

Probabilités

Table des matières

1. Cadre général de la théorie des probabilités	2
1.1. Espace probabilisé général	2
1.1.1. Exemples d'espaces probabilisés	4
1.1.1.1. Univers $\Omega = \mathbb{N}$	4
1.1.1.2. Univers $\Omega = \mathbb{R}$	4
1.1.1.3. Univers $\Omega = \mathbb{R}^d$	5
1.1.2. Classe monotone	5
1.2. Variables et vecteurs aléatoires	6
1.2.1. Loi d'un vecteur aléatoire	7
1.3. Fonction de répartition	7
1.3.1. Reconnaître une densité de probabilité	8
1.3.2. Reconnaître une loi discrète	8
2. Espérance	9
2.1. Calculs de l'espérance	9
2.1.1. Définition et formule de transfert	9
2.1.2. Variance	9
2.1.3. Covariance	9
2.1.4. Concentration	10
2.2. Application au calcul de lois	11
2.2.1. Méthode de la fonction muette	11
3. Indépendance	12
3.1. Vecteurs aléatoires indépendants	12
3.1.1. Critère d'indépendance pour des vecteurs discrets	13
3.1.2. Critère d'indépendance pour des vecteurs à densité	13
3.2. Somme de variables aléatoires indépendantes	14
3.2.1. Cas de variables aléatoires discrètes	14
3.2.2. Cas de variables aléatoires à densité	14
3.3. Familles d'événements et de variables indépendantes	14
4. Convergence de suites de variables aléatoires	15
4.1. Convergence en loi	15
4.2. Convergence en probabilité	17
4.2.1. Stabilité de la convergence en probabilité	17

1. Cadre général de la théorie des probabilités

1.1. Espace probabilisé général

Définition 1.1. Soit Ω un ensemble. On appelle *tribu* sur Ω une famille \mathcal{F} de parties de Ω vérifiant :

- (1) \mathcal{F} est non-vidé : $\emptyset \in \mathcal{F}$,
- (2) la stabilité par passage au complémentaire : $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$,
- (3) la stabilité par union dénombrable : $\forall (A_n)_{n \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}, \bigcup_{n \geq 1} A_n \in \mathcal{F}$.

On dit que le couple (Ω, \mathcal{F}) est un *espace probabilisable* où Ω est l'univers et \mathcal{F} sont les événements.

Définition 1.2. Soit (Ω, \mathcal{F}) un espace probabilisable. On appelle *mesure de probabilité* sur (Ω, \mathcal{F}) une mesure $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}_+$ vérifiant $\mathbb{P}(\Omega) = 1$.

On dit que le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est un *espace probabilisé*.

Remarque 1.3. Dans le cadre discret, on avait souvent $\mathcal{F} := \mathcal{P}(\Omega)$. Dans le cadre général, on aura souvent $\mathcal{F} \subsetneq \mathcal{P}(\Omega)$.

Définition 1.4. Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements sur $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que $(A_n)_{n \in \mathbb{N}}$ est un *système complet* si elle vérifie :

- (1) les A_n sont disjoints deux à deux : $\forall i, j \in \mathbb{N}, i \neq j \Rightarrow A_i \cap A_j = \emptyset$,
- (2) la probabilité de l'union des A_n est 1 : $\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = 1$.

Proposition 1.5. Soit $(A_n)_{n \in \mathbb{N}}$ un système complet sur $(\Omega, \mathcal{F}, \mathbb{P})$. Alors on a

$$\forall B \in \mathcal{F}, \mathbb{P}(B) = \sum_{n=1}^{+\infty} \mathbb{P}(B \cap A_n).$$

Démonstration. On pose $C := \bigcup_{n \geq 1} A_n$, puisque $\mathbb{P}(C) = 1$, on a $\mathbb{P}(C^c) = 0$ d'où $\mathbb{P}(B \cap C^c) = 0$. Soit $B \in \mathcal{F}$, on en déduit

$$\mathbb{P}(B) = \mathbb{P}(B \cap C) + \underbrace{\mathbb{P}(B \cap C^c)}_{=0} = \mathbb{P}\left(\bigcup_{n \geq 1} B \cap A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(B \cap A_n).$$

□

Corollaire 1.6. Soit $(A_n)_{n \in \mathbb{N}}$ un système complet sur $(\Omega, \mathcal{F}, \mathbb{P})$. Alors pour tout $B \in \mathcal{F}$ on a

- (1) $\mathbb{P}(B) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n) \mathbb{P}(B|A_n)$,
- (2) $\forall i \geq 1, \mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i) \mathbb{P}(B|A_i)}{\sum_{n=1}^{+\infty} \mathbb{P}(A_n) \mathbb{P}(B|A_n)}$.

Théorème 1.7. (Continuité de la mesure de probabilité) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

(1) Soit $(A_n)_{n \in \mathbb{N}}$ une suite croissante d'événements. Alors on a

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n \geq 1} A_n\right).$$

(2) Soit $(A_n)_{n \in \mathbb{N}}$ une suite décroissante d'événements. Alors on a

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n \geq 1} A_n\right).$$

Démonstration.

(1) Pour tout $n \geq 1$, on pose $B_n := A_n \setminus A_{n-1}$ avec $A_0 = \emptyset$, tel que les $(B_n)_{n \in \mathbb{N}}$ forment un système complet sur $\bigcup_{n \geq 1} A_n$, on en déduit alors

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \mathbb{P}\left(\bigcup_{n \geq 1} B_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(B_n) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n) - \mathbb{P}(A_{n-1})$$

on reconnaît une somme télescopique et on a donc

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n) - \mathbb{P}(A_0) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

(2) On obtient directement le résultat par passage au complémentaire.

□

Définition 1.8. Soit $(A)_{n \in \mathbb{N}}$ une suite d'événements de $(\Omega, \mathcal{F}, \mathbb{P})$.

- On appelle *limite supérieure* de la suite $(A)_{n \in \mathbb{N}}$ la valeur

$$\limsup_{n \rightarrow +\infty} A_n := \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$$

intuitivement on considère les éléments qui appartiennent à une infinité d'événements.

- On appelle *limite inférieure* de la suite $(A)_{n \in \mathbb{N}}$ la valeur

$$\liminf_{n \rightarrow +\infty} A_n := \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k.$$

Corollaire 1.9. Soit $(A)_{n \in \mathbb{N}}$ une suite d'événements de $(\Omega, \mathcal{F}, \mathbb{P})$. Alors on a

$$\begin{aligned} \mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) &= \lim_{m \rightarrow +\infty} \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=m}^n A_k\right) \\ \mathbb{P}\left(\liminf_{n \rightarrow +\infty} A_n\right) &= \lim_{m \rightarrow +\infty} \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcap_{k=m}^n A_k\right) \end{aligned}$$

Proposition 1.10. Soit $(A)_{n \in \mathbb{N}}$ une suite d'événements de $(\Omega, \mathcal{F}, \mathbb{P})$. Alors on a

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

Démonstration. On sait que le résultat est vérifié pour un nombre fini d'événements. Par passage à la limite et par continuité de la mesure \mathbb{P} on a

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \lim_{m \rightarrow +\infty} \mathbb{P}\left(\bigcup_{n=1}^m A_n\right) \leq \lim_{m \rightarrow +\infty} \sum_{n=1}^m \mathbb{P}(A_n) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

□

Définition 1.11. Soit A un événement de $(\Omega, \mathcal{F}, \mathbb{P})$.

- On dit que A est *négligeable* si $\mathbb{P}(A) = 0$.
- On dit que A est *presque-sûr* si $\mathbb{P}(A) = 1$.

Corollaire 1.12. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Alors

- L'union dénombrable d'événements négligeables est négligeable.
- L'intersection dénombrable d'événements presque-sûrs est presque-sûre.

Proposition 1.13. Soit \mathcal{A} une famille d'événements de $(\Omega, \mathcal{F}, \mathbb{P})$. Alors il existe une unique tribu $\sigma(\mathcal{A})$ telle que $\sigma(\mathcal{A})$ soit la plus petite tribu contenant \mathcal{A} .

Démonstration. Il existe au moins une tribu contenant \mathcal{A} , à savoir $\mathcal{P}(\Omega)$. Alors l'intersection de toutes les tribus contenant \mathcal{A} est une tribu et convient. \square

Définition 1.14. Soit \mathcal{A} une famille d'événements de $(\Omega, \mathcal{F}, \mathbb{P})$. On appelle *tribu engendrée* par \mathcal{A} , notée $\sigma(\mathcal{A})$, la tribu de la Proposition 1.13.

Exemple 1.15. Soit A un événement de $(\Omega, \mathcal{F}, \mathbb{P})$. Alors $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$.

1.1.1. Exemples d'espaces probabilisés

Définition 1.16. Soit (E, \mathcal{O}) un espace topologique. On appelle *tribu borélienne* sur E , notée $\mathcal{B}(E)$, la tribu engendrée par les intervalles ouverts de E , c'est-à-dire $\mathcal{B}(E) := \sigma(\mathcal{O})$.

Lemme 1.17. Soit $(\mu)_{n \in \mathbb{N}}$ une suite de mesures de probabilité sur (Ω, \mathcal{F}) et $(\lambda)_{n \in \mathbb{N}}$ une suite de nombres réels positifs telle que $\sum_{n=1}^{+\infty} \lambda_n = 1$. Alors $\mu = \sum_{n=1}^{+\infty} \lambda_n \mu_n$ est une mesure de probabilité sur (Ω, \mathcal{F}) .

1.1.1.1. Univers $\Omega = \mathbb{N}$

Se référer au cours de *Probabilités* de deuxième année.

1.1.1.2. Univers $\Omega = \mathbb{R}$

Exemple 1.18. (Mesure de Dirac) Soit $x \in \mathbb{R}$, l'application $\delta_x : \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{R}_+$ définie par

$$\forall A \in \mathcal{B}(\mathbb{R}), \delta_x(A) = \begin{cases} 0 & \text{si } x \notin A \\ 1 & \text{si } x \in A \end{cases}$$

est une mesure de probabilité sur \mathbb{R} .

Exemple 1.19. (Mesure uniforme sur $\{1, \dots, n\}$) L'application $\mu = \frac{1}{n} \sum_{k=1}^n \delta_k$ est une mesure uniforme sur \mathbb{R} .

Exemple 1.20. (Mesure de Poisson) Soit $\lambda > 0$, l'application $\mu = \sum_{n=1}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} \delta_n$ est une mesure de Poisson sur \mathbb{R} .

Définition 1.21. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction borélienne. On dit que f est une *densité de probabilité* sur \mathbb{R} si elle vérifie :

- (1) pour λ -presque tout $x \in \mathbb{R}$, $f(x) \geq 0$,
- (2) $\int_{\mathbb{R}} f(x) d\lambda(x) = 1$.

Lemme 1.22. Soit f une densité de probabilité sur \mathbb{R} . Alors l'application $\mu_f : \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{R}_+$ définie par $\forall A \in \mathcal{B}(\mathbb{R}), \mu_f(A) = \int_A f(x) d\lambda(x)$ est une mesure de probabilité sur \mathbb{R} .

Démonstration. On a bien $\forall A \in \mathcal{B}(\mathbb{R}), \mu_f(A) \geq 0$. De plus $\mu_f(\mathbb{R}) = 1$. Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'éléments de $\mathcal{B}(\mathbb{R})$ deux à deux disjoints. On pose $A := \bigcup_{n \geq 1} A_n$, alors $\mathbb{1}_A = \sum_{n=1}^{+\infty} \mathbb{1}_{A_n}$ et

$$\mu_f(A) = \int_A f(x) d\lambda(x) = \int_{\mathbb{R}} \mathbb{1}_A(x) f(x) d\lambda(x) = \int_{\mathbb{R}} \sum_{n=1}^{+\infty} \mathbb{1}_{A_n}(x) f(x) d\lambda(x)$$

d'après le théorème de convergence monotone on a

$$\mu_f(A) = \lim_{m \rightarrow +\infty} \int_{\mathbb{R}} \sum_{n=1}^m \mathbb{1}_{A_n}(x) f(x) d\lambda(x) = \lim_{m \rightarrow +\infty} \sum_{n=1}^m \mu_f(A_n) = \sum_{n=1}^{+\infty} \mu_f(A_n).$$

Donc μ_f est bien une mesure de probabilité sur \mathbb{R} . \square

Remarque 1.23. On dit que μ_f est la *mesure de densité* f .

Proposition 1.24. Soit f et g deux densités de probabilités sur \mathbb{R} . Alors les mesures de densité μ_f et μ_g sont égales si et seulement si f et g sont égales presque partout.

Démonstration.

\Rightarrow : Supposons que $\mu_f = \mu_g$. On pose

$$A_+ := \{x \in \mathbb{R} \mid f(x) > g(x)\}$$

$$A_- := \{x \in \mathbb{R} \mid f(x) < g(x)\}$$

ces deux ensembles sont boréliens car f et g sont boréliennes. Par construction

$$\int_{A_+} f - g \, d\lambda = \mu_f(A_+) - \mu_g(A_+) = 0 = \int_{A_-} f - g \, d\lambda$$

de plus $A := \{x \in \mathbb{R} \mid |f(x) - g(x)| > 0\} = A_+ \cup A_-$, on en déduit

$$\int_A |f - g| \, d\lambda = \int_A (f - g) \mathbb{1}_{A_+} + (g - f) \mathbb{1}_{A_-} \, d\lambda = 0$$

donc $f - g = 0$ presque partout et $f = g$ presque partout.

\Leftarrow : Si $f = g$ presque partout, alors il est évident que $\mu_f = \mu_g$. □

Exemple 1.25. (Loi uniforme) Soit $c, d \in \mathbb{R}$ avec $c < d$. Alors la fonction $f : \mathbb{R} \rightarrow \mathbb{R}; x \mapsto \frac{\mathbb{1}_{[c,d]}(x)}{d-c}$ est une densité de probabilité. En particulier, pour tout $[a, b] \subset [c, d]$

$$\mu_f([a, b]) = \int_{[a,b]} f(x) \, d\lambda(x) = \frac{b-a}{d-c}.$$

On note la probabilité associée $\mathcal{U}([c, d])$.

Exemple 1.26. (Loi exponentielle) Soit $\lambda > 0$. Alors la fonction $f : \mathbb{R} \rightarrow \mathbb{R}; x \mapsto \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x)$ est une densité de probabilité. On note la probabilité associée $\mathcal{E}(\lambda)$.

Exemple 1.27. (Loi normale) La fonction $f : \mathbb{R} \rightarrow \mathbb{R}; x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ est une densité de probabilité. On note la probabilité associée $\mathcal{N}(0, 1)$.

1.1.1.3. Univers $\Omega = \mathbb{R}^d$

On peut étendre les exemples de \mathbb{R} , ainsi que les définitions de densité et de mesures de probabilité associée.

1.1.2. Classe monotone

Définition 1.28. Soit \mathcal{C} une famille de parties d'un ensemble Ω . On dit que \mathcal{C} est une *classe monotone* si elle vérifie :

- (1) $\Omega \in \mathcal{C}$,
- (2) $\forall A, B \in \mathcal{C}, A \subset B \Rightarrow B \setminus A \in \mathcal{C}$,
- (3) $\forall (A_n)_{n \in \mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$ croissante, $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{C}$.

Remarque 1.29. Une tribu est une classe monotone, la réciproque est fausse.

Lemme 1.30. Soit \mathcal{C} une classe monotone. Alors \mathcal{C} est une tribu si et seulement si elle est stable par intersection finie, c'est-à-dire :

$$\forall A_1, \dots, A_n \in \mathcal{C}, \bigcap_{k=1}^n A_k \in \mathcal{C}.$$

Démonstration.

\Rightarrow : Si \mathcal{C} est une tribu elle est stable par intersection finie.

\Leftarrow : Supposons que \mathcal{C} est stable par intersection finie. Soit $(A)_{n \in \mathbb{N}}$ une suite d'éléments de \mathcal{C} . Puisque \mathcal{C} est stable par passage au complémentaire, \mathcal{C} est aussi stable par union finie, en effet

$$A, B \in \mathcal{C} \Rightarrow A^c, B^c \in \mathcal{C} \Rightarrow A^c \cap B^c \in \mathcal{C} \Rightarrow A \cup B = (A^c \cap B^c)^c \in \mathcal{C}$$

on a donc pour tout $N \in \mathbb{N}$, $\bigcup_{n=0}^N A_n \in \mathcal{C}$, et par union croissante

$$\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{N \in \mathbb{N}} \underbrace{\bigcup_{n=0}^N A_n}_{\text{croissante}} \in \mathcal{C}$$

donc \mathcal{C} est bien une tribu. □

Définition 1.31. Soit \mathcal{A} une famille de parties d'un ensemble Ω . On appelle *classe monotone engendrée* par \mathcal{A} , notée $\mathcal{C}(\mathcal{A})$, l'intersection de toutes les classes monotones contenant \mathcal{A} .

Théorème 1.32. (Théorème de la classe monotone) Soit \mathcal{A} une famille de parties d'un ensemble Ω . Si \mathcal{A} est stable par intersection finie, alors $\mathcal{C}(\mathcal{A}) = \sigma(\mathcal{A})$.

Démonstration. Soit $A \in \mathcal{C}(\mathcal{A})$, on pose $\mathcal{C}_A := \{B \in \mathcal{C}(\mathcal{A}) \mid A \cap B \in \mathcal{C}(\mathcal{A})\}$. Puisque \mathcal{C}_A est une classe monotone contenant A , on a $\mathcal{C}_A = \mathcal{C}(\mathcal{A})$. Donc $\mathcal{C}(\mathcal{A})$ est stable par intersection finie. D'après le **Lemme 1.30** $\mathcal{C}(\mathcal{A})$ est une tribu. □

Corollaire 1.33. Soit μ et ν deux mesures de probabilités sur (Ω, \mathcal{F}) . S'il existe une famille de parties \mathcal{A} stable par intersection finie sur laquelle μ et ν coïncident, alors elles coïncident sur $\sigma(\mathcal{A})$.

1.2. Variables et vecteurs aléatoires

Définition 1.34. Soit (Ω, \mathcal{F}) un espace probabilisable. On appelle *vecteur aléatoire* une application borélienne $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Dans le cas $d = 1$, on dit que X est une *variable aléatoire*.

Proposition 1.35. Soit (Ω, \mathcal{F}) un espace probabilisable.

(1) Une application $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire si et seulement si

$$\forall t \in \mathbb{R}, X^{-1}([-\infty, t]) \in \mathcal{F}$$

(2) Une application $X = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ est un vecteur aléatoire si et seulement si X_1, \dots, X_d sont des variables aléatoires.

(3) Soit $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire et $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^n$ une application borélienne. Alors $\varphi \circ X$ est un vecteur aléatoire.

Démonstration.

(1) \Rightarrow : Si X est une variable aléatoire, alors X est mesurable et le résultat est évident.

\Leftarrow : Si pour tout $t \in \mathbb{R}$ on a $X^{-1}([-\infty, t]) \in \mathcal{F}$. Alors puisque la famille $\{[-\infty, t] \mid t \in \mathbb{R}\}$ engendre $\mathcal{B}(\mathbb{R})$, X est mesurable. Donc X est une variable aléatoire.

(2) On obtient le résultat par projection en appliquant (1) à X_1, \dots, X_n .

(3) On obtient le résultat par composition de fonctions boréliennes. □

Proposition 1.36. Soit $(X)_{n \in \mathbb{N}}$ une suite de variables aléatoires sur (Ω, \mathcal{F}) .

(1) Si les applications $S := \sup_{n \in \mathbb{N}} X_n$ et $I := \inf_{n \in \mathbb{N}} X_n$ sont finies, alors S et I sont des variables aléatoires.

(2) Si $(X)_{n \in \mathbb{N}}$ converge simplement vers une limite finie X , alors X est une variable aléatoire.

Démonstration.

- (1) On remarque que pour tout $t \in \mathbb{R}$ on a $S^{-1}(]-\infty, t]) = \bigcap_{n \in \mathbb{N}} X^{-1}(]-\infty, t])$ et que l'on peut écrire de la même manière pour I .
- (2) On remarque que $X = \lim_{n \rightarrow +\infty} X_n = \limsup_{n \rightarrow +\infty} X_n = \inf_{m \rightarrow +\infty} (\sup_{n \geq m} X_n)$.

□

1.2.1. Loi d'un vecteur aléatoire

Proposition 1.37. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. Alors l'application $\mathbb{P}_X : \mathcal{B}(\mathbb{R}^d) \rightarrow \mathbb{R}_+; A \mapsto \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$ est une mesure de probabilité sur \mathbb{R}^d .

Définition 1.38. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. On appelle *loi de X*, notée \mathbb{P}_X , la mesure de probabilité de la Proposition 1.37. On dit aussi que X suit la loi \mathbb{P}_X .

Définition 1.39. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. On appelle *atomes de X*, noté \mathcal{V}_X , l'ensemble

$$\mathcal{V}_X := \{x \in \mathbb{R}^d \mid \mathbb{P}_X(\{x\}) > 0\}.$$

Exemple 1.40. (Loi de Bernoulli) On considère $(\Omega, \mathcal{F}, \mathbb{P})$ avec $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ et \mathbb{P} la mesure uniforme sur $[0, 1]$. On prend $X = \mathbb{1}_{[0, p]}$ avec $p \in [0, 1]$. Soit $A \in \mathcal{B}(\mathbb{R})$, alors

$$\begin{aligned} \mathbb{P}_X(A) &= \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X^{-1}(A \cap \{0\})) + \mathbb{P}(X^{-1}(A \cap \{1\})) \\ &= \delta_0(A)\mathbb{P}(X^{-1}(0)) + \delta_1(A)\mathbb{P}(X^{-1}(1)) = \delta_0(A)(1 - p) + \delta_1(A)p \end{aligned}$$

donc $\mathbb{P}_X = \delta_0(1 - p) + \delta_1 p$.

Proposition 1.41. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. Si X admet une densité $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$, alors les variables aléatoires X_1, \dots, X_d admettent des densités $f_1, \dots, f_d : \mathbb{R} \rightarrow \mathbb{R}_+$ avec

$$\forall i \in \{1, \dots, d\}, f_i(x) := \int_{\mathbb{R}^{d-1}} f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) d\lambda(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d).$$

Démonstration. Il suffit d'appliquer le théorème de Fubini. □

1.3. Fonction de répartition

Définition 1.42. Soit \mathbb{P} une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On appelle *fonction de répartition*, notée $F_{\mathbb{P}}$, la fonction $F_{\mathbb{P}} : \mathbb{R} \rightarrow \mathbb{R}_+; t \mapsto \mathbb{P}(]-\infty, t])$.

Définition 1.43. Soit $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ un espace probabilisé, et X une variable aléatoire. On appelle *fonction de répartition de X*, notée F_X , la fonction de répartition liée à \mathbb{P}_X .

Proposition 1.44. Soit $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ un espace probabilisé, et $X, Y : \mathbb{R} \rightarrow \mathbb{R}$ deux variables aléatoires. Alors X et Y ont la même loi si et seulement si elles ont la même fonction de répartition.

Démonstration.

\Rightarrow : Supposons que $\mathbb{P}_X = \mathbb{P}_Y$. Alors on a

$$\forall t \in \mathbb{R}, F_X(t) = \mathbb{P}_X(]-\infty, t]) = \mathbb{P}_Y(]-\infty, t]) = F_Y(t)$$

donc $F_X = F_Y$.

\Leftarrow : Supposons que $F_X = F_Y$. Alors on pose $\mathcal{A} := \{]-\infty, t] \mid t \in \mathbb{R}\}$ qui est stable par intersection avec $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$, on pose $\mathcal{C} := \{A \in \mathcal{B}(\mathbb{R}) \mid \mathbb{P}_X(A) = \mathbb{P}_Y(A)\}$ qui est une classe monotone, alors d'après le théorème de la classe monotone $\mathcal{C} = \mathcal{B}(\mathbb{R})$. Donc $\mathbb{P}_X = \mathbb{P}_Y$ □

1.3.1. Reconnaître une densité de probabilité

Proposition 1.45. Soit $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ un espace probabilisé et $X : \mathbb{R} \rightarrow \mathbb{R}$ une variable aléatoire. Alors si la fonction de répartition de X est C^1 par morceaux, X admet une densité de probabilité définie pour tout $x \in \mathbb{R}$ par $f(x) = F'_X(x)$ si F_X est dérivable en x et $f(x) = 0$ sinon.

Démonstration. Puisque F_X est C^k par morceaux, il existe une suite croissante $(a_n)_{n \in \mathbb{Z}}$ telle que

$$\lim_{n \rightarrow +\infty} a_n = -\lim_{n \rightarrow -\infty} a_n = +\infty$$

et pour tout $n \in \mathbb{Z}$, F_X soit dérivable sur $]a_n, a_{n+1}[$. Soit $n \in \mathbb{Z}$, alors

$$\forall s, t \in]a_n, a_{n+1}[, \int_s^t f(x) dx = F_X(t) - F_X(s)$$

et par passage à la limite pour $s \rightarrow a_n$ et $t \rightarrow a_{n+1}$ on a

$$\int_{a_n}^{a_{n+1}} f(x) dx = F_X(a_n) - F_X(a_{n+1}).$$

Soit $t \in \mathbb{R}$, alors il existe $n \in \mathbb{Z}$ tel que $t \in]a_n, a_{n+1}[$ et

$$\int_{-\infty}^t f(x) dx = \sum_{k=-\infty}^n \int_{a_k}^{a_{k+1}} f(x) dx + \int_{a_n}^t f(x) dx$$

on reconnaît une somme télescopique et on a donc

$$\int_{-\infty}^t f(x) dx = F_X(t) - \underbrace{\lim_{k \rightarrow -\infty} F_X(a_k)}_{=0} = F_X(t) = \mathbb{P}_X([-\infty, t]).$$

□

1.3.2. Reconnaître une loi discrète

Définition 1.46. Soit $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ un espace probabilisé et $X : \mathbb{R} \rightarrow \mathbb{R}$ une variable aléatoire. On appelle *saut* de la fonction de répartition de X , noté Δ_X , la fonction définie par

$$\forall t \in \mathbb{R}, \Delta_X(t) := F_X(t) - \lim_{x \rightarrow t^-} F_X(x).$$

Lemme 1.47. Soit $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ un espace probabilisé et $X : \mathbb{R} \rightarrow \mathbb{R}$ une variable aléatoire. Alors l'ensemble des points de discontinuités, noté $\mathcal{D}_X := \{t \in \mathbb{R} \mid \Delta_X(t) > 0\}$, est dénombrable avec $\sum_{t \in \mathcal{D}_X} \Delta_X(t) \leq 1$ de plus X est discrète si et seulement si $\sum_{t \in \mathcal{D}_X} \Delta_X(t) = 1$.

2. Espérance

2.1. Calculs de l'espérance

2.1.1. Définition et formule de transfert

Définition 2.1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}_d$ un vecteur aléatoire. On appelle *espérance* de X , notée $\mathbb{E}[X]$, la valeur

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

Remarque 2.2. Pour que l'intégrale précédente ait du sens dans \mathbb{R} on a besoin que :

- $X \geq 0$ presque sûrement,
- X soit intégrable sur $(\Omega, \mathcal{F}, \mathbb{P})$.

Théorème 2.3. (Formule de transfert) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow \mathbb{R}_d$ un vecteur aléatoire et $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ une application mesurable. Alors

$$\mathbb{E}[\varphi(X)] = \int_{\Omega} \varphi(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}_X(x).$$

Remarque 2.4. Pour que l'intégrale précédente ait du sens on a besoin que :

- $\varphi(X)$ soit intégrable sur $(\Omega, \mathcal{F}, \mathbb{P})$, c'est-à-dire que φ soit intégrable sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X)$.

Proposition 2.5. (Cas discret) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow \mathbb{R}_d$ un vecteur aléatoire et $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ une application mesurable. Alors

$$\mathbb{E}[\varphi(X)] = \sum_{\omega \in \mathcal{V}_X} \varphi(\omega) \mathbb{P}(X = \omega).$$

Proposition 2.6. (Cas à densité) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow \mathbb{R}_d$ un vecteur aléatoire à densité $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ et $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ une application mesurable. Alors

$$\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}^d} \varphi(x) f(x) d\lambda(x).$$

2.1.2. Variance

Définition 2.7. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}_d$ un vecteur aléatoire. On appelle *variance* de X , notée $V(X)$, la valeur

$$V(X) := \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Proposition 2.8. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}_d$ un vecteur aléatoire. Alors la variance de X vérifie les propriétés suivantes :

- (1) $V(X)$ ne dépend que de X .
- (2) $V(X) \geq 0$, avec égalité si et seulement si X est constante.
- (3) $\forall a, b \in \mathbb{R}, V(aX + b) = a^2 V(X)$.
- (4) $V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

2.1.3. Covariance

Définition 2.9. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X, Y : \Omega \rightarrow \mathbb{R}_d$ deux vecteurs aléatoires. On appelle *covariance* de X et X , notée $\text{Cov}(X, Y)$, la valeur

$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Proposition 2.10. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X, Y : \Omega \rightarrow \mathbb{R}_d$ deux vecteurs aléatoires. Alors la covariance vérifie les propriétés suivantes :

- (1) Cov est bilinéaire symétrique.
- (2) $\text{Cov}(X, X) = V(X)$.
- (3) $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.
- (4) $\text{Cov}(X, Y) \leq \sqrt{V(X)V(Y)}$, avec égalité si et seulement si X et Y sont en relation affine.
- (5) $V(X + Y) = V(X) + V(Y) + 2 \text{Cov}(X, Y)$.

2.1.4. Concentration

Définition 2.11. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X, Y : \Omega \rightarrow \mathbb{R}_d$ deux vecteurs aléatoires. Alors si $\text{Cov}(X, Y) = 0$ on dit que X et Y sont *non corrélées*.

Corollaire 2.12. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X, Y : \Omega \rightarrow \mathbb{R}_d$ deux vecteurs aléatoires. Alors si X et Y sont non-corrélées, on a $V(X + Y) = V(X) + V(Y)$.

Définition 2.13. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}_d$ des vecteurs aléatoires. On appelle *moyenne empirique* de X_1, \dots, X_n , notée \bar{X}_n , le vecteur aléatoire

$$\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k.$$

Proposition 2.14. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}_d$ des vecteurs aléatoires. Alors l'espérance de \bar{X}_n est donnée par

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k]$$

et sa variance par

$$V(\bar{X}_n) = \frac{1}{n^2} \sum_{k=1}^n V(X_k)$$

Proposition 2.15. (Inégalité de Markov et de Bienaymé-Chebychev) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire.

- (1) Si $X \geq 0$ presque sûrement, alors on a

$$\forall \varepsilon > 0, \mathbb{P}(X > \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}.$$

- (2) Si X est intégrable, alors on a

$$\forall \varepsilon > 0, \mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

Démonstration.

- (1) Soit $\varepsilon > 0$, on remarque que l'on a toujours l'inégalité

$$\varepsilon \mathbb{1}_{\{X \geq \varepsilon\}} \leq X$$

par passage à l'espérance on trouve

$$\varepsilon \mathbb{E}[\mathbb{1}_{\{X \geq \varepsilon\}}] \leq \mathbb{E}[X]$$

ce qui donne bien l'inégalité de Markov.

- (2) On applique l'inégalité de Markov à $(X - \mathbb{E}[X])^2$.

□

Proposition 2.16. (Inégalité de Jensen) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire intégrable et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe bornée inférieurement. Alors

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

Théorème 2.17. (Inégalité de Hoeffding) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ des variables aléatoires indépendantes de sorte que pour tout $k \in \{1, \dots, n\}$, il existe $a_k, b_k \in \mathbb{R}$ tels que $a_k \leq X_k \leq b_k$ presque sûrement. Si on note $S_n := X_1 + \dots + X_n$, alors

$$\forall t > 0, \max(\mathbb{P}(S_n - \mathbb{E}[S_n] \geq t), \mathbb{P}(S_n - \mathbb{E}[S_n] \leq -t)) < \exp\left(-\frac{t^2}{\sum_{k=1}^n (b_k - a_k)^2}\right).$$

2.2. Application au calcul de lois

2.2.1. Méthode de la fonction muette

Proposition 2.18. (Méthode) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}_d$ un vecteur aléatoire de densité $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$. Alors pour toute fonction borélienne positive

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

en particulier pour tout $A \in \mathcal{F}$ en prenant $h := \mathbb{1}_A$ on trouve

$$\mathbb{P}(X \in A) = \mathbb{E}[\mathbb{1}_A(X)] = \int_A f(x) d\lambda(x).$$

ce qui montre que X est de densité f .

3. Indépendance

3.1. Vecteurs aléatoires indépendants

Définition 3.1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}^{d_i}$ des vecteurs aléatoires. On dit que X_1, \dots, X_n sont *indépendants*, noté $X_1 \perp \dots \perp X_n$, si

$$\forall A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}^{d_1}) \times \dots \times \mathcal{B}(\mathbb{R}^{d_n}), \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i).$$

Lemme 3.2. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et $X : \Omega \rightarrow \mathbb{R}^p$ et $Y : \Omega \rightarrow \mathbb{R}^q$ deux vecteurs aléatoires. Alors les assertions suivantes sont équivalentes :

- (1) X et Y sont indépendants.
- (2) $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$.
- (3) Pour toutes fonctions boréliennes positives g et h , $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$

Démonstration.

(1) \Rightarrow (2) : Soit $A \in \mathcal{B}(\mathbb{R}^p)$ et $B \in \mathcal{B}(\mathbb{R}^q)$. Puisque X et Y sont indépendants on a

$$\begin{aligned} \mathbb{P}_{(X,Y)}(A \times B) &= \mathbb{P}((X, Y) \in A \times B) \\ &= \mathbb{P}(X \in A, Y \in B) \\ &= \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \\ &= \mathbb{P}_X(A)\mathbb{P}_Y(B) = (\mathbb{P}_X \otimes \mathbb{P}_Y)(A \times B) \end{aligned}$$

par unicité de la mesure produit $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$.

(2) \Rightarrow (3) : Soit g et h deux fonctions boréliennes positives. Alors par la formule de transfert, en posant $\varphi : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}_+; (x, y) \mapsto g(x)h(y)$, on a

$$\begin{aligned} \mathbb{E}[\varphi(X, Y)] &= \int_{\mathbb{R}^p \times \mathbb{R}^q} \varphi(x, y) d\mathbb{P}_{(X,Y)}(x, y) \\ &= \int_{\mathbb{R}^p \times \mathbb{R}^q} g(x)h(y) d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x, y) \end{aligned}$$

en appliquant Fubini, on trouve

$$\begin{aligned} \mathbb{E}[\varphi(X, Y)] &= \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} g(x)h(y) d\mathbb{P}_X(x) d\mathbb{P}_Y(y) \\ &= \int_{\mathbb{R}^q} h(y) \int_{\mathbb{R}^p} g(x) d\mathbb{P}_X(x) d\mathbb{P}_Y(y) \\ &= \int_{\mathbb{R}^p} g(x) d\mathbb{P}_X(x) \int_{\mathbb{R}^q} h(y) d\mathbb{P}_Y(y) \\ &= \mathbb{E}[g(X)]\mathbb{E}[h(Y)]. \end{aligned}$$

(3) \Rightarrow (1) : Soit $A \in \mathcal{B}(\mathbb{R}^p)$ et $B \in \mathcal{B}(\mathbb{R}^q)$. Il suffit de prendre $g := \mathbb{1}_A$ et $h := \mathbb{1}_B$ pour obtenir l'indépendance. \square

Proposition 3.3. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

- (1) Soit $X : \Omega \rightarrow \mathbb{R}^p$ et $Y : \Omega \rightarrow \mathbb{R}^q$ deux vecteurs aléatoires indépendants. Alors pour toutes fonctions boréliennes f et g , $f(X)$ et $g(Y)$ sont indépendants.
- (2) Soit $X_1, \dots, X_m : \Omega \rightarrow \mathbb{R}^{d_i}$ des vecteurs aléatoires indépendants. Alors pour tout $1 \leq n < m$, (X_1, \dots, X_n) et (X_{n+1}, \dots, X_m) sont indépendants.

Démonstration.

- (1) Soit f et g deux fonctions boréliennes. Alors il suffit d'appliquer le point (3) du [Lemme 3.2](#) aux compositions de f et g avec des fonctions boréliennes positives pour obtenir l'indépendance de $f(X)$ et $g(Y)$.
- (2) Soit $1 \leq n < m$. Alors il suffit d'appliquer le point (2) du [Lemme 3.2](#) pour obtenir l'indépendance de (X_1, \dots, X_n) et (X_{n+1}, \dots, X_m) .

□

3.1.1. Critère d'indépendance pour des vecteurs discrets

Proposition 3.4. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et $X : \Omega \rightarrow \mathbb{R}^p$ et $Y : \Omega \rightarrow \mathbb{R}^q$ deux vecteurs aléatoires discrets. Alors s'ils existent des fonctions $f : \mathcal{V}_X \rightarrow \mathbb{R}_+$ et $g : \mathcal{V}_Y \rightarrow \mathbb{R}_+$ telles que

$$\forall (x, y) \in \mathcal{V}_X \times \mathcal{V}_Y, \mathbb{P}(X = x, Y = y) = f(x)g(y)$$

alors X et Y sont indépendants, et il existe $c > 0$ tel que

$$\mathbb{P}_X(\{x\}) = cf(x) \text{ et } \mathbb{P}_Y(\{y\}) = \frac{1}{c}g(y).$$

Démonstration. Soit $A \in \mathcal{B}(\mathbb{R}^p)$ et $B \in \mathcal{B}(\mathbb{R}^q)$ alors

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \sum_{x \in \mathcal{V}_X \cap A} \sum_{y \in \mathcal{V}_Y \cap B} \mathbb{P}(X = x, Y = y) \\ &= \sum_{x \in \mathcal{V}_X \cap A} \sum_{y \in \mathcal{V}_Y \cap B} f(x)g(y) \\ &= \sum_{x \in \mathcal{V}_X \cap A} f(x) \sum_{y \in \mathcal{V}_Y \cap B} g(y) \end{aligned}$$

en particulier si on pose $B := \mathbb{R}^q$ et $c := \sum_{y \in \mathcal{V}_Y \cap B} g(y)$, on trouve

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A, Y \in \mathbb{R}^q) = c \sum_{x \in \mathcal{V}_X \cap A} f(x)$$

d'où pour tout $x \in \mathcal{V}_X$, $\mathbb{P}_X(\{x\}) = cf(x)$. On fait la même chose avec $A := \mathbb{R}^p$ et $d := \sum_{x \in \mathcal{V}_X \cap A} f(x)$. Mais $\mathbb{P}(X \in \mathbb{R}^p, Y \in \mathbb{R}^q) = c \times d = 1$, donc $d = \frac{1}{c}$. Enfin

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \sum_{x \in \mathcal{V}_X \cap A} \mathbb{P}(X = x) \sum_{y \in \mathcal{V}_Y \cap B} \mathbb{P}(Y = y) \\ &= \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \end{aligned}$$

donc X et Y sont indépendants.

□

3.1.2. Critère d'indépendance pour des vecteurs à densité

Proposition 3.5. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et $X : \Omega \rightarrow \mathbb{R}^p$ et $Y : \Omega \rightarrow \mathbb{R}^q$ deux vecteurs aléatoires à densités respectives f_X et f_Y .

- (1) Si X et Y sont indépendantes. Alors le vecteur (X, Y) admet une densité f vérifiant :

$$\forall (x, y) \in \mathbb{R}^p \times \mathbb{R}^q, f(x, y) = f_X(x)f_Y(y)$$

- (2) Si (X, Y) admet une densité f de la forme :

$$\forall (x, y) \in \mathbb{R}^p \times \mathbb{R}^q, f(x, y) = g(x)h(y)$$

où g et h sont boréliennes. Alors X et Y sont indépendantes et il existe $c > 0$ tel que

$$f_X = cg \text{ et } f_Y = ch.$$

Démonstration.

- (1) Supposons que X et Y sont indépendantes. Soit $\varphi : \mathbb{R}^{p+q} \rightarrow \mathbb{R}_+$ une fonction borélienne, alors par la formule de transfert

$$\mathbb{E}[\varphi(X, Y)] = \int_{\mathbb{R}^{p+q}} \varphi(x, y) d\mathbb{P}_{(X, Y)}(x, y)$$

puisque X et Y sont indépendantes on a

$$\begin{aligned} \mathbb{E}[\varphi(X, Y)] &= \int_{\mathbb{R}^p \times \mathbb{R}^q} \varphi(x, y) d\mathbb{P}_X(x) \otimes \mathbb{P}_Y(y) \\ &= \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} \varphi(x, y) d\mathbb{P}_X(x) d\mathbb{P}_Y(y) \end{aligned}$$

et puisque X et Y admettent des densités

$$\begin{aligned} \mathbb{E}[\varphi(X, Y)] &= \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} \varphi(x, y) f_X(x) d\lambda(x) f_Y(y) d\lambda(y) \\ &= \int_{\mathbb{R}^{p+q}} \varphi(x, y) f_X(x) f_Y(y) d\lambda(x, y). \end{aligned}$$

Donc (X, Y) admet bien une densité $(x, y) \mapsto f_X(x) f_Y(y)$.

- (2) La réciproque se montre une nouvelle fois en appliquant le théorème de Fubini

□

3.2. Somme de variables aléatoires indépendantes

3.2.1. Cas de variables aléatoires discrètes

Proposition 3.6. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et $X, Y : \Omega \rightarrow \mathbb{N}$ deux variables aléatoires discrètes indépendantes à valeurs entières. On pose $S := X + Y$. Alors on a

$$\forall n \in \mathbb{N}, \mathbb{P}(S = n) = \sum_{k \in \mathbb{N}} \mathbb{P}(X = k) \mathbb{P}(Y = n - k)$$

3.2.2. Cas de variables aléatoires à densité

Proposition 3.7. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et $X, Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires à densités respectives f_X et f_Y . On pose $S := X + Y$. Alors la densité de S est donnée par

$$\forall t \in \mathbb{R}, f(t) := \int_{\mathbb{R}} f_X(x) f_Y(t - x) d\lambda(x).$$

3.3. Familles d'événements et de variables indépendantes

Définition 3.8. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

- (1) Soit $(X_i : \Omega \rightarrow \mathbb{R})_{i \in I}$ une famille de variables aléatoires. On dit que la famille est *indépendante* si les variables de tout sous-ensemble fini sont indépendantes.
- (2) Soit $(A_i)_{i \in I}$ une famille d'événements. On dit que la famille est *indépendante* si la famille de variables aléatoires $(\mathbb{1}_{A_i})_{i \in I}$ est indépendante.

Lemme 3.9. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(A_i)_{i \in I}$ une famille d'événements. Alors la famille est indépendante si et seulement si pour tout sous-ensemble J de I fini, on a

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{i \in J} \mathbb{P}(A_j).$$

Théorème 3.10. (Premier lemme de Borel-Cantelli) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(A_n)_{n \in \mathbb{N}}$ une suite d'événements. Si la série $\sum \mathbb{P}(A_n)$ converge, alors $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 0$.

Démonstration. Puisque la suite $(U_{k \geq n}(A_k))_{n \in \mathbb{N}}$ est décroissante, on a

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k \geq n} A_k\right) \leq \lim_{n \rightarrow +\infty} \sum_{k \geq n} \mathbb{P}(A_k) = 0.$$

□

Théorème 3.11. (Second lemme de Borel-Cantelli) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(A_n)_{n \in \mathbb{N}}$ une suite d'événements indépendants. Si la série $\sum \mathbb{P}(A_n)$ diverge, alors $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 1$.

Démonstration. Puisque la suite $(U_{k \geq n}(A_k))_{n \in \mathbb{N}}$ est décroissante, on a

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k \geq n} A_k\right)$$

mais pour tout $n, m \in \mathbb{N}$ avec $n \geq m$, puisque A_n, A_{n+1}, \dots, A_m sont indépendants, on a

$$\mathbb{P}\left(\bigcup_{k=n}^m A_k\right) = 1 - \mathbb{P}\left(\bigcap_{k=n}^m A_k^c\right) = 1 - \prod_{k=n}^m \mathbb{P}(A_k^c) = 1 - \prod_{k=n}^m (1 - \mathbb{P}(A_k))$$

en utilisant l'inégalité $\forall x \in \mathbb{R}, 1 - x \leq e^{-x}$, on a

$$\mathbb{P}\left(\bigcup_{k=n}^m A_k\right) \geq 1 - \prod_{k=n}^m e^{-\mathbb{P}(A_k)} = 1 - e^{-\sum_{k=n}^m \mathbb{P}(A_k)} \xrightarrow{m \rightarrow +\infty} 1.$$

□

4. Convergence de suites de variables aléatoires

4.1. Convergence en loi

Définition 4.1. Soit (Ω, \mathcal{F}) un espace probabilisable, $(X_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ continue et bornée, on a

$$\mathbb{E}[f(X_n)] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[f(X)].$$

Théorème 4.2. Soit (Ω, \mathcal{F}) un espace probabilisable, $(X_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. Alors la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si pour tout $t \in \mathbb{R}$ tels que $\lim_{x \rightarrow t-} F_X(x) = F_X(t)$, on a

$$F_{X_n}(t) \xrightarrow{n \rightarrow +\infty} F_X(t).$$

Remarque 4.3. On peut caractériser la convergence en loi par la fonction de répartition.

Théorème 4.4. (Théoreme de Portemanteau) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, M un espace métrique, $(X_n : \Omega \rightarrow M)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires et $X : \Omega \rightarrow M$ un vecteur aléatoire. Alors les énoncés suivants sont équivalents :

(1) $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X .

(2) Pour toute fonction $f : M \rightarrow \mathbb{R}$ uniformément continue et bornée, on a

$$\mathbb{E}[f(X_n)] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[f(X)].$$

(3) Pour tout fermé $F \in M$, on a

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F).$$

(4) Pour tout ouvert $O \in M$, on a

$$\liminf_{n \rightarrow +\infty} \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O).$$

(5) Pour tout $A \in M$ tel que $\mathbb{P}(X \in \partial A) = 0$, on a

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A).$$

Démonstration.

(1) \Rightarrow (2) : Soit $f : M \rightarrow \mathbb{R}$ une fonction uniformément continue et bornée. Alors en particulier f est continue et bornée et on a bien

$$\mathbb{E}[f(X_n)] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[f(X)].$$

(2) \Rightarrow (3) : Soit $F \in M$ un fermé. On pose la fonction $f : \mathbb{R} \rightarrow [0, 1]$ définie par :

$$\forall x \in \mathbb{R}, f(x) = \begin{cases} 1 & \text{si } x \leq 0 \\ 1 - x & \text{si } 0 < x < 1 \\ 0 & \text{si } x \geq 1 \end{cases}$$

Pour tout $k \geq 1$, on pose $F_k := \{x \in M \mid d(x, F) \leq \frac{1}{k}\}$ et $f_k : M \rightarrow [0, 1]; x \mapsto f(kd(x, F))$ qui vaut :

$$\forall x \in M, f_k(x) = \begin{cases} 1 - kd(x, F) & \text{si } x \in F_k \\ 0 & \text{sinon} \end{cases}$$

On remarque que pour tout $x \in M$, on a

$$\mathbb{1}_F(x) \leq f_k(x) \leq \mathbb{1}_{F_k}(x)$$

par passage à l'espérance et à la limite, on obtient

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(X_n \in F) \leq \limsup_{n \rightarrow +\infty} \mathbb{E}[f_k(X_n)]$$

puisque f_k est uniformément continue et bornée, on a

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(X_n \in F) \leq \mathbb{E}[f_k(X)] \leq \mathbb{P}(X \in F_k)$$

puisque la suite $(F_k)_{k \in \mathbb{N} \setminus \{0\}}$ est décroissante et que les F_k sont fermés, par passage à la limite et continuité de la mesure, on a

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F).$$

(3) \Rightarrow (4) : Passage au complémentaire.

(3) + (4) \Rightarrow (5) : $\limsup_{n \rightarrow +\infty} \mathbb{P}(X_n \in A) = \liminf_{n \rightarrow +\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$.

(5) \Rightarrow (1) : Soit $f : M \rightarrow \mathbb{R}$ une fonction continue et bornée. Alors

$$E[f(X_n)] = \int_{\mathbb{R}} \mathbb{P}(f(X_n) > x) dx = \int_{\mathbb{R}} \mathbb{P}(X_n \in f^{-1}(]x, +\infty[)) dx$$

et par passage à la limite, on a

$$E[f(X)] = \int_{\mathbb{R}} \mathbb{P}(f(X) > x) dx = \int_{\mathbb{R}} \mathbb{P}(X \in f^{-1}(]x, +\infty[)) dx$$

mais puisque f est continue, on a $\partial f^{-1}(]x, +\infty[) = f^{-1}(\{x\})$ et on a

$$\sum_{x \in \mathbb{R}} \mathbb{P}(X \in \partial f^{-1}(]x, +\infty[)) = \sum_{x \in \mathbb{R}} \mathbb{P}(X \in f^{-1}(x)) < +\infty$$

donc $\{x \in \mathbb{R} \mid \mathbb{P}(X \in \partial f^{-1}(]x, +\infty[)) > 0\}$ est dénombrable et on a

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \in f^{-1}(]x, +\infty[)) = \mathbb{P}(X \in f^{-1}(]x, +\infty[))$$

enfin par convergence dominée $E[f(X_n)] = E[f(X)]$. □

Lemme 4.5. Soit (Ω, \mathcal{F}) un espace probabilisable, $(X_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires qui converge en loi vers une constante $a \in \mathbb{R}^d$. Alors

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(X_n \in B(a, \varepsilon)) = 1.$$

Démonstration. Soit $\varepsilon > 0$. Puisque $B(a, \varepsilon)$ est ouvert, d'après la propriété (4) du [Théorème 4.4](#) on a

$$\liminf_{n \rightarrow +\infty} \mathbb{P}(X_n \in B(a, \varepsilon)) \geq \mathbb{P}(a \in B(a, \varepsilon)) = 1.$$

□

Lemme 4.6. Soit (Ω, \mathcal{F}) un espace probabilisable, $(X_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ et $(Y_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ deux suites de vecteurs aléatoires qui convergent en loi respectivement vers un vecteur aléatoire $X : \Omega \rightarrow \mathbb{R}^d$ et une constante $a \in \mathbb{R}^d$. Alors la suite $((X_n, a))_{n \in \mathbb{N}}$ converge en loi vers (X, a) .

4.2. Convergence en probabilité

Définition 4.7. Soit (Ω, \mathcal{F}) un espace probabilisable, $(X_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X si pour tout $\varepsilon > 0$, elle vérifie :

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Proposition 4.8. Soit (Ω, \mathcal{F}) un espace probabilisable, $(X_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires, $X : \Omega \rightarrow \mathbb{R}^d$ et $Y : \Omega \rightarrow \mathbb{R}^d$ deux vecteurs aléatoires. Si $(X_n)_{n \in \mathbb{N}}$ converge vers X et Y en probabilité, alors $X = Y$ presque sûrement.

4.2.1. Stabilité de la convergence en probabilité

Définition 4.9. Soit (Ω, \mathcal{F}) un espace probabilisable et $(X_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires. On dit que $(X_n)_{n \in \mathbb{N}}$ est tendue si pour tout $\varepsilon > 0$, il existe un compact $K \in \mathbb{R}$ tel que :

$$\forall n \in \mathbb{N}, \mathbb{P}(X_n \in K) \geq 1 - \varepsilon.$$

Lemme 4.10. Soit (Ω, \mathcal{F}) un espace probabilisable, $(X_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. Si $(X_n)_{n \in \mathbb{N}}$ converge vers X en loi et X est presque sûrement fini, alors $(X_n)_{n \in \mathbb{N}}$ est tendue.

Proposition 4.11. Soit (Ω, \mathcal{F}) un espace probabilisable, $(X_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ et $(Y_n : \Omega \rightarrow \mathbb{R}^d)_{n \in \mathbb{N}}$ deux suites de vecteurs aléatoires qui convergent en probabilité respectivement vers deux vecteurs aléatoires $X : \Omega \rightarrow \mathbb{R}^d$ et $Y : \Omega \rightarrow \mathbb{R}^d$. Alors pour toute fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ continue, on a que $(f(X_n, Y_n))_{n \in \mathbb{N}}$ converge vers $f(X, Y)$ en probabilité. En particulier :

- (1) Pour tout $a, b \in \mathbb{R}$, on a que $(aX_n + bY_n)_{n \in \mathbb{N}}$ converge vers $aX + bY$ en probabilité.
- (2) Pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue, on a que $(f(X_n))_{n \in \mathbb{N}}$ converge vers $f(X)$ en probabilité.
- (3) $(X_n Y_n)_{n \in \mathbb{N}}$ converge vers XY en probabilité.