pose $Y = AX + b$ où A est une matrice $p \times n$ et b un vecteur de \mathbb{R}^p on a, pour tout u dans \mathbb{R}^p :

$$\varphi_Y(u) = e^{i\langle u, b \rangle} \varphi_X(A'u),$$

où A' est la matrice transposée de A ;

- vi) Si X est une v.a.r. dans L^p , où p est un entier, alors φ_X est dérivable p -fois et

$$\forall k \leq p : \varphi_X^{(k)}(0) = i^k \mathbb{E}X^k.$$

Preuve.

- i) On a :

$$\begin{aligned} |\varphi_X(t)| &= \left| \int e^{i\langle t, x \rangle} dP_X(x) \right| \leq \int |e^{i\langle t, x \rangle}| dP_X(x) \\ &= \int dP_X(x) = 1 \end{aligned}$$

et

$$\varphi_X(0) = \mathbb{E}(e^0) = 1.$$

- ii) Par définition, on a :

$$\varphi_X(0, \dots, 0, t_j, 0, \dots, 0) = \mathbb{E}(e^{it_j X_j}) = \varphi_{X_j}(t_j).$$

- iii) On peut écrire :

$$\begin{aligned} \varphi_X(-t) &= \mathbb{E} \left(e^{i\langle -t, X \rangle} \right) = \mathbb{E} \left(e^{-i\langle t, X \rangle} \right) \\ &= \mathbb{E} \left(\overline{e^{i\langle t, X \rangle}} \right) = \overline{\mathbb{E} \left(e^{i\langle t, X \rangle} \right)} = \overline{\varphi_X(t)}. \end{aligned}$$

iv) Admis.

v)

$$\begin{aligned}\varphi_Y(u) &= \mathbb{E}\left(e^{i\langle u, Y \rangle}\right) = \mathbb{E}\left(e^{i\langle u, AX \rangle} e^{i\langle u, b \rangle}\right) \\ &= e^{i\langle u, b \rangle} \mathbb{E}\left(e^{i\langle A'u, X \rangle}\right) = e^{i\langle u, b \rangle} \varphi_X(A'u).\end{aligned}$$

vi) Admis aussi. □

Théorème 4.2.5 *La fonction caractéristique caractérise la loi d'une variable aléatoire. Autrement dit, si deux variables aléatoires ont même fonction caractéristique, elles ont même loi.*

Théorème 4.2.6 *Soit φ une fonction complexe de la variable réelle. Si φ est intégrable, i.e. si*

$$\int_{\mathbb{R}} |\varphi(t)| dt < +\infty$$

et si la fonction définie par

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi(u) e^{-iux} du$$

est aussi intégrable, alors φ est la fonction caractéristique de la variable aléatoire X ayant pour loi P_X de densité f_X par rapport à la mesure de Lebesgue.

Proposition 4.2.7 *Soit (X, Y) un couple de variables aléatoires indépendantes. Alors, pour tout réel t , on a :*

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t).$$

Preuve. Par définition, et grâce à l'indépendance entre X et Y , on peut écrire :

$$\varphi_{X+Y}(t) = \mathbb{E}\left(e^{it(X+Y)}\right) = \mathbb{E}\left(e^{itX}\right) \mathbb{E}\left(e^{itY}\right) = \varphi_X(t) \varphi_Y(t) \quad \square$$

Nous attirons à nouveau l'attention du lecteur sur le fait que la réciproque est fausse.

Théorème 4.2.8 *Une famille $(X_j)_{j=1, \dots, n}$ constitue une famille de v.a.r. indépendantes si, et seulement si, pour tout $t = (t_1, \dots, t_n)$ dans \mathbb{R}^n , on a :*

$$\varphi_{(X_1, \dots, X_n)}(t) = \prod_{j=1}^n \varphi_{X_j}(t_j).$$

Chapitre 5

Vecteurs gaussiens

5.1 Exemple fondamental

Considérons n variables aléatoires X_1, \dots, X_n indépendantes et de loi respectivement $N(m_1, \sigma_1^2), \dots, N(m_n, \sigma_n^2)$.

Pour $i = 1, \dots, n$, la variable aléatoire X_i est donc de densité

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left\{ -\frac{1}{2} \left(\frac{x - m_i}{\sigma_i} \right)^2 \right\}$$

par rapport à la mesure de Lebesgue λ sur \mathbb{R} .

En raison de l'indépendance des variables aléatoires X_i , la densité conjointe du vecteur X_1, \dots, X_n est :

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \frac{1}{\prod_{i=1}^n \sigma_i} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m_i}{\sigma_i} \right)^2 \right\}.$$

D'après leur définition donnée au chapitre II, le vecteur espérance du vecteur $X = (X_1, \dots, X_n)$ et sa matrice de covariance sont :

$$\mathbb{E}X = m = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} \quad \text{et} \quad \Gamma_X = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix}.$$

Notons que la matrice Γ_X est diagonale en raison de l'indépendance des v.a.r. $(X_i)_{i=1, \dots, n}$. Comme toutes les variances σ_i sont strictement positives, on obtient aisément la matrice inverse

$$\Gamma_X^{-1} = \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_n^2 \end{pmatrix}.$$

On peut alors réécrire la densité conjointe du vecteur $X = (X_1, \dots, X_n)$ sous la forme

$$f_X(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \times \frac{1}{\sqrt{\det(\Gamma_X)}} \exp \left\{ -\frac{1}{2} (x - m)' \Gamma_X^{-1} (x - m) \right\},$$

puisque

$$\begin{aligned} & (x - m)' \Gamma_X^{-1} (x - m) \\ = & (x_1 - m_1, \dots, x_n - m_n) \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_n^2 \end{pmatrix} \begin{pmatrix} x_1 - m_1 \\ \vdots \\ x_n - m_n \end{pmatrix} \\ = & \sum_{i=1}^n \frac{(x_i - m_i)^2}{\sigma_i^2}. \end{aligned}$$

Intéressons-nous maintenant à la fonction caractéristique du vecteur X . Toujours en raison de l'indépendance, on a, pour tout $\lambda = (\lambda_1, \dots, \lambda_n)$ de \mathbb{R}^n :

$$\varphi_X(\lambda) = \varphi_{X_1, \dots, X_n}(\lambda_1, \dots, \lambda_n) = \prod_{j=1}^n \varphi_{X_j}(\lambda_j).$$

Or, on a vu dans le chapitre précédent que la fonction caractéristique d'une v.a.r. de loi $N(m_j, \sigma_j^2)$ est :

$$\varphi_{X_j}(\lambda_j) = \exp \left\{ i\lambda_j m_j - \frac{1}{2} \lambda_j^2 \sigma_j^2 \right\}$$

d'où on tire :

$$\begin{aligned} \varphi_{X_1, \dots, X_n}(\lambda) &= \exp \left\{ i \sum_{j=1}^n \lambda_j m_j - \frac{1}{2} \sum_{j=1}^n \lambda_j^2 \sigma_j^2 \right\} \\ &= \exp \left\{ i\lambda' m - \frac{1}{2} \lambda' \Gamma_X \lambda \right\}. \end{aligned}$$

Remarquons enfin que toute combinaison linéaire des X_j , pour $j = 1, \dots, n$, est de loi normale dans \mathbb{R} . Une combinaison linéaire des X_j s'écrit en effet de manière générale sous la forme :

$$\langle \lambda, X \rangle = \lambda' X$$

où $\lambda = (\lambda_1, \dots, \lambda_n)$ est un vecteur de \mathbb{R}^n . Il vient alors, pour tout u dans \mathbb{R} :

$$\begin{aligned} \varphi_{\langle \lambda, X \rangle}(u) &= \mathbb{E} (e^{iu \langle \lambda, X \rangle}) = \mathbb{E} (e^{i \langle u\lambda, X \rangle}) \\ &= \varphi_X(u\lambda) = \varphi_X(u\lambda_1, \dots, u\lambda_n) \\ &= \exp \left\{ iu\lambda' m - \frac{1}{2} u^2 \lambda' \Gamma_X \lambda \right\}. \end{aligned}$$

La fonction caractéristique de la v.a.r. $\langle \lambda, X \rangle$ est donc de la forme :

$$\varphi_{\langle \lambda, X \rangle}(u) = \exp \left\{ iua - \frac{1}{2} u^2 b \right\},$$

avec $a = \lambda' m$ et $b = \lambda' \Gamma_X \lambda$. Par caractérisation, la v.a.r. $\langle \lambda, X \rangle$ est donc de loi $N(\lambda' m, \lambda' \Gamma_X \lambda)$.

5.2 Définition

Définition 5.2.1 Un vecteur aléatoire $X = (X_1, \dots, X_n)$ de \mathbb{R}^n est dit vecteur gaussien si, pour tout $\lambda = (\lambda_1, \dots, \lambda_n)$ de \mathbb{R}^n , la v.a.r.

$$\lambda'X = \sum_{i=1}^n \lambda_i X_i$$

est une v.a.r. de loi normale. Autrement dit, si toute combinaison linéaire des composantes de (X_1, \dots, X_n) est de loi normale.

Si son vecteur des espérances est m et sa matrice de covariance est Γ_X , on note $X \sim N_n(m, \Gamma_X)$.

Remarquons que l'on peut en particulier en déduire que toutes les composantes du vecteur X sont des v.a.r. de loi normale. En revanche, la réciproque est fautive. Un vecteur dont toutes les composantes sont de loi normale, n'est pas nécessairement un vecteur gaussien.

La définition précédente implique également que tout sous vecteur d'un vecteur gaussien est encore un vecteur gaussien.

Proposition 5.2.2 Si X est un vecteur gaussien de vecteur des espérances $m = (m_1, \dots, m_n)$ et de matrice de covariance Γ_X , alors, pour tout λ dans \mathbb{R}^n , la v.a.r. $\lambda'X = \langle \lambda, X \rangle$ est de loi $N(\lambda'm, \lambda'\Gamma_X\lambda)$.

Preuve. On utilise d'abord le fait que, par définition d'un vecteur gaussien, la v.a.r. $\lambda'X$ est de loi normale. Il ne reste plus qu'à calculer son espérance et sa variance. On utilise alors les résultats vus au chapitre IV, pour obtenir :

$$\begin{aligned} \mathbb{E}(\lambda'X) &= \lambda'\mathbb{E}X = \lambda'm \\ \text{et} \quad \Gamma_{\lambda'X} &= \lambda'\Gamma_X\lambda. \end{aligned} \quad \square$$

On peut aussi caractériser un vecteur gaussien par sa fonction caractéristique, grâce à la proposition suivante.

Proposition 5.2.3 Pour qu'un vecteur X de \mathbb{R}^n soit un vecteur gaussien, il faut et il suffit qu'il existe un vecteur m de \mathbb{R}^n et une matrice Γ symétrique et positive de dimension $n \times n$ tels que, pour tout vecteur λ de \mathbb{R}^n , on ait :

$$\varphi_X(\lambda_1, \dots, \lambda_n) = \exp \left\{ i\lambda'm - \frac{1}{2}\lambda'\Gamma\lambda \right\}.$$

Dans ce cas, on a : $\mathbb{E}X = m$ et $\Gamma_X = \Gamma$.

Preuve. Supposons que X soit un vecteur gaussien. Toute v.a.r. de la forme $\lambda'X$, pour λ dans \mathbb{R}^n , est donc de loi $N(\lambda'm, \lambda'\Gamma_X\lambda)$. Ainsi sa fonction caractéristique est :

$$\varphi_{\lambda'X}(u) = \mathbb{E}(e^{iu\lambda'X}) = \exp \left\{ iu\lambda'm - \frac{1}{2}u^2\lambda'\Gamma_X\lambda \right\}$$

En posant $u = 1$ dans cette équation, on obtient :

$$\mathbb{E}(e^{i\lambda'X}) = \exp \left\{ i\lambda'm - \frac{1}{2}\lambda'\Gamma_X\lambda \right\},$$

Ce qui est bien l'expression annoncée pour la fonction caractéristique.

Réciproquement, soit X un vecteur aléatoire dans \mathbb{R}^n ayant pour fonction caractéristique

$$\varphi_X(\lambda) = \exp \left\{ i\lambda'm - \frac{1}{2}\lambda'\Gamma_X\lambda \right\} = \mathbb{E} \left(e^{i\langle \lambda, X \rangle} \right),$$

pour tout λ dans \mathbb{R}^n . Notons maintenant $Y = \langle \lambda, X \rangle$ la variable aléatoire réelle dont la fonction caractéristique est, pour tout u dans \mathbb{R} :

$$\begin{aligned} \varphi_Y(u) &= \mathbb{E}(e^{iuY}) = \mathbb{E}(e^{iu\lambda'X}) = \mathbb{E}(e^{i\langle \lambda u, X \rangle}) \\ &= \exp \left\{ iu\lambda'm - \frac{1}{2}u^2\lambda'\Gamma_X\lambda \right\} \\ &= \exp \left\{ iua - \frac{1}{2}u^2b \right\} \end{aligned}$$

où $a = \lambda'm$ et $b = \lambda'\Gamma_X\lambda$. Par caractérisation, la v.a.r. Y est donc de loi $N(\lambda'm, \lambda'\Gamma_X\lambda)$. On a donc démontré que toute combinaison linéaire des composantes du vecteur X est de loi normale, et par définition il s'agit bien d'un vecteur gaussien. \square

Notons que, dans tout ce qui précède, la matrice Γ_X n'est pas supposée inversible. En revanche, la définition d'un vecteur gaussien par sa densité, par rapport à la mesure de Lebesgue dans \mathbb{R}^n , n'est possible que si cette matrice est inversible, comme l'affirme la proposition suivante.

Proposition 5.2.4 *Soit X un vecteur gaussien dans \mathbb{R}^n d'espérance m et de matrice des covariances Γ_X . Lorsque Γ_X est inversible, le vecteur aléatoire X est dit vecteur aléatoire gaussien non dégénéré et sa loi est absolument continue par rapport à la mesure de Lebesgue dans \mathbb{R}^n et admet pour densité*

$$f_X(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(\Gamma_X)}} \exp \left\{ -\frac{1}{2}(x - m)'\Gamma_X^{-1}(x - m) \right\}.$$

Un vecteur gaussien de matrice de covariance Γ_X telle que $\det(\Gamma_X) = 0$ (i.e. Γ_X non inversible) est dit dégénéré et n'admet pas de densité par rapport à la mesure de Lebesgue dans \mathbb{R}^n .

5.3 Propriétés des vecteurs aléatoires gaussiens

5.3.1 Transformation linéaire d'un vecteur gaussien

Proposition 5.3.1 *La transformée d'un vecteur gaussien de \mathbb{R}^n par une application linéaire de \mathbb{R}^n vers \mathbb{R}^p est encore un vecteur gaussien.*

Preuve. Soit X un vecteur gaussien de \mathbb{R}^n , de vecteur des espérances m et de matrice de covariance Γ_X . Soit A la matrice associée à une transformation linéaire quelconque de \mathbb{R}^n vers \mathbb{R}^p . La matrice A est donc de dimension $p \times n$. Calculons la fonction caractéristique du vecteur aléatoire $Y = AX$.

D'après ce que l'on a vu au chapitre précédent, pour tout λ de \mathbb{R}^p , on a :

$$\begin{aligned}\varphi_Y(\lambda) &= \varphi_{AX}(\lambda) = \mathbb{E} \left(e^{i\langle \lambda, AX \rangle} \right) = \mathbb{E} \left(e^{i\langle A' \lambda, X \rangle} \right) \\ &= \varphi_X(A' \lambda) = \exp \left\{ i\lambda' A m - \frac{1}{2} \lambda' A \Gamma_X A' \lambda \right\}.\end{aligned}$$

Par caractérisation, le vecteur Y est donc un vecteur gaussien dans \mathbb{R}^p de vecteur des espérances $A m$ et de matrice de covariance $A \Gamma_X A'$, i.e.

$$Y \sim N_p(A m, A \Gamma_X A'). \quad \square$$

5.3.2 Vecteur gaussien et indépendance

On a vu précédemment (au chapitre III) que, d'une manière générale, l'indépendance entraîne la non corrélation mais que la réciproque est fautive. Dans le cas d'un vecteur gaussien il y a équivalence, comme le montre la proposition suivante.

Proposition 5.3.2 *Soit X un vecteur gaussien dans \mathbb{R}^n . Pour que ses composantes X_1, \dots, X_n soient indépendantes, il faut et il suffit que la matrice de covariance soit diagonale.*

Preuve. Il suffit, bien sûr, de montrer la réciproque. Supposons donc que Γ_X soit diagonale, i.e.

$$\Gamma_X = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix}.$$

Comme X est un vecteur gaussien de loi $N_n(m, \Gamma_X)$, chacune de ses composantes X_j , pour $j = 1, \dots, n$, est de loi normale $N(m_j, \sigma_j^2)$ et de fonction caractéristique :

$$\varphi_{X_j}(\lambda_j) = \exp \left\{ i\lambda_j m_j - \frac{1}{2} \sigma_j^2 \lambda_j^2 \right\},$$

pour tout λ_j dans \mathbb{R} .

Par ailleurs, la fonction caractéristique du vecteur X est, pour tout λ dans \mathbb{R}^n :

$$\begin{aligned} \varphi_X(\lambda) &= \exp \left\{ i\lambda' m - \frac{1}{2} \lambda' \Gamma_X \lambda \right\} \\ &= \exp \left\{ i \sum_{j=1}^n \lambda_j m_j - \frac{1}{2} \sum_{j=1}^n \lambda_j^2 \sigma_j^2 \right\} \\ &= \exp \left\{ \sum_{j=1}^n \left\{ i\lambda_j m_j - \frac{1}{2} \lambda_j^2 \sigma_j^2 \right\} \right\} \\ &= \prod_{j=1}^n \exp \left\{ i\lambda_j m_j - \frac{1}{2} \lambda_j^2 \sigma_j^2 \right\} = \prod_{j=1}^n \varphi_{X_j}(\lambda_j). \end{aligned}$$

Un résultat du chapitre IV permet d'en déduire l'indépendance. \square

Corollaire 5.3.3 *Si le couple (X, Y) est un vecteur gaussien, on a*

$$X \perp\!\!\!\perp Y \Leftrightarrow \text{Cov}(X, Y) = 0.$$

Preuve. Immédiate. \square

Nous attirons l'attention du lecteur sur le fait que deux variables aléatoires réelles gaussiennes et non corrélées ne sont pas nécessairement indépendantes. Pour s'assurer qu'elles le soient il faut pour cela qu'elles constituent un couple gaussien.

Contre-exemple. Considérons une v.a.r. X de loi $N(0, 1)$ et ε une v.a.r. discrète de loi définie par :

$$p(\varepsilon = 1) = \frac{1}{2} \quad \text{et} \quad p(\varepsilon = -1) = \frac{1}{2}$$

et telle que les v.a.r. ε et X soient indépendantes. On pose $Y = \varepsilon X$ et calculons la loi de Y . On a :

$$\begin{aligned} F_Y(y) &= P(\varepsilon X \leq y) \\ &= P(\{\varepsilon X \leq y\} \cap \{\varepsilon = 1\}) + P(\{\varepsilon X \leq y\} \cap \{\varepsilon = -1\}) \\ &= P(\{X \leq y\} \cap \{\varepsilon = 1\}) + P(\{-X \leq y\} \cap \{\varepsilon = -1\}) \\ &= P(X \leq y)P(\varepsilon = 1) + P(X \geq -y)P(\varepsilon = -1) \\ &= \frac{1}{2}P(X \leq y) + \frac{1}{2}P(X \leq y) = F_X(y). \end{aligned}$$

Ainsi la v.a.r. Y est de loi $N(0, 1)$.

Par ailleurs, puisque X et Y sont centrées et que ε et X sont indépendantes, on a :

$$\text{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = \mathbb{E}(\varepsilon X^2) = \mathbb{E}\varepsilon\mathbb{E}X^2 = 0$$

Les v.a.r. X et Y sont donc non corrélées et cependant elles ne sont pas indépendantes. En effet, en raisonnant par l'absurde, supposons qu'elles le soient. D'après ce que l'on a vu au début de ce chapitre, le couple (X, Y) serait gaussien et $X + Y$ serait alors de loi normale et donc absolument continue.

Or, en notant que $X + Y = (1 + \varepsilon)X$, on a :

$$P(X + Y = 0) = P(1 + \varepsilon = 0) = P(\varepsilon = -1) = \frac{1}{2},$$

ce qui contredit le fait que la v.a.r. $X + Y$ soit absolument continue. Les v.a.r. X et Y ne sont donc pas indépendantes. \diamond

Chapitre 6

Convergences

6.1 Convergence en loi

6.1.1 Définition

Définition 6.1.1 Soit (X_n) et X des vecteurs aléatoires à valeurs dans l'espace probabilisable $(\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p})$. On dit que la suite (X_n) converge en loi vers X si, pour toute fonction h de \mathbb{R}^p vers \mathbb{R} , continue et bornée, on a

$$\lim_{n \rightarrow +\infty} \mathbb{E}h(X_n) = \mathbb{E}h(X).$$

On note $X_n \xrightarrow{\mathcal{L}} X$ et on dit aussi parfois que la loi de X_n converge vers celle de X .

Théorème 6.1.2 (Théorème de Slutsky)

Soit (X_n) et X des vecteurs aléatoires dans \mathbb{R}^p , tels que (X_n) converge en loi vers X . Si g est une application continue de \mathbb{R}^p vers \mathbb{R}^q , alors on a :

$$g(X_n) \xrightarrow{\mathcal{L}} g(X).$$

6.1.2 Caractérisation de la convergence en loi

Voyons dans un premier temps une condition nécessaire et suffisante pour la convergence en loi dans le cas de v.a.r.

Proposition 6.1.3 Soit (X_n) et X des v.a.r. de fonction de répartition (F_n) et F respectivement. La suite (X_n) converge en loi vers X si, et seulement si,

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x),$$

en tout point x où F est continue.

Exemple. On considère la suite (X_n) de v.a.r. telle que, pour tout n , la v.a.r. X_n ait pour loi

$$P\left(X_n = 2 + \frac{1}{n}\right) = 1,$$

i.e. la loi de X_n est la dirac en $2 + \frac{1}{n}$ ($P_{X_n} = \delta_{2+\frac{1}{n}}$). En raison de la convergence de la suite $(2 + 1/n)$ vers 2, on a :

$$\begin{aligned} \forall x > 2, \exists n_0 : \forall n > n_0, 2 + \frac{1}{n} < x \\ \forall x > 2, \exists n_0 : \forall n > n_0, F_n(x) = P(X_n \leq x) = 1. \end{aligned}$$

Par ailleurs, pour tout $x \leq 2$, on a :

$$F_n(x) = P(X_n \leq 2) = 0.$$

Définissons alors X la v.a.r. de loi δ_2 . Sa fonction de répartition est alors :

$$F_X(x) = \begin{cases} 0 & \text{si } x < 2 \\ 1 & \text{si } x \geq 2 \end{cases}.$$

On remarque que la fonction F_X est continue sur $\mathbb{R} \setminus \{2\}$ et que, sur cet ensemble, on a :

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x).$$

Ainsi, d'après la proposition précédente, on a la convergence de X_n vers X .

Il est intéressant de noter que la convergence des fonctions de répartition n'a pas lieu au point de discontinuité de F puisque l'on a, pour tout n ,

$$F_n(2) = 0 \neq F(2) = 1. \quad \diamond$$

Théorème 6.1.4 *Soit (X_n) et X des vecteurs aléatoires de \mathbb{R}^p , absolument continus de densité (f_n) et f par rapport à la mesure de Lebesgue dans \mathbb{R}^p . Si on a, λ_p -presque-partout,*

$$\lim_n f_n = f,$$

alors

$$X_n \xrightarrow{\mathcal{L}} X.$$

Théorème 6.1.5 (Théorème de Paul Lévy)

1) *Si (X_n) est une suite de variables aléatoires dans \mathbb{R}^p convergeant en loi vers une variable aléatoire X dans \mathbb{R}^p , alors la suite (φ_{X_n}) des fonctions caractéristiques associée à la suite (X_n) converge en tout point vers la fonction caractéristique φ_X de X , i.e.*

$$X_n \xrightarrow{\mathcal{L}} X \quad \Rightarrow \quad \forall x \in \mathbb{R}^p, \varphi_{X_n}(x) \rightarrow \varphi_X(x).$$

2) *Soit (X_n) est une suite de variables aléatoires dans \mathbb{R}^p . Si la suite (φ_{X_n}) de ses fonctions caractéristiques converge simplement vers une fonction φ continue en 0, alors φ est la fonction caractéristique d'une variable aléatoire X et X_n converge en loi vers X , i.e.*

$$\varphi_{X_n}(x) \rightarrow \varphi(x), \quad \forall x \in \mathbb{R}^p \quad \Rightarrow \quad X_n \xrightarrow{\mathcal{L}} X,$$

où X est une variable aléatoire de fonction caractéristique φ .

Théorème 6.1.6 (Théorème de Cramer-Wold)

Soit (X_n) et X des vecteurs aléatoires dans \mathbb{R}^p . On a alors l'équivalence suivante :

$$X_n \xrightarrow{\mathcal{L}} X \quad \Leftrightarrow \quad \forall u \in \mathbb{R}^p : u'X_n \xrightarrow{\mathcal{L}} u'X.$$

Preuve. Supposons en premier lieu que X_n converge en loi vers X . La fonction g de \mathbb{R}^p vers \mathbb{R} définie par $g(x) = u'x$, pour u dans \mathbb{R}^p , est une forme linéaire. Elle est donc continue. Ainsi, d'après le théorème de Slutsky, on a la convergence

$$u'X_n \xrightarrow{\mathcal{L}} u'X.$$

Réciproquement, supposons que pour tout u dans \mathbb{R}^p , on ait

$$u'X_n \xrightarrow{\mathcal{L}} u'X.$$

Le théorème de Paul Lévy, nous donne alors la convergence

$$\varphi_{u'X_n}(t) \rightarrow \varphi_{u'X}(t),$$

pour tout t dans \mathbb{R} . Celle-ci prise en $t = 1$, nous donne :

$$\varphi_{u'X_n}(1) = \varphi_{X_n}(u) \rightarrow \varphi_X(u) = \varphi_{u'X}(1)$$

dont on tire, en utilisant la réciproque du théorème de Paul Lévy, la convergence $X_n \xrightarrow{\mathcal{L}} X$ □

6.1.3 Approximation de lois

Concrètement, un des intérêts de la notion de convergence en loi d'une suite de v.a. (X_n) vers X est d'approcher la loi de X_n , qui est souvent inconnue ou difficilement utilisable, par la loi de X .

a) Approximation de la loi binomiale par la loi de Poisson.

Proposition 6.1.7 Soit (p_n) une suite de nombres réels strictement positifs tels que :

$$\lim_{n \rightarrow +\infty} np_n = \lambda,$$

où λ est un réel strictement positif. Si, pour tout n , X_n est une v.a.r. de loi $B(n, p_n)$, alors (X_n) converge en loi vers une v.a.r. X de loi de Poisson de paramètre λ .

Preuve. On admet le résultat suivant dans le corps des complexes :

$$\text{si } z_n \rightarrow z \text{ alors } \lim_{n \rightarrow +\infty} \left(1 + \frac{z_n}{n}\right)^n = e^z.$$

On a déjà vu que si X_n est de loi binomiale $B(n, p_n)$ alors sa fonction caractéristique est :

$$\varphi_{X_n}(t) = (1 - p_n + p_n e^{it})^n = \left(1 - \frac{np_n(1 - e^{it})}{n}\right)^n.$$

Or, on a :

$$\lim_{n \rightarrow +\infty} np_n(1 - e^{it}) = \lambda(1 - e^{it}),$$

par suite de l'hypothèse sur les p_n . On a donc la convergence, pour tout t dans \mathbb{R} ,

$$\lim_{n \rightarrow +\infty} \varphi_{X_n}(t) = e^{\lambda(e^{it}-1)} = \varphi_X(t),$$

où la v.a.r. X est de loi de Poisson $\mathcal{P}(\lambda)$. □

b) **Théorème de la limite centrale** (central limit en anglais)

Le théorème suivant est fondamental et très souvent utilisé en Statistique. Notons auparavant $z_{n,1}, \dots, z_{n,p}$ les p -coordonnées d'un vecteur z_n de \mathbb{R}^p . Notons également \bar{z}_n le vecteur des moyennes des composantes des n premiers vecteurs de la suite (z_n) , i.e.

$$\bar{z}_n = \begin{pmatrix} \frac{1}{n} \sum_{j=1}^n z_{j,1} \\ \vdots \\ \frac{1}{n} \sum_{j=1}^n z_{j,p} \end{pmatrix}.$$

Théorème 6.1.8 (Théorème de la limite centrale multidimensionnel)

Soit (Z_n) une suite de vecteurs aléatoires dans $(\mathbb{R}^p, B_{\mathbb{R}^p})$, indépendants, de même loi de moyenne μ et de matrice de covariance Σ . On a alors :

$$\sqrt{n}(\bar{Z}_n - \mu) \xrightarrow{\mathcal{L}} N_p(0, \Sigma).$$

Pour démontrer ce théorème nous utiliserons son cas particulier, correspondant au cas unidimensionnel.

Théorème 6.1.9 (Théorème de la limite centrale unidimensionnel)

Soit (X_n) une suite de v.a.r. indépendantes, dans L^2 et de même loi de moyenne μ et de variance σ^2 . On a alors

$$\sqrt{n} \left(\frac{\frac{1}{n} \sum_{j=1}^n X_j - \mu}{\sigma} \right) = \frac{\sum_{j=1}^n X_j - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} N(0, 1).$$

Preuve (du théorème unidimensionnel). Notons

$$\begin{aligned} Y_n &= \frac{\sqrt{n} \left(\frac{1}{n} \sum_{j=1}^n X_j - \mu \right)}{\sigma} = \frac{\sqrt{n}}{n} \left(\sum_{j=1}^n \frac{(X_j - \mu)}{\sigma} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n U_j \end{aligned}$$

où les v.a.r. U_j sont définies par

$$U_j = \frac{X_j - \mu}{\sigma},$$

pour $j = 1 \dots, n$. Ces dernières sont, par hypothèse, centrées réduites, de même loi et indépendantes. On peut alors écrire :

$$\begin{aligned} \varphi_{Y_n}(t) &= \mathbb{E}(e^{itY_n}) = \mathbb{E}\left(e^{it \frac{1}{\sqrt{n}} \sum_{j=1}^n U_j}\right) \\ &= \prod_{j=1}^n \varphi_{U_j}\left(\frac{t}{\sqrt{n}}\right) = \left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n, \end{aligned}$$

où φ est la fonction caractéristique des v.a.r. U_j . Or, en utilisant les propriétés de la fonction caractéristique vues au chapitre IV, on a :

$$\begin{aligned} \varphi'(0) &= i \mathbb{E}U_j = 0 \\ \text{et} \quad \varphi''(0) &= i^2 \mathbb{E}(U_j^2) = -\text{Var } U_j = -1. \end{aligned}$$

Le développement de Taylor de φ à l'ordre 2 et en 0, est alors :

$$\begin{aligned} \varphi(u) &= \varphi(0) + \varphi'(0)u + \frac{1}{2} \varphi''(0)u^2 + u^2\varepsilon(u) \\ &= 1 - \frac{u^2}{2} + u^2\varepsilon(u), \end{aligned}$$

avec ε une fonction telle que $\lim_{u \rightarrow 0} \varepsilon(u) = 0$.

Ainsi, on peut écrire :

$$\varphi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + \frac{t^2}{n} \varepsilon\left(\frac{t}{\sqrt{n}}\right)$$

et

$$\varphi_{Y_n}(t) = \left(1 - \frac{t^2}{2n} + \frac{t^2}{n} \varepsilon\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{\frac{t^2}{2} + t^2\varepsilon\left(\frac{t}{\sqrt{n}}\right)}{n}\right)^n,$$

dont on tire aisément la convergence :

$$\varphi_{Y_n}(t) \rightarrow e^{-t^2/2},$$

quand n tend vers $+\infty$. Reconnaisant, à la limite, la fonction caractéristique d'une loi $N(0, 1)$, le théorème de Paul-Lévy nous donne la convergence

$$Y_n \xrightarrow{\mathcal{L}} X,$$

où X est une v.a.r. de loi $N(0, 1)$. □

Preuve (du théorème multidimensionnel). Pour tout u dans \mathbb{R}^p , notons $X_n = u'Z_n$. Par hypothèse, les v.a.r. constituant la suite (X_n) sont donc indépendantes, de même loi, d'espérance $\mathbb{E}X_n = u'\mu$ et de variance

$$\text{Var } X_n = u'\Sigma u$$

En utilisant le théorème unidimensionnel, il vient alors :

$$\sqrt{n} \left(\frac{\frac{1}{n} \sum_{j=1}^n X_j - u'\mu}{\sqrt{u'\Sigma u}} \right) \xrightarrow{\mathcal{L}} N(0, 1).$$

Cette convergence peut être réécrite sous la forme :

$$\sqrt{n} \left(\frac{1}{n} \sum_{j=1}^n X_j - u'\mu \right) \xrightarrow{\mathcal{L}} N(0, u'\Sigma u) \Leftrightarrow u' \cdot \sqrt{n}(\bar{Z}_n - \mu) \xrightarrow{\mathcal{L}} u' \cdot Z,$$

où Z est un vecteur gaussien $N_p(0, \Sigma)$.

Ce résultat étant vrai pour tout u dans \mathbb{R}^p , le théorème de Cramer-Wold nous permet de conclure que

$$\sqrt{n}(\bar{Z}_n - \mu) \xrightarrow{\mathcal{L}} Z,$$

quand n tend vers $+\infty$. □

6.2 Convergence en probabilité

6.2.1 Définition

a) Cas des variables aléatoires réelles

Définition 6.2.1 On dit que la suite (X_n) de v.a.r. converge en probabilité vers la variable aléatoire X , si

$$\forall \varepsilon > 0, \quad P(|X_n - X| > \varepsilon) \longrightarrow 0, \text{ quand } n \rightarrow +\infty.$$

On note $X_n \xrightarrow{P} X$.

Remarquons qu'il est équivalent de dire

$$\lim_{n \rightarrow +\infty} P(|X_n - X| > \varepsilon) = 0$$

et

$$\lim_{n \rightarrow +\infty} P(|X_n - X| \leq \varepsilon) = 1.$$

Le théorème suivant nous donne une condition suffisante pour avoir la convergence en probabilité vers une constante.

Proposition 6.2.2 Soit (X_n) une suite de v.a.r. dans L^2 . Si on a

$$\lim_{n \rightarrow +\infty} \mathbb{E}X_n = a \quad \text{et} \quad \lim_{n \rightarrow +\infty} \text{Var } X_n = 0$$

alors

$$X_n \xrightarrow{P} a.$$

Preuve. Grâce à l'inégalité de Bienaymé-Tchebychev, on peut écrire, pour tout $\varepsilon > 0$

$$P(|X_n - a| > \varepsilon) \leq \frac{\mathbb{E}(X_n - a)^2}{\varepsilon^2}.$$

Or, on a déjà vu que

$$\mathbb{E}(X_n - a)^2 = \text{Var } X_n + (\mathbb{E}X_n - a)^2.$$

D'où :

$$\forall \varepsilon > 0, \quad P(|X_n - a| > \varepsilon) \leq \frac{\text{Var } X_n + (\mathbb{E}X_n - a)^2}{\varepsilon^2}$$

et en utilisant les deux hypothèses, on a bien :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} P(|X_n - a| > \varepsilon) = 0$$

et donc $X_n \xrightarrow{P} a$. □

Proposition 6.2.3 (Loi faible des grands nombres)

Soit (X_n) une suite de v.a.r. dans L^1 , indépendantes et identiquement distribuées de moyenne μ .

On a alors la convergence suivante :

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{P} \mu.$$

Théorème 6.2.4 (Théorème de Slutsky) Soit (X_n) et X des v.a.r. Si (X_n) converge en probabilité vers la v.a. X et si g est une application continue de \mathbb{R} vers \mathbb{R} , alors

$$g(X_n) \xrightarrow{P} g(X).$$

b) Cas des vecteurs aléatoires

Définition 6.2.5 Soit (X_n) et X des vecteurs aléatoires à valeurs dans \mathbb{R}^p . On dit que (X_n) converge en probabilité vers le vecteur aléatoire X si ses p composantes $X_{n,1}, \dots, X_{n,p}$ convergent en probabilité vers les composantes X_1, \dots, X_p de X .

Le théorème suivant permet de donner une définition équivalente à cette convergence en probabilité .

Théorème 6.2.6 Soit $\| \cdot \|$ une norme quelconque dans \mathbb{R}^p et (X_n) et X des vecteurs aléatoires dans \mathbb{R}^p . La suite (X_n) converge en probabilité vers la v.a. X si, et seulement si,

$$\| X_n - X \| \xrightarrow{P} 0, \text{ quand } n \rightarrow +\infty.$$

Preuve. Démontrons ce résultat pour la norme supérieure dans \mathbb{R}^p . Supposons en premier lieu que X_n converge en probabilité vers X et notons

$$Y = \| X_n - X \| = \max_i | X_{n,i} - X_i |.$$

De l'inégalité

$$\{Y > \varepsilon\} \subset \bigcup_{i=1}^p \{| X_{n,i} - X_i | > \varepsilon\}.$$

on tire

$$P(Y > \varepsilon) \leq \sum_{i=1}^p P\{| X_{n,i} - X_i | > \varepsilon\}.$$

Par hypothèse, le terme de droite de cette dernière l'inégalité converge en probabilité vers 0, d'où on tire la convergence

$$Y \xrightarrow{P} 0, \text{ quand } n \rightarrow +\infty.$$

Réciproquement, supposons que la v.a.r. Y converge en probabilité vers 0. Ayant, pour tout $i = 1, \dots, p$,

$$\{|X_{n,i} - X_i| > \varepsilon\} \subset \{Y > \varepsilon\}$$

et donc

$$P\{|X_{n,i} - X_i| > \varepsilon\} \leq P(Y > \varepsilon), \quad \forall i = 1, \dots, p$$

on a bien la convergence de X_n vers X en probabilité. \square

Proposition 6.2.7 *Considérons des suites (X_n) et (Y_n) de v.a.r. Si on a les convergences :*

$$\begin{aligned} X_n &\xrightarrow{P} X \\ \text{et} \quad Y_n &\xrightarrow{P} Y \end{aligned}$$

et si g est une fonction continue de \mathbb{R}^2 dans \mathbb{R} , alors

$$g(X_n, Y_n) \xrightarrow{P} g(X, Y).$$

Preuve. Par hypothèse et par définition de la convergence en probabilité d'un vecteur aléatoire, on a la convergence jointe

$$(X_n, Y_n) \xrightarrow{P} (X, Y).$$

Le théorème de Slutsky (également vrai pour les vecteurs aléatoires) entraîne alors le résultat. \square

Corollaire 6.2.8 *Soit toujours (X_n) et (Y_n) des suites de v.a.r. Si on a les convergences :*

$$\begin{aligned} X_n &\xrightarrow{P} X \\ \text{et} \quad Y_n &\xrightarrow{P} Y \end{aligned}$$

alors

$$X_n + \lambda Y_n \xrightarrow{P} X + \lambda Y,$$

pour tout λ dans \mathbb{R} , et

$$X_n \cdot Y_n \xrightarrow{P} X \cdot Y$$

Preuve. Immédiate. \square

6.2.2 Convergence en probabilité et convergence en loi

Théorème 6.2.9 *La convergence en probabilité entraîne la convergence en loi, i.e.*

$$X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{\mathcal{L}} X.$$

Preuve. Admise. □

La réciproque est fausse sauf quand la limite est une constante.

Proposition 6.2.10 *Si (X_n) est une suite de v.a.r. convergeant en loi vers une constante a dans \mathbb{R} , alors elle converge également en probabilité, i.e.*

$$X_n \xrightarrow{\mathcal{L}} a \Rightarrow X_n \xrightarrow{P} a.$$

Preuve. Notons F_n la fonction de répartition de la v.a.r. X_n , pour tout n , et F celle de la variable aléatoire X déterministe égale à a . On a

$$P(X = a) = 1 \quad \text{et} \quad F_X(x) = \mathbb{1}_{[a, +\infty[}(x).$$

Notons que la fonction F est continue sur $\mathbb{R} \setminus \{a\}$ et que, comme X_n converge en loi vers X , on a :

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x),$$

pour tout x différent de a .

Or, pour tout ε strictement positif, on peut écrire :

$$\begin{aligned} P(|X_n - a| < \varepsilon) &= P(-\varepsilon < X_n - a < \varepsilon) \\ &= P(X_n < a + \varepsilon) - P(X_n \leq a - \varepsilon) \\ &\leq P(X_n \leq a + \varepsilon) - P(X_n \leq a - \varepsilon) \\ &= F_n(a + \varepsilon) - F_n(a - \varepsilon). \end{aligned}$$

D'après la convergence de (F_n) vers F sur $\mathbb{R} \setminus \{a\}$, on a :

$$\lim_{n \rightarrow +\infty} P(|X_n - X| < \varepsilon) = 1 - 0 = 1,$$

toujours pour tout ε strictement positif, ce qui achève la démonstration. □

Proposition 6.2.11 *Supposons que l'on ait les convergences :*

$$\begin{aligned} &X_n \xrightarrow{\mathcal{L}} X \\ \text{et} \quad &Y_n \xrightarrow{P} a, \end{aligned}$$

pour a dans \mathbb{R} . Alors on a :

$$\begin{aligned} i) \quad & X_n + Y_n \xrightarrow{\mathcal{L}} X + a \\ ii) \quad & X_n \cdot Y_n \xrightarrow{\mathcal{L}} X \cdot a \\ iii) \quad & \frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{X}{a}, \text{ si } a \neq 0. \end{aligned}$$

Preuve (idée). Le plus dur est de montrer la convergence conjointe

$$(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, a).$$

La suite est alors une simple utilisation du théorème de Slutsky. \square

6.3 Convergence presque sûre

6.3.1 Définition

Définition 6.3.1 On dit que la suite (X_n) de v.a.r. converge presque sûrement vers X s'il existe un élément A de la tribu \mathcal{A} tel que $P(A) = 1$ et

$$\forall \omega \in A : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega).$$

On note

$$X_n \xrightarrow{p.s.} X.$$

6.3.2 Critères de convergence p.s.

Théorème 6.3.2 La suite de v.a.r. (X_n) converge presque sûrement vers X si la suite de v.a.r. (Y_m) définie par :

$$Y_m = \sup_{n \geq m} |X_n - X|$$

converge en probabilité vers 0.

Proposition 6.3.3 Si, pour tout ε strictement positif, la série de terme général $P[|X_n| > \varepsilon]$ est convergente, i.e.

$$\forall \varepsilon > 0, \quad \sum_n P[|X_n| > \varepsilon] < +\infty,$$

alors (X_n) converge presque sûrement vers zéro.

6.3.3 Convergence presque sûre et convergence en probabilité

Théorème 6.3.4 *La convergence presque sûre entraîne celle en probabilité.*

Preuve. Supposons que (X_n) converge presque sûrement vers la v.a.r. X . D'après le théorème précédent, on a, pour

$$Y_m = \sup_{n \geq m} |X_n - X|,$$

la convergence

$$Y_m \xrightarrow{P} 0,$$

quand m tend vers $+\infty$. Or, à l'aide de l'inclusion

$$\{|X_m - X| > \varepsilon\} \subset \left\{ \sup_{n \geq m} |X_n - X| > \varepsilon \right\} = \{Y_m > \varepsilon\},$$

on peut en déduire que

$$P(|X_m - X| > \varepsilon) \leq P(Y_m > \varepsilon) \xrightarrow{m \rightarrow +\infty} 0,$$

ce qui achève la démonstration. \square

On admet enfin le résultat suivant.

Proposition 6.3.5 *Si la suite (X_n) converge en probabilité vers X , il existe une sous suite $(X_{n_k})_k$ qui converge presque sûrement vers X .*

6.3.4 Loi forte des grands nombres

Théorème 6.3.6 *Soit (X_n) une suite de v.a.r. indépendantes, de même loi et dans L^1 . Notons μ l'espérance de ces v.a.r. On a alors*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} \mu.$$

Notons que l'on peut obtenir le même résultat sans qu'il soit nécessaire que les X_n aient même loi pour tout n .

Théorème 6.3.7 *Soit (X_n) une suite de v.a.r. indépendantes et dans L^2 . Si*

$$\lim_{n \rightarrow +\infty} \mathbb{E}X_n = \mu$$

et si

$$\sum_{n=1}^{+\infty} \frac{\text{Var } X_n}{n^2} < +\infty,$$

alors

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} \mu.$$

6.4 Convergence dans L^p

Définition 6.4.1 Soit (X_n) une suite de v.a.r. dans L^p . On dit qu'elle converge dans L^p vers une v.a.r. X si

$$\|X_n - X\|_p \xrightarrow{n \rightarrow +\infty} 0.$$

Proposition 6.4.2 Soit p et q des réels tels que : $1 \leq p < q$. Si (X_n) converge dans L^q vers X , alors la convergence a également lieu dans L^p .

Preuve. Immédiate en utilisant l'inégalité vue au chapitre III

$$\|X_n - X\|_p \leq \|X_n - X\|_q$$

□

Le corollaire suivant est alors évident.

Corollaire 6.4.3 Si on a :

$$X_n \xrightarrow{L^2} X$$

alors

$$X_n \xrightarrow{L^1} X.$$

Proposition 6.4.4 La convergence dans L^1 entraîne celle en probabilité.

Preuve. Remarquons que l'on a :

$$\begin{aligned} \|X_n - X\|_{L^1} &= \mathbb{E} |X_n - X| \\ &= \int_{|X_n - X| > \varepsilon} |X_n - X| dP + \int_{|X_n - X| \leq \varepsilon} |X_n - X| dP \\ &\geq \int_{|X_n - X| > \varepsilon} |X_n - X| dP \geq \varepsilon P(|X_n - X| > \varepsilon). \end{aligned}$$

La convergence de (X_n) vers X dans L^1 entraîne alors que, pour tout ε strictement positif, on a :

$$P(|X_n - X| > \varepsilon) \xrightarrow{n \rightarrow +\infty} 0,$$

ce qui est bien le résultat annoncé.

□

Proposition 6.4.5 *Soit (X_n) une suite de v.a.r. dans L^2 . Sous les hypothèses :*

$$\lim_{n \rightarrow +\infty} \mathbb{E}X_n = \mu$$

et

$$\lim_{n \rightarrow +\infty} \text{Var } X_n = 0,$$

on a la convergence de (X_n) vers μ dans L^2 .

Preuve. Il suffit de remarquer que l'on peut écrire :

$$\begin{aligned} \mathbb{E}(X_n - \mu)^2 &= \mathbb{E}(X_n - \mathbb{E}X_n + \mathbb{E}X_n - \mu)^2 \\ &= \text{Var } X_n + (\mathbb{E}X_n - \mu)^2, \end{aligned}$$

qui, par hypothèse, converge vers 0, quand n tend vers $+\infty$. \square

Remarquons que lorsqu'on a montré que ces hypothèses suffisaient à entraîner la convergence en probabilité, on avait en fait montré la convergence dans L^2 ce qui, on vient de le voir, entraîne celle en probabilité.

Théorème 6.4.6 (Loi des grands nombres dans L^2)

Soit (X_n) une suite de v.a.r. dans L^2 , de même loi et non corrélées de moyenne μ et variance σ^2 . On a alors :

$$\overline{X}_n \xrightarrow{L^2} \mu.$$

Preuve. D'une part on a, pour tout n :

$$\mathbb{E}\overline{X}_n = \mu$$

et

$$\text{Var}\overline{X}_n = \frac{1}{n^2} \text{Var} \sum_{i=1}^n X_i = \frac{n}{n^2} \text{Var } X = \frac{\sigma^2}{n}$$

qui converge vers 0 quand n tend vers $+\infty$. La proposition précédente permet alors de conclure. \square

6.5 Résumé

$$\begin{array}{c} L^q \Rightarrow_{q \geq p \geq 2} L^p \Rightarrow L^2 \Rightarrow L^1 \\ \searrow \quad \nearrow \\ \quad p \Rightarrow \mathcal{L} \\ \text{p.s.} \end{array}$$

Index

- Absolue continuité
 - d'un vecteur aléatoire, 48
 - d'une mesure par rapport à une autre, 66
 - d'une v.a.r., 39
- Aléatoire
 - (variable), 19
 - (vecteur), 19
- Bayes (formule de), 21
- Beppo-Lévi (théorème de), 65
- Bienaymé-Tchebychev (inégalité de), 79
- Centile, 35
- Changement de variables (formule du)
 - multidimensionnel, 52
 - unidimensionnel, 45
- Coefficient de corrélation, 83
- Convergence
 - dans L^p , 114
 - dominée (théorème de la), 65
 - en loi, 102
 - en probabilité
 - de v.a.r., 108
 - de vecteurs aléatoires, 109
 - monotone (théorème de la), 65
 - presque sûre, 112
- Covariance
 - (matrice de), 83
 - de deux v.a.r., 81
- Cramer-Wold (théorème de), 104
- Densité, 39
 - marginale, 49
 - conditionnelle, 50–51, 69
 - conjointe, 48, 69
 - d'une mesure p/r à une autre, 66
 - marginale, 69
- Décile, 35
- Ecart-type, 78
- Espace
 - L^1 , 75
 - L^2 , 77
 - L^p , 75
 - fondamental, 8
 - mesurable, 10
 - probabilisable, 10
 - probabilisé, 13
 - probabilisé produit, 29
 - produit, 27
- Espérance conditionnelle, 84
- Espérance mathématique
 - d'un vecteur aléatoire, 80
 - d'une v.a.r., 72
- Fonction
 - Bêta, 43
 - caractéristique, 90
 - de répartition, 32
 - de répartition conjointe, 47
 - de répartition marginale, 47
 - Gamma, 42
 - génératrice des moments, 86

- intégrable p/r à une mesure, 60
 - mesurable, 18
 - étagée, 57
- Formule
 - de Bayes, 21
 - du changement de variables, 45, 52
- Gaussien (*vecteur*), 96
- Hölder (*inégalité de*), 76
- Indépendance
 - de deux événements, 22
 - de tribus, 25
 - de variables aléatoires, 25, 53
 - mutuelle d'événements, 24
- Intégrale
 - complexe, 89
 - de Lebesgue, 63
 - p/r à la mesure de Dirac, 61
 - p/r à la mesure de Lebesgue, 63
 - p/r à une mesure, 57
 - p/r à une mesure discrète, 62
 - par rapport à une mesure, 65
- Inégalité
 - de Bienaymé-Tchebychev, 79
 - de Hölder, 76
 - de Markov, 79
 - de Minkowski, 77
- Jacobien, 52
- Jensen (*inégalité de*), 74
- Laplace (*transformée de*), 86, 89
- Lebesgue (*théorème de*), 65
- Limite centrale (*théorème de la*), 105
- Loi, 19
 - absolument continue, 39, 48, 66
 - binomiale, 36
 - binomiale négative, 37
- Bêta, 43
 - conditionnelle, 50–51
 - conjointe, 30
 - continue, 39
 - de Bernoulli, 36
 - de Cauchy, 44
 - de Dirac, 35
 - de Fisher, 43
 - de la somme de deux v.a.r., 56
 - de Poisson, 38
 - de probabilité, 19
 - de Student, 43
 - discrète, 35
 - du χ^2 , 42
 - exponentielle, 42
 - faible des grands nombres, 109
 - forte des grands nombres, 113, 115
 - gamma, 42
 - géométrique, 37
 - hypergéométrique, 38
 - log-normale, 43
 - marginale, 30
 - multinomiale, 48
 - normale, 41
 - uniforme, 41
 - uniforme discrète, 36
- Lévy (*théorème de*), 103
- Markov (*inégalité de*), 79
- Matrice de covariance, 83
- Mesurable
 - (*espace*), 10
 - (*fonction*), 18
 - (*pavé*), 27
- Mesure, 13
 - concentrée sur un ensemble, 15
 - continue, 15
 - de comptage, 17
 - de Dirac, 15

- de Lebesgue, 17
- discrète, 15
- Minkowski (*inégalité de*), 77
- Moment
 - absolu d'ordre p , 76
 - centré d'ordre p , 76
 - d'ordre p , 76
- Médiane, 34
- Norme
 - L^1 , 75
 - L^p , 76
- Négligeabilité, 64
- Presque (*propriété vraie*)
 - partout, 64
 - sûrement, 72
- Probabilité, 13
 - conditionnelle, 20
 - discrète, 16
 - image, 19
 - produit, 28
- Produit de convolution, 56
- Produit scalaire dans L^2 , 77
- Quantile, 34
- Quartile, 35
- Slutsky (*théorème de*), 102, 109
- Tchebychev (*inégalité de*), 79
- Transformée de Laplace
 - d'un vecteur, 89
 - d'une v.a.r., 86
- Transport (*théorème du*), 72
- Tribu, 9
 - borélienne, 13
 - engendrée, 12
 - grossière, 10
 - produit, 27
 - trace, 11
- Variable aléatoire, 19
 - absolument continue, 39, 66
 - complexe, 89
 - continue, 39
 - discrète, 19
 - marginale, 47
 - réelle, 19
- Variance, 78
- Vecteur gaussien, 96