

Probabilités 2

Julien Poisat
(d'après les notes de José Trashorras)

Université Paris-Dauphine, Département MIDO, L2 MIE

18 janvier 2021

Table des matières

1	Probabilités, variables aléatoires et espérance	3
1.1	Ensemble fondamental	3
1.2	Tribus et mesures	4
1.2.1	Tribus	4
1.2.2	Tribu borélienne	4
1.2.3	Mesures	5
1.2.4	Propriétés	6
1.2.5	Indépendance et probabilité conditionnelle	7
1.3	Variable aléatoire	8
1.3.1	Mesurabilité	8
1.3.2	Loi d'une variable aléatoire	10
1.3.3	Caractérisation	10
1.4	Espérance	12
1.4.1	Principe général	12
1.4.2	Cas particuliers importants	13
1.4.3	Théorème de transfert.	16
1.4.4	Interversion limite-intégrale ou limite-espérance.	17
2	Espaces produits	19
2.1	Le produit de tribus	19
2.2	Le produit de mesures	20
2.3	Théorème d'interversion de Fubini	21
2.3.1	Cas général	21
2.3.2	Cas particuliers	21
2.3.3	Exemples	23
3	Vecteurs aléatoires	25
3.1	Introduction	25
3.2	Un exemple simple	25
3.3	Loi jointe d'un couple	27
3.4	Variance et covariance	28
3.5	Variables aléatoires indépendantes	28
3.6	Inégalités de Hölder et de Cauchy-Scharwz	30
3.7	Couples de variables discrètes	30
3.8	Couples à densité	31
3.9	Calcul de la densité d'un couple	32
3.9.1	Rappels en une dimension	33
3.9.2	Formule du changement de variable	33
3.9.3	Exemple	34
3.10	Vecteurs aléatoires, $d \geq 3$	35

3.10.1	Produit de plus de deux mesures	35
3.10.2	Matrice de dispersion	35
3.10.3	Indépendance	36
3.10.4	Indépendance affine	37
4	Théorèmes limites	39
4.1	Introduction	39
4.2	Suites de variables indépendantes	39
4.3	Fonctions caractérisant la loi d'une variable ou d'un vecteur aléatoire	39
4.3.1	Fonction caractéristique d'une variable aléatoire réelle	39
4.3.2	Fonction caractéristique d'un vecteur aléatoire	40
4.4	Quatre types de convergence	40
4.5	Théorèmes limites fondamentaux	42
4.5.1	Lois des grands nombres	42
4.5.2	Le théorème de la limite centrale	44
4.6	Combinaison de limites	44
5	Variables gaussiennes, vecteurs gaussiens	47
5.1	Introduction	47
5.2	Variables gaussiennes	47
5.3	Vecteurs gaussiens	48

Chapitre 1

Probabilités, variables aléatoires et espérance

Ce chapitre est constitué en grande partie de rappels du premier semestre.

1.1 Ensemble fondamental

Le rôle de l'ensemble fondamental Ω est de donner la liste de tous les résultats possibles de l'expérience à laquelle il est associé.

Définition 1.1. *On appelle ensemble fondamental associé à une expérience aléatoire \mathcal{E} l'ensemble Ω de tous les résultats possibles de cette expérience. Les éléments de Ω sont appelés résultats élémentaires de l'expérience.*

Exemple 1.1. *Dans chacun des cas suivants, déterminer l'ensemble fondamental :*

1. Lancer d'une pièce de monnaie.
2. Lancer d'un dé à six faces numérotées.
3. Tirages successifs et avec remise de deux boules dans une urne contenant une boule verte, une boule bleue et une boule rouge.
4. Tirages successifs et sans remise de deux boules dans une urne contenant une boule verte, une boule bleue et une boule rouge.
5. Tirages simultanés de deux boules dans une urne contenant une boule verte, une boule bleue et une boule rouge.
6. Nombre de passagers du métro à Paris pendant un an.
7. Durée de vie d'un individu.
8. Lancer dans un jeu de fléchettes.

Quelques remarques :

1. Il faut que la définition de Ω permette de garder le plus d'information possible sur l'expérience aléatoire associée. Par exemple dans l'exemple 3 on utilise des couples plutôt que des paires pour conserver l'information sur l'ordre de sortie des boules.
2. Parfois, étant donnée une expérience aléatoire, il n'y a pas d'ensemble Ω dont le choix s'impose : définir un ensemble fondamental relève alors d'une décision subjective. Dans ce cas il faut se souvenir qu'il n'existe pas de modèle vrai ou faux mais seulement des modèles plus ou moins efficaces au regard de l'utilisation que l'on en fait.

Il est important de savoir identifier les propriétés de Ω en termes de cardinal et de dimension car les outils que l'on utilise ensuite en dépendent.

1.2 Tribus et mesures

Soit Ω l'ensemble fondamental associé à une expérience, dont les sous-ensembles sont appelés *événements* ou *parties de Ω* , et $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω . Les événements dont on va chercher à calculer la probabilité (c'est-à-dire, à *mesurer*) forme un sous-ensemble de $\mathcal{P}(\Omega)$ qui doit satisfaire une structure particulière, appelée *tribu*.

1.2.1 Tribus

Définition 1.2. On appelle *tribu sur un ensemble Ω* tout ensemble \mathcal{A} de parties de Ω qui vérifie les conditions suivantes

1. $\Omega \in \mathcal{A}$.
2. Pour tout $A \in \mathcal{A}$ on a $A^c \in \mathcal{A}$.
3. Pour toute suite $(A_n)_{n \geq 1}$ d'éléments de \mathcal{A} on a $\cup_{n \geq 1} A_n \in \mathcal{A}$.

Un ensemble muni d'une tribu est appelé *espace mesurable*.

Exemple 1.2. Soit $\Omega = \{1, 2, 3\}$. La partie $\mathcal{A} = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$ de l'ensemble des parties de Ω est une tribu sur Ω . En revanche, $\mathcal{B} = \{\emptyset, \{1\}, \{1, 3\}, \{1, 2, 3\}\}$ n'est pas une tribu. Pourquoi ?

Exemple 1.3. L'ensemble des parties $\{\emptyset, \Omega\}$ est appelée *tribu triviale*.

Exemple 1.4. L'ensemble de toutes les parties $\mathcal{P}(\Omega)$ est une tribu de Ω . Nous utiliserons presque toujours cette tribu lorsque Ω est fini ou dénombrable.

Proposition 1.1. Soit Ω un ensemble et \mathcal{A} une tribu de parties de Ω . Alors :

1. $\emptyset \in \mathcal{A}$.
2. Pour tous $A, B \in \mathcal{A}$ on a $A \cup B \in \mathcal{A}$ et $A \cap B \in \mathcal{A}$.
3. Pour toute suite $(A_n)_{n \geq 1}$ d'éléments de \mathcal{A} on a $\cap_{n \geq 1} A_n \in \mathcal{A}$.

Proposition 1.2. Soit Ω un ensemble non vide et $(\mathcal{A}_i)_{i \in I}$ une famille quelconque de tribus sur Ω . L'ensemble de parties de Ω défini par $\cap_{i \in I} \mathcal{A}_i$ est une tribu sur Ω .

1.2.2 Tribu borélienne

Une classe de parties naturelle de \mathbb{R} est celle des intervalles (par exemple, ceux de la forme $]a, b[$, $a, b \in \mathbb{R}$). Malheureusement, cette classe ne forme pas une tribu (pourquoi?). Une manière de contourner le problème est de considérer la plus petite tribu contenant de tels intervalles. Cela conduit à la notion de *tribu engendrée* :

Définition 1.3. Soit Ω un ensemble non vide et $\mathcal{F} = (A_i)_{i \in I}$ une famille quelconque de parties de Ω . On appelle *tribu engendrée par \mathcal{F}* et on note $\sigma(\mathcal{F})$ l'intersection de toutes les tribus sur Ω qui contiennent \mathcal{F} .

Notons que $\sigma(\mathcal{F})$ est bien une tribu d'après la proposition 1.2.

Remarque 1.1. La tribu $\sigma(\mathcal{F})$ est la plus petite tribu au sens de l'inclusion contenant \mathcal{F} . (Pourquoi?)

Exemple 1.5. Soit Ω un ensemble non vide et A une partie de Ω . On a $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$. On écrit parfois $\sigma(A)$ au lieu de $\sigma(\{A\})$ pour alléger les notations.

D'après ce qui précède, nous pouvons maintenant donner la définition suivante :

Définition 1.4. On appelle tribu borélienne sur \mathbb{R} la tribu engendrée par la classe \mathcal{I} des intervalles de la forme $]a, b[$, où $a, b \in \mathbb{R}$ vérifient $a < b$. On note $\mathcal{B}(\mathbb{R})$ cette tribu et ses éléments sont appelés boréliens.

Proposition 1.3. La tribu borélienne est également engendrée par les classes de parties de \mathbb{R} suivantes (liste non exhaustive) :

1. $\mathcal{I}_1 = \{]a, b[, a, b \in \mathbb{Q}\}$;
2. $\mathcal{I}_2 = \{[a, b], a, b \in \mathbb{Q}\}$;
3. $\mathcal{I}_3 = \{]-\infty, a], a \in \mathbb{R}\}$;
4. $\mathcal{I}_4 = \{]-\infty, a[, a \in \mathbb{R}\}$;
5. $\mathcal{I}_5 = \{[a, +\infty[, a \in \mathbb{R}\}$;
6. $\mathcal{I}_6 = \{\mathcal{O}, \mathcal{O} \text{ est un ouvert de } \mathbb{R} \text{ muni de la topologie habituelle}\}$.

1.2.3 Mesures

Maintenant que nous avons défini l'ensemble fondamental et une classe d'événements appropriée, il est temps de définir la notion de *mesure de probabilité*. Nous commençons par introduire la notion un peu plus générale de *mesure* :

Définition 1.5. Soit (Ω, \mathcal{A}) un espace mesurable. On appelle mesure sur (Ω, \mathcal{A}) toute application μ définie sur \mathcal{A} , à valeurs dans $\mathbb{R}_+ \cup \{\infty\}$ et telle que pour toute suite $(A_n)_{n \geq 1}$ d'éléments de \mathcal{A} deux à deux disjoints on a

$$\mu(\cup_{n \geq 1} A_n) = \sum_{n \geq 1} \mu(A_n). \quad (1.1)$$

On dit alors que $(\Omega, \mathcal{A}, \mu)$ est un espace mesuré.

Remarque 1.2. La série $\sum_{n \geq 1} \mu(A_n)$ est bien définie comme série dont les termes sont à valeurs dans $\mathbb{R}_+ \cup \{\infty\}$. Il s'agit de la limite (fine ou infinie) de la suite croissante $(\sum_{n=1}^N \mu(A_n))_{N \geq 1}$, où par convention, $a + \infty = \infty$ pour tout $a \in \mathbb{R}_+ \cup \{\infty\}$.

On considérera plusieurs types de mesures.

Définition 1.6. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré.

1. Si $\mu(\Omega) < \infty$ on dira que μ est une mesure finie.
2. Si $\mu(\Omega) = 1$ on dira que μ est une **mesure de probabilité** (ou loi de probabilité), et l'espace mesuré est alors appelé **espace de probabilité**.

Voici quelques exemples importants :

Exemple 1.6 (Mesure de Dirac). Soit $a \in \mathbb{R}$. L'application δ_a qui à toute partie A de \mathbb{R} associe

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon} \end{cases} \quad (1.2)$$

est une mesure de probabilité sur \mathbb{R} (peu importe la tribu) appelée mesure de Dirac en a . Cette mesure de probabilité correspond à une expérience qui renvoie toujours a comme résultat.

Exemple 1.7 (Mesure de comptage). Sur l'espace mesuré $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ soit μ défini par

$$\mu(A) = \text{card}(A) = \sum_{n \in \mathbb{N}} \delta_n(A), \quad A \subset \mathbb{N}. \quad (1.3)$$

On appelle μ la mesure de comptage sur \mathbb{N} , notée par la suite $\delta_{\mathbb{N}}$. La mesure de comptage n'est pas une mesure finie.

Exemple 1.8 (Loi de probabilité sur \mathbb{N}). Pour n'importe quelle suite de réels positifs $(p_i)_{i \geq 0}$ vérifiant $\sum_{i \geq 0} p_i = 1$, on peut définir une mesure de probabilité \mathbb{P} sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ par

$$\mathbb{P}(A) = \sum_{i \in A} p_i, \quad A \subset \mathbb{N}. \quad (1.4)$$

Les réels p_i sont appelés probabilités ponctuelles. Cela marche si l'on remplace \mathbb{N} par n'importe quel ensemble au plus dénombrable.

Exemple 1.9 (Loi uniforme sur un espace fini). Soit E un ensemble fini. La loi uniforme sur $(E, \mathcal{P}(E))$ est la loi de probabilité définie par

$$\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(E)}, \quad A \subset E. \quad (1.5)$$

Passons enfin au cas continu. Nous admettrons le résultat suivant :

Théorème 1.1 (Admis). Il existe une unique mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ vérifiant $\lambda(]a, b[) = b - a$ pour tout intervalle $]a, b[$ de \mathbb{R} . On l'appelle mesure de Lebesgue.

Le mesure de Lebesgue sur \mathbb{R} correspond à la notion intuitive de longueur. Remarquons que cette mesure est invariante par translation.

Exemple 1.10 (Loi uniforme sur un intervalle). Soit $a < b$. La loi uniforme sur l'intervalle $[a, b]$ est la loi de probabilité sur \mathbb{R} définie par

$$\mathbb{P}(A) = \frac{\lambda(A \cap [a, b])}{b - a}, \quad A \in \mathcal{B}(\mathbb{R}). \quad (1.6)$$

1.2.4 Propriétés

Rappelons tout d'abord les deux notions d'incompatibilité :

Définition 1.7. (i) On dit des éléments d'une famille $(A_i)_{i \in I}$ d'événements qu'ils sont deux à deux incompatibles dès que pour tous $i, j \in I$ tels que $i \neq j$, on a $A_i \cap A_j = \emptyset$.
(ii) On dit des éléments d'une famille $(A_i)_{i \in I}$ d'événements qu'ils sont incompatibles dès que $\bigcap_{i \in I} A_i = \emptyset$.

Voici quelques propriétés élémentaires d'une mesure de probabilité.

Proposition 1.4. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Alors :

1. $\mathbb{P}(\emptyset) = 0$
2. Pour tout $n \in \mathbb{N}$ et toute famille A_1, A_2, \dots, A_n d'éléments de \mathcal{A} deux à deux incompatibles on a

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n).$$

3. Pour tout $A \in \mathcal{A}$ on a $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
4. Pour tous $A, B \in \mathcal{A}$ tels que $A \subset B$ on $\mathbb{P}(A) \leq \mathbb{P}(B)$.
5. Pour tout $A \in \mathcal{A}$ on a $0 \leq \mathbb{P}(A) \leq 1$.
6. Pour tous $A, B \in \mathcal{A}$ on a

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

En particulier $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

Proposition 1.5. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Toute suite croissante $(A_n)_{n \geq 1}$ d'éléments de \mathcal{A} , i.e. toute suite telle que si $i \leq j$ alors $A_i \subset A_j$, vérifie

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(A_n) &= \sup_{n \geq 1} \mathbb{P}(A_n) \\ &= \mathbb{P}(\cup_{n \geq 1} A_n). \end{aligned}$$

Corollaire 1.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Toute suite décroissante $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , i.e. toute suite telle que si $i \geq j$ alors $A_i \subset A_j$, vérifie

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(A_n) &= \inf_{n \in \mathbb{N}} \mathbb{P}(A_n) \\ &= \mathbb{P}(\cap_{n \geq 1} A_n). \end{aligned}$$

Exemple 1.11. Montrer que $\{a\}$ est un borélien de \mathbb{R} et que $\lambda(\{a\}) = 0$ pour tout $a \in \mathbb{R}$, où λ est la mesure de Lebesgue sur \mathbb{R} . On pourra commencer par montrer que le corollaire 1.1 reste valide si la mesure de probabilité \mathbb{P} est remplacée par une mesure μ et que $\mu(A_n)$ est finie à partir d'un certain rang.

1.2.5 Indépendance et probabilité conditionnelle

On va à présent introduire la notion d'indépendance. Il s'agit d'une notion importante qui est souvent mal distinguée de la notion d'incompatibilité.

Définition 1.8. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. On dit de deux événements $A, B \in \mathcal{A}$ qu'ils sont indépendants si et seulement si ils vérifient $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Il faut garder en tête que deux événements sont indépendants ou pas suivant la probabilité que l'on utilise comme référence, alors que la notion d'incompatibilité est intrinsèque aux événements considérés.

Exemple 1.12. On lance deux dés à six faces de couleurs différentes, l'un rouge et l'autre bleu. On considère les événements E : "Le dé rouge tombe sur 4" et F : "La somme des deux résultats est 7". Les événements E et F sont-ils indépendants ?

Définition 1.9. On dit des éléments d'une famille $(A_i)_{i \in I}$ qu'ils sont mutuellement indépendants dès qu'ils vérifient que pour toute sous-famille finie i_1, \dots, i_k de I ($k \geq 2$) on a

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_k}). \quad (1.7)$$

Exemple 1.13. Pour vérifier que A_1, A_2 et A_3 sont mutuellement indépendants il faut donc vérifier quatre conditions. Lesquelles ?

Attention : si les événements $(A_i)_{i \in I}$ sont mutuellement indépendants alors ils sont aussi indépendants deux à deux (pourquoi ?) alors que la réciproque n'est pas nécessairement vraie. (Trouver un contre-exemple !)

Intuitivement, deux événements A et B sont indépendants dès que la réalisation de l'un d'eux n'influe pas sur la probabilité d'occurrence de l'autre. C'est pour tenir compte du gain d'information que donne la connaissance du fait qu'un certain événement a lieu que l'on introduit les *probabilités conditionnelles*.

Définition 1.10. Soit $A \in \mathcal{A}$ tel que $\mathbb{P}(A) \neq 0$. On appelle probabilité conditionnelle sachant A et on note $\mathbb{P}(\cdot|A)$ l'application définie sur \mathcal{A} par

$$B \in \mathcal{A} \mapsto \mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Les probabilités conditionnelles permettent de simplifier le calcul de certaines probabilités quand on a identifié un système complet d'événements.

Définition 1.11. Soit A_1, A_2, \dots, A_n (resp. $(A_n)_{n \geq 1}$) une famille (resp. une suite) d'éléments de \mathcal{A} . On dit qu'elle constitue un système complet d'événements dès qu'elle vérifie

1. $\sum_{i=1}^n \mathbb{P}(A_i) = 1$ (resp. $\sum_{n \geq 1} \mathbb{P}(A_n) = 1$).
2. Les A_1, A_2, \dots, A_n (resp. $(A_n)_{n \geq 1}$) sont deux à deux incompatibles.

Le lien entre probabilités conditionnelles et systèmes complets d'événements est donné par la formule des probabilités totales et la formule de Bayes :

Proposition 1.6. Soit $A \in \mathcal{A}$ tel que $\mathbb{P}(A) \neq 0$.

1. $\mathbb{P}(\cdot|A)$ est une mesure de probabilité sur \mathcal{A} .
2. A et B sont indépendants si et seulement si on a $\mathbb{P}(B|A) = \mathbb{P}(B)$.
3. **Formule des probabilités totales :** Soit A_1, A_2, \dots, A_n (resp. $(A_n)_{n \geq 1}$) un système complet d'événements. Alors pour tout $B \in \mathcal{A}$ on a

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B \cap A_i) \quad (1.8)$$

(resp. $\mathbb{P}(B) = \sum_{n \geq 1} \mathbb{P}(B \cap A_n)$). Si pour tout $i = 1, \dots, n$ on a $\mathbb{P}(A_i) > 0$ (resp. pour tout $n \geq 1$ on a $\mathbb{P}(A_n) > 0$) alors

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i) \quad (1.9)$$

(resp. $\mathbb{P}(B) = \sum_{i=1}^{\infty} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$).

4. **Formule de Bayes ou "formule de la probabilité des causes" :** Soient A et B deux événements. Si $\mathbb{P}(B) \neq 0$ alors on a

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \quad (1.10)$$

Ainsi, d'après le premier point de la proposition précédente, une probabilité conditionnelle vérifie toutes les propriétés d'une probabilité, en particulier ceux de la proposition 1.4.

1.3 Variable aléatoire

1.3.1 Mesurabilité

Définition 1.12. Soient (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables. On dit d'une application $f : E \rightarrow F$ qu'elle est $\mathcal{E} - \mathcal{F}$ mesurable si et seulement si pour tout $B \in \mathcal{F}$ on a

$$f^{-1}(B) = \{x \in E : f(x) \in B\} \in \mathcal{E}.$$

Remarque 1.3. Dans le cas où $\mathcal{F} = \sigma(\mathcal{F}_0)$, \mathcal{F}_0 étant un ensemble quelconque de parties de F , il suffit que $f^{-1}(B) \in \mathcal{E}$ pour tout $B \in \mathcal{F}_0$ pour que f soit mesurable. En effet, on peut montrer dès lors que l'ensemble des parties $\{B \subset F : f^{-1}(B) \in \mathcal{E}\}$ forme une tribu. Comme cette tribu contient \mathcal{F}_0 , elle contient $\sigma(\mathcal{F}_0) = \mathcal{F}$.

Si l'espace de départ $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de probabilité alors toute fonction mesurable

$$X : \omega \in (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow X(\omega) \in (E, \mathcal{E}) \quad (1.11)$$

est appelée *variable aléatoire* (v.a.). On utilise d'habitude des lettres majuscules pour de tels objets. Une variable aléatoire réelle (v.a.r) est une variable aléatoire

$$X : \omega \in (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow X(\omega) \in (\mathbb{R}, \mathcal{B}(\mathbb{R})). \quad (1.12)$$

Ainsi, la première chose à retenir est qu'une variable aléatoire n'est pas une variable mais une fonction (mesurable). En particulier, une variable aléatoire (réelle) bornée n'est rien d'autre qu'une fonction (mesurable) bornée :

Définition 1.13. Une variable aléatoire réelle $X : \omega \in (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow X(\omega) \in (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite bornée s'il existe $M \in \mathbb{R}_+$ tel que, pour tout $\omega \in \Omega$, $|X(\omega)| \leq M$.

Il faut bien noter l'ordre des quantificateurs dans la définition précédente.

En pratique, les fonctions rencontrées dans ce cours seront souvent assez régulières (typiquement, continues sauf éventuellement en un nombre fini ou dénombrable de points). Ceci garantit leur mesurabilité, comme le montre la proposition suivante.

Proposition 1.7. Toute fonction $(\mathbb{R}, \mathcal{B}(\mathbb{R})) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ continue (éventuellement, par morceaux*) est mesurable.

Notons que si X est une v.a $(\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$ et que $f : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une fonction mesurable alors $f(X)$ est une v.a.r (pourquoi?) et insistons sur le fait que $f(X)$ est une fonction (composée) :

$$f(X) : \omega \in (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow f(X(\omega)) \in (\mathbb{R}, \mathcal{B}(\mathbb{R})). \quad (1.13)$$

La notion de mesurabilité est stable par les opérations élémentaires sur les fonctions. Ainsi, si f et g sont deux fonctions réelles et mesurables (par rapport aux mêmes tribus d'arrivée et de départ), alors $f + g$, αf (α est un réel) et $\max(f, g)$ sont aussi mesurables. On admettra que si $(f_n)_{n \geq 1}$ est une suite de fonctions réelles mesurables alors $\sup_{n \geq 1} (f_n)$ et $\inf_{n \geq 1} (f_n)$ sont mesurables.

Voici un exemple de fonctions mesurables que l'on rencontrera souvent.

Exemple 1.14. Soit (E, \mathcal{E}) un espace mesurable et $A \in \mathcal{E}$. La fonction indicatrice de A , notée 1_A et définie par

$$1_A : x \in (E, \mathcal{E}) \mapsto \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon,} \end{cases} \quad (1.14)$$

est mesurable de (E, \mathcal{E}) dans \mathbb{R} , peu importe la tribu que l'on met sur \mathbb{R} .

On s'autorisera parfois la notation $1_{\{x \in A\}}$ à la place de $1_A(x)$. Cette fonction est très souvent utilisée pour des expériences n'ayant que deux issues possibles (sondage à deux modalités, par exemple). Attention à ne pas confondre fonction indicatrice et dirac...

*. Une fonction définie sur un segment $f : [a, b] \mapsto \mathbb{R}$ est dite continue par morceaux ssi il existe une subdivision $\sigma = (x_0, \dots, x_n)$ telle que pour tout i , la restriction de f à (x_i, x_{i+1}) se prolonge sur $[x_i, x_{i+1}]$ en fonction continue. Une fonction f définie sur un intervalle I est continue par morceaux si sa restriction à tout segment $J \subset I$ est continue par morceaux.

1.3.2 Loi d'une variable aléatoire

Soit $X : \omega \in (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow X(\omega) \in (E, \mathcal{E})$ une variable aléatoire. Soit $B \in \mathcal{E}$. Par définition, l'ensemble $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ est un élément de \mathcal{A} .

Définition 1.14. La loi de la variable aléatoire X est la mesure de probabilité notée \mathbb{P}_X et définie sur \mathcal{E} par

$$\forall B \in \mathcal{E}, \quad \mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)). \quad (1.15)$$

Exercice 1.1. Montrer qu'il s'agit bien d'une mesure de probabilité.

On dit aussi que \mathbb{P}_X est la mesure induite par \mathbb{P} et X .

Pour simplifier les notations, on écrit généralement $\{X \in B\}$ à la place de $X^{-1}(B)$ et $\mathbb{P}(X \in B)$ à la place de $\mathbb{P}(X^{-1}(B))$, c'est-à-dire $\mathbb{P}_X(B)$.

Dans certains (voire la plupart) des exercices, on donne la loi de X sans préciser l'espace de départ $(\Omega, \mathcal{A}, \mathbb{P})$.

Nous verrons plus tard dans ce cours deux types de v.a.r : les v.a. discrètes et les v.a. continues.

Exemple 1.15. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $A \in \mathcal{A}$. Montrer que la variable aléatoire 1_A suit une loi de Bernoulli dont on précisera le paramètre.

1.3.3 Caractérisation

Dans cette section nous donnons des outils pour montrer que deux variables aléatoires ont même loi ou bien qu'une variable aléatoire a une certaine loi.

Egalité de deux mesures de probabilité.

De la même manière que deux applications linéaires définies sur le même espace vectoriel sont égales ssi elles coïncident sur une base de cet espace, il n'est pas toujours nécessaire de considérer tous les éléments d'une tribu pour comparer deux mesures de probabilité :

Théorème 1.2 (de Dynkin - **Admis**). Soit (E, \mathcal{E}) un espace mesurable tel que $\mathcal{E} = \sigma(\mathcal{C})$ où \mathcal{C} est un ensemble de parties de E stable par intersection finie[†]. Si ν et μ sont deux mesures de probabilité sur (E, \mathcal{E}) telles que pour tout $A \in \mathcal{C}$ on a $\mu(A) = \nu(A)$ alors ν et μ coïncident sur \mathcal{E} .

Remarque 1.4. Le théorème est en fait toujours vrai si les mesures ν et μ sont seulement σ -finies (cf définition 2.2).

Remarque 1.5. On trouve aussi ce théorème sous le nom de théorème des classes monotones.

Egalité en loi de deux variables aléatoires.

Le théorème 1.2 a comme conséquence immédiate :

Corollaire 1.2. Soient X et Y deux variables aléatoires ayant même espace d'arrivée (E, \mathcal{E}) , avec \mathcal{E} engendré par une classe de parties \mathcal{C} stable par intersection finie. Alors X et Y ont même loi (i.e. $\mathbb{P}_X = \mathbb{P}_Y$) ssi $\mathbb{P}_X(A) = \mathbb{P}_Y(A)$ pour tout $A \in \mathcal{C}$.

Attention : Si les variables aléatoires X et Y sont égales, au sens où $X(\omega) = Y(\omega)$ pour tout $\omega \in \Omega$ alors X et Y ont même loi. La réciproque n'est pas vraie.

[†]. Montrer que sans cette hypothèse la proposition devient fausse.

Fonction de répartition.

Dans le cas de variables aléatoires réelles, on obtient par le corollaire 1.2 et la proposition 1.3, en choisissant $\mathcal{C} = \{] - \infty, x], x \in \mathbb{R} \}$ (que faut-il vérifier sur cette classe de parties?) :

Proposition 1.8. *Deux variables aléatoires réelles X et Y ont même loi ssi $\mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x)$ pour tout $x \in \mathbb{R}$.*

Cela nous conduit naturellement à la définition suivante :

Définition 1.15. *La fonction de répartition d'une v.a.r. X est la fonction*

$$F_X : x \in \mathbb{R} \mapsto \mathbb{P}(X \leq x) \in [0, 1]. \quad (1.16)$$

Mise en garde : une fonction de répartition est définie sur tout \mathbb{R} . Ainsi, lorsque l'expression d'une fonction de répartition est demandée, il faut donner celle-ci sur tout \mathbb{R} pour répondre totalement à la question.

Une fonction de répartition possède les propriétés suivantes :

Proposition 1.9. *Toute fonction de répartition est croissante, continue à droite et limitée à gauche (abbr. càdlàg) et converge vers 0 (resp 1) en $-\infty$ (resp. $+\infty$).*

Remarque 1.6. *Le fait que la fonction de répartition soit limitée à gauche est une conséquence directe de la croissance et du fait que la fonction est bornée supérieurement (par un).*

Remarque 1.7. *Une fonction F est continue à droite en x lorsque $F(x) = \lim_{y \rightarrow x^+} F(y)$ et elle est limitée à gauche en x lorsque $\lim_{y \rightarrow x^-} F(y)$ existe. On note habituellement, lorsque ces quantités existent,*

$$\begin{aligned} F(x^-) &= \lim_{y \rightarrow x^-} F(y), \\ F(x^+) &= \lim_{y \rightarrow x^+} F(y). \end{aligned} \quad (1.17)$$

Il s'agit de la limite de $F(y)$ lorsque y tend vers x en prenant des valeurs inférieures (resp. supérieures) à x .

Insistons sur le fait qu'une fonction de répartition est bien définie pour tout v.a.r, qu'elle soit discrète ou continue. Néanmoins, celle-ci a une allure différente selon ces deux cas de figure.

Exemple 1.16. *Calculer la fonction de répartition d'une loi de Bernoulli de paramètre $\frac{1}{2}$ et montrer que celle-ci a deux discontinuités, en 0 et 1. De tels points de discontinuité sont appelés atomes.*

Dans le cas où x est un point de discontinuité de F_X , la quantité $F_X(x)$ est représentée graphiquement par \bullet sur la courbe de F_X tandis que $F_X(x^-)$ est représentée par \circ .

Exemple 1.17. *Calculer la fonction de répartition d'une loi uniforme sur $[0, 1]$.*

On obtient finalement la caractérisation suivante :

Proposition 1.10. *Deux variables aléatoires réelles X et Y ont même loi ssi elles ont même fonction de répartition.*

En particulier, pour montrer qu'une v.a.r a une loi connue, il suffit de calculer sa fonction de répartition. C'est particulièrement utile pour les v.a.r. continues.

En fait, la classe des intervalles de la forme $] - \infty, x]$ n'est pas la seule à engendrer $\mathcal{B}(\mathbb{R})$. En considérant d'autres classes, on obtient :

Proposition 1.11. Soient X et Y deux variables aléatoires réelles. Les propositions suivantes sont équivalentes :

1. X et Y ont même loi ;
2. $\mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x)$ pour tout $x \in \mathbb{R}$;
3. $\mathbb{P}(X > x) = \mathbb{P}(Y > x)$ pour tout $x \in \mathbb{R}$;
4. $\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq Y \leq b)$ pour tout $a \leq b$;
5. $\mathbb{P}(a < X \leq b) = \mathbb{P}(a < Y \leq b)$ pour tout $a \leq b$;
6. ...

Variables aléatoires discrètes

Pour les variables aléatoires à valeurs dans un espace fini ou dénombrable, c'est plus simple :

Proposition 1.12. Soient X et Y deux variables aléatoires à valeurs dans un espace fini ou dénombrable E muni de la tribu $\mathcal{P}(E)$. Alors X et Y ont même loi ssi

$$\forall i \in E, \quad \mathbb{P}(X = i) = \mathbb{P}(Y = i). \quad (1.18)$$

Ainsi, pour déterminer la loi d'une variable aléatoire discrète il suffit de déterminer ses probabilités ponctuelles.

1.4 Espérance

L'espérance correspond à la valeur moyenne « théorique » d'une variable aléatoire réelle, c'est-à-dire à la valeur que l'on peut « espérer ».

Calculer l'espérance d'une variable aléatoire signifie intégrer cette fonction par rapport à sa mesure (ou loi) de probabilité. L'intégrale d'une fonction par rapport à une mesure (appelée intégrale de Lebesgue) est une notion plus générale que la notion d'intégrale vue en cours d'analyse. En fait, celle-ci englobe à la fois la notion d'intégrale d'une fonction continue (ou continue par morceaux) sur un intervalle (intégrale de Riemann) et la notion de série numérique ! En pratique, nous ne considérerons que ces deux cas particuliers.

1.4.1 Principe général

La construction de l'intégrale de Lebesgue sera vue et approfondie en troisième année. Toutefois, nous retiendrons pour ce cours ces quelques points :

Soit (E, \mathcal{E}, μ) un espace mesuré et une fonction $f : E \rightarrow \mathbb{R}$ mesurable. L'intégrale de f par rapport à μ , notée $\int f d\mu$, $\int f(x) d\mu(x)$ ou $\int f(x) \mu(dx)$, est définie comme suit :

- Si f est positive et ne prend qu'un nombre fini de valeurs, c'est-à-dire si f s'écrit $\sum_{i=1}^n a_i 1_{A_i}$ avec $a_i \geq 0$ (on dit que f est étagée positive) alors on définit

$$\int f d\mu = \sum_{i=1}^n a_i \mu(A_i) \in [0, +\infty] \quad (1.19)$$

- Si f est positive alors on définit

$$\int f d\mu = \sup \left\{ \int g d\mu : g \leq f, g \text{ étagée positive} \right\} \in [0, +\infty]. \quad (1.20)$$

- Si f est de signe quelconque et $\int |f|d\mu$ est fini (on dit alors que f est *intégrable*) alors on définit

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu, \quad (1.21)$$

où f_+ et f_- désignent les parties positive et négative de f , respectivement. Dans l'équation précédente, toutes les quantités sont finies.

Voici les trois propriétés à connaître pour survivre :

Proposition 1.13 (Admise). *Soit (E, \mathcal{E}, μ) un espace mesuré, $f, g : E \rightarrow \mathbb{R}$ deux fonctions intégrables et $\lambda \in \mathbb{R}$. On a*

1. (Linéarité) $\int_E (f + \lambda g) d\mu = \int_E f d\mu + \lambda \int_E g d\mu$;
2. (Monotonie) Si pour tout $x \in E$ on a $f(x) \leq g(x)$ alors $\int_E f d\mu \leq \int_E g d\mu$;
3. (Inégalité triangulaire) $|\int_E f d\mu| \leq \int_E |f| d\mu$.

Remarque 1.8. *Les propriétés de linéarité et de monotonie restent vraies si f et g sont positives (pourvu que $\lambda \geq 0$ pour le premier point).*

Définition 1.16. *Soit X une variable aléatoire réelle définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ positive ou intégrable (par rapport à \mathbb{P}). On note*

$$\mathbb{E}(X) = \int X d\mathbb{P} = \int X(\omega) d\mathbb{P}(\omega) \quad (1.22)$$

l'espérance de X .

Ainsi, on peut toujours parler de l'espérance de X quand X est une v.a.r positive, mais celle-ci peut être finie ou infinie. Si X est de signe quelconque, il faut d'abord s'assurer que X soit intégrable, c'est-à-dire que $\mathbb{E}(|X|)$ soit finie, pour pouvoir parler de l'espérance de X . Si X est bien intégrable, alors $\mathbb{E}(X)$ est finie.

Bien entendu, les trois propriétés de la proposition 1.13 se transposent avec la notation « espérance ».

Exemple 1.18. *Intuitivement, quelle est l'espérance d'une constante ? Quelle propriété de cette section permet de confirmer votre réponse ?*

Exemple 1.19. *A l'aide des propositions précédentes, montrer que toute v.a.r bornée est intégrable.*

Remarque 1.9 (Restriction d'une intégrale). *Si f est intégrable par rapport à une mesure μ et si A est un ensemble mesurable alors $f1_A$ est également intégrable et on note*

$$\int_A f d\mu = \int f1_A d\mu. \quad (1.23)$$

Nous admettrons que cette quantité est nulle lorsque $\mu(A) = 0$.

1.4.2 Cas particuliers importants

En pratique, pour mener les calculs, nous utiliserons ces quelques cas particuliers, à connaître.

Intégration d'une fonction indicatrice.

En appliquant (1.19), on a pour tout ensemble mesurable A ,

$$\int 1_A d\mu = \mu(A), \quad \text{ou encore} \quad \mathbb{E}(1_{\{X \in A\}}) = \mathbb{P}(X \in A). \quad (1.24)$$

Intégration par rapport à un dirac.

Soit $x \in \mathbb{R}$ et $f: \mathbb{R} \rightarrow \mathbb{R}$. Si l'on considère la mesure $\mu = \delta_x$, alors $\int f d\mu = f(x)$.

Intégration par rapport à la mesure de comptage (séries)

Si μ est la mesure de comptage et $f: \mathbb{N} \rightarrow \mathbb{R}$ positive alors $\int f d\mu = \sum_{n \geq 1} f(n)$, qui peut être fini ou infini. Si f est de signe quelconque alors f est intégrable par rapport à μ ssi $\sum_{n \geq 1} |f(n)| < \infty$. Ainsi, dans ce cas de figure, la notion d'intégrabilité correspond à une notion de convergence de série.

Intégration par rapport à la mesure de Lebesgue (intégrales usuelles).

Pour rappel, on note λ la mesure de Lebesgue. Pour des fonctions suffisamment régulières, intégrer par rapport à la mesure de Lebesgue revient à l'intégrale usuelle (celle vue en analyse, dite de Riemann) et on pourra alors utiliser toutes les techniques d'intégration déjà vues. Les trois propositions suivantes sont admises.

Proposition 1.14 (Lien avec l'intégrale définie). *Soit f une fonction continue (par morceaux, éventuellement) sur un segment $[a, b]$. Alors*

$$\int_{[a,b]} f(x) d\lambda(x) = \int_a^b f(x) dx. \quad (1.25)$$

Proposition 1.15 (Lien avec l'intégrale impropre - cas positif). *Soit f une fonction **positive** et continue (par morceaux, éventuellement) sur un intervalle $[a, b[$ où $b \leq \infty$. Alors l'intégrale impropre $\int_a^b f(x) dx$ est convergente ssi $\int_{[a,b[} f(x) d\lambda(x) < \infty$, et alors*

$$\int_{[a,b[} f(x) d\lambda(x) = \int_a^b f(x) dx. \quad (1.26)$$

Proposition 1.16 (Lien avec l'intégrale impropre - cas général). *Soit f une fonction continue (par morceaux, éventuellement) sur un intervalle $[a, b[$ où $b \leq \infty$. Alors f est intégrable sur $[a, b[$ par rapport à la mesure de Lebesgue ssi l'intégrale impropre $\int_a^b |f(x)| dx$ est convergente, et alors*

$$\int_{[a,b[} f(x) d\lambda(x) = \int_a^b f(x) dx. \quad (1.27)$$

Le piège : La convergence de l'intégrale impropre $\int_a^b f(x) dx$ n'entraîne pas l'intégrabilité de f .

Exemple 1.20. *L'intégrale impropre $\int_0^\infty \frac{\sin x}{x} dx$ est bien définie (et finie) mais la fonction $x \mapsto \frac{\sin x}{x}$ n'est pas intégrable sur \mathbb{R}_+ .*

Espérance d'une variable aléatoire réelle discrète.

Soit X une variable aléatoire réelle à valeurs dans un espace fini ou dénombrable $\mathcal{X} = \{x_1, x_2, \dots\}$. La v.a X est intégrable ssi

$$\mathbb{E}(|X|) = \sum_{i \geq 1} |x_i| \mathbb{P}(X = x_i) < \infty. \quad (1.28)$$

Si X est positive ou intégrable, alors

$$\mathbb{E}(X) = \sum_{i \geq 1} x_i \mathbb{P}(X = x_i). \quad (1.29)$$

Les lois discrètes usuelles à connaître : Bernoulli, binomiale, géométrique, Poisson.

Espérance d'une variable aléatoire réelle continue.

Définition 1.17. Dans ce cours, on appellera densité de probabilité (sur \mathbb{R}) une fonction f positive, continue (éventuellement par morceaux) et telle que $\int_{-\infty}^{\infty} f(x)dx = 1$.

Remarque 1.1. La définition implique qu'une densité de probabilité est intégrable sur \mathbb{R} .

Remarque 1.2. L'hypothèse de continuité peut être remplacée par l'hypothèse plus faible de mesurabilité.

Définition 1.18. On dit qu'une mesure de probabilité \mathbb{P} sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une mesure (ou loi) de densité f (sous-entendu : par rapport à la mesure de Lebesgue λ) si pour tout $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(A) = \int_A f(x)d\lambda(x). \quad (1.30)$$

Dans le cas où A est un intervalle $[a, b]$, on a simplement, d'après ce qui précède :

$$\mathbb{P}([a, b]) = \int_a^b f(x)dx. \quad (1.31)$$

On trouve parfois les notations

$$\mathbb{P}(dx) = f(x)dx, \quad d\mathbb{P}(x) = f(x)dx. \quad (1.32)$$

Naturellement, on dit qu'une v.a.r. X a une densité f si sa loi \mathbb{P}_X a pour densité f . On parle dans ce cas de *variable aléatoire continue*.

Remarque 1.3. Une densité doit être définie sur tout \mathbb{R} . Le support d'une densité de probabilité f (et par extension, le support d'une variable aléatoire réelle suivant cette densité) est le plus petit sous-ensemble fermé de \mathbb{R} (cf. cours analyse) dont le complémentaire est de probabilité nulle. Par exemple, la densité d'une loi exponentielle de paramètre $\lambda > 0$ s'écrit

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0, +\infty[}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon,} \end{cases} \quad (1.33)$$

et son support est $[0, +\infty[$.

Les lois à densité usuelles (à connaître) sont les lois uniforme, exponentielle et normale (ou gaussienne).

Soit X une variable aléatoire réelle continue de densité f . Elle est intégrable ssi

$$\mathbb{E}(|X|) = \int_{-\infty}^{\infty} |x|f(x)dx < \infty. \quad (1.34)$$

Si X est positive ou intégrable, alors

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx. \quad (1.35)$$

Exemple 1.21. Montrer qu'une variable aléatoire suivant la loi normale centrée réduite est intégrable et d'espérance nulle.

Lien entre densité et fonction de répartition.

Soit X une v.a.r. de densité f . D'après la définition, on a

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in]-\infty, x]) = \int_{]-\infty, x]} f(t) d\lambda(t) = \int_{-\infty}^x f(t) dt. \quad (1.36)$$

Ainsi, si f est continue sur \mathbb{R} , F_X est dérivable sur \mathbb{R} d'après le théorème fondamental de l'analyse. De manière équivalente, F_X est une primitive de f (plus précisément, l'unique primitive dont la limite en $-\infty$ vaut 0). Si f est seulement continue par morceaux alors F reste dérivable, sauf aux points de discontinuité de f .

Exemple 1.22. *Etudier la dérivabilité des fonctions de répartition associées à la loi $\mathcal{N}(0, 1)$ et la loi $\mathcal{U}([0, 1])$.*

1.4.3 Théorème de transfert.

Supposons que l'on connaisse la loi d'une v.a.r. X et soit $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ mesurable. Le théorème de transfert nous permet de calculer l'espérance de la v.a.r. $\varphi(X)$ **sans avoir besoin de déterminer sa loi**.

Proposition 1.1 (Admis). *Soit X une v.a.r. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable.*

1. *Si φ est positive on a*

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

2. *Si φ n'est pas de signe constant mais que $\mathbb{E}[|\varphi(X)|] < \infty$ l'égalité ci-dessus est encore vraie.*

Voici maintenant deux cas particuliers :

— Soit X une variable aléatoire réelle discrète à valeurs dans $\mathcal{X} = \{x_1, x_2, \dots\}$. Si $\varphi(X)$ est positive ou intégrable, alors

$$\mathbb{E}(\varphi(X)) = \sum_{i \geq 1} \varphi(x_i) \mathbb{P}(X = x_i). \quad (1.37)$$

— Soit X une variable aléatoire réelle continue de densité f . Si $\varphi(X)$ est positive ou intégrable, alors

$$\mathbb{E}(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x) f(x) dx. \quad (1.38)$$

La proposition 1.1 admet une sorte de réciproque :

Proposition 1.2. *Soit X une v.a.r. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$. S'il existe une mesure de probabilité Q sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour toute application $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ mesurable bornée on ait*

$$\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) dQ \quad (1.39)$$

alors nécessairement $Q = \mathbb{P}_X$.

1.4.4 Interversions limite-intégrale ou limite-espérance.

Comme cela a été vu en analyse, il n'est pas toujours possible d'intervertir le signe « limite » et le signe « intégrale » (au sens de Riemann). De la même manière, l'intervention limite-intégrale (au sens de Lebesgue) ou limite-espérance n'est pas toujours licite. Les deux théorèmes suivants, admis, donnent deux situations dans lesquelles cela est toutefois possible.

Théorème 1.3 (Convergence monotone). *Soit $(f_n)_{n \geq 1}$ une suite croissante de fonctions mesurables positives et f la fonction définie par $f(x) = \lim_{n \rightarrow \infty} f_n(x)$. Alors $(\int_E f_n d\mu)_{n \geq 1}$ est une suite croissante qui admet $\int_E f d\mu$ pour limite.*

Théorème 1.4 (Convergence dominée). *Soit $(f_n)_{n \geq 1}$ une suite de fonctions mesurables convergant simplement vers la fonction f . S'il existe une fonction g positive et intégrable par rapport à μ et si $|f_n| \leq g$ à partir d'un certain rang, alors f est intégrable par rapport à μ et*

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu. \quad (1.40)$$

Bien sûr, ces deux théorèmes se transposent naturellement au cas de variables aléatoires réelles.

Exemple 1.23. *Soit X une v.a.r positive. On tronque X en posant $X_n = \min(X, n)$ pour tout $n \in \mathbb{N}$. Montrer que $E(X_n)$ converge vers $E(X)$ quand $n \rightarrow \infty$.*

A noter également que l'hypothèse de convergence simple dans les théorèmes de convergence dominée et monotone peut-être remplacée par l'hypothèse plus faible de convergence presque-sûre :

Définition 1.19 (Convergence presque-sûre). *On dit que la suite de fonction mesurables $(f_n)_{n \geq 1}$ converge μ -presque-sûrement (μ -p.s) vers f si l'ensemble des éléments x tels que $(f_n(x))_{n \geq 1}$ ne converge pas vers $f(x)$ est de mesure nulle par rapport à μ (on admet que cet ensemble est bien mesurable).*

Remarque 1.10. *Plus généralement, on dit qu'un événement arrive μ -presque-sûrement si $\mu(A^c) = 0$. Montrer que si $(A_n)_{n \geq 1}$ est une famille dénombrable d'événements μ -presque-sûrs alors l'événement $\bigcap_{n \geq 1} A_n$ arrive μ -p.s. Est-ce toujours vrai si l'on considère une famille d'événements quelconque ?*

Chapitre 2

Espaces produits

Dans ce chapitre nous posons le problème de l'intégration des fonctions définies sur n'importe quel ensemble ayant une structure d'espace mesuré. Ainsi si E est un espace produit, c'est-à-dire $E = E_1 \times E_2$, il est possible de définir la notion d'intégrale pour des fonctions réelles mesurables f définies sur E une fois que l'on s'est donné une tribu \mathcal{E} sur E et une mesure μ sur (E, \mathcal{E}) . Ce que l'on va voir dans ce paragraphe est un cas particulier : que se passe-t-il quand on nous donne deux espaces mesurés $(E_1, \mathcal{E}_1, \mu_1)$ et $(E_2, \mathcal{E}_2, \mu_2)$? Existe-t-il une tribu sur $E_1 \times E_2$ qui soit naturellement en rapport avec \mathcal{E}_1 et \mathcal{E}_2 ? Si oui existe-t-il une mesure sur cette tribu qui soit naturellement en rapport avec μ_1 et μ_2 ? On va voir que la réponse à ces deux questions est oui et qu'elles donnent lieu à la tribu produit $\mathcal{E}_1 \otimes \mathcal{E}_2$ et à la mesure produit $\mu_1 \otimes \mu_2$.

2.1 Le produit de tribus

Définition 2.1. Soient (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) deux espaces mesurables. On appelle tribu produit et on note $\mathcal{E}_1 \otimes \mathcal{E}_2$ la tribu engendrée par la classe des pavés :

$$\mathcal{C} = \{B_1 \times B_2, \text{ où } B_1 \in \mathcal{E}_1 \text{ et } B_2 \in \mathcal{E}_2\}.$$

Remarque 2.1. Soient $E_1 = \{0, 1\}$ muni de la tribu $\mathcal{E}_1 = \mathcal{P}(E_1) = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$ et $E_2 = \{a, b\}$ muni de la tribu $\mathcal{E}_2 = \mathcal{P}(E_2) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$. Alors

$$E_1 \times E_2 = \{(0, a), (0, b), (1, a), (1, b)\}$$

et, par exemple $\{(0, a)\} = \{0\} \times \{a\} \in \mathcal{C}$. Par conséquent $\{(0, a)\} \in \mathcal{E}_1 \otimes \mathcal{E}_2$, donc

$$\{(0, a)\}^c = \{(0, b), (1, a), (1, b)\} \in \mathcal{E}_1 \otimes \mathcal{E}_2$$

puisque $\mathcal{E}_1 \otimes \mathcal{E}_2$ étant une tribu, le complémentaire de tout élément de $\mathcal{E}_1 \otimes \mathcal{E}_2$ est dans $\mathcal{E}_1 \otimes \mathcal{E}_2$. Cependant, $\{(0, a)\}^c \notin \mathcal{C}$: les éléments de $\mathcal{E}_1 \otimes \mathcal{E}_2$ ne sont pas tous des pavés mais tous les pavés sont évidemment des éléments de $\mathcal{E}_1 \otimes \mathcal{E}_2$.

Proposition 2.1. Soient (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) deux espaces mesurables. La tribu $\mathcal{E}_1 \otimes \mathcal{E}_2$ est la plus petite tribu sur $E_1 \times E_2$ qui rend mesurables les deux applications

$$p_1 : E_1 \times E_2 \rightarrow E_1 \quad \text{et} \quad p_2 : E_1 \times E_2 \rightarrow E_2 \\ (x_1, x_2) \mapsto x_1 \quad \quad \quad (x_1, x_2) \mapsto x_2.$$

Proposition 2.2. (Propriété de section) Soient (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) deux espaces mesurables. Tout $B \in \mathcal{E}_1 \otimes \mathcal{E}_2$ vérifie la propriété suivante : pour tout $x_1 \in E_1$ on a

$$B_2(x_1) = \{x_2 \in E_2, (x_1, x_2) \in B\} \in \mathcal{E}_2$$

et pour tout $x_2 \in E_2$ on a

$$B_1(x_2) = \{x_1 \in E_1, (x_1, x_2) \in B\} \in \mathcal{E}_1.$$

Remarque 2.2. Il est naturel de se demander quel est le rapport entre les deux tribus sur \mathbb{R}^2 que sont $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ et $\mathcal{B}(\mathbb{R}^2)$, définie comme la tribu engendrée par les ouverts de \mathbb{R}^2 . Nous admettrons que ce sont les mêmes.

Voici pour terminer cette partie deux propositions concernant les fonctions définies sur, ou à valeurs dans un espace produit.

Proposition 2.3. Une fonction f est mesurable de (F, \mathcal{F}) dans $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ si et seulement si $f_1 = p_1 \circ f$ est mesurable de (F, \mathcal{F}) dans (E_1, \mathcal{E}_1) et $f_2 = p_2 \circ f$ est mesurable de (F, \mathcal{F}) dans (E_2, \mathcal{E}_2) .

Proposition 2.4. Soit f une application mesurable de $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ dans (F, \mathcal{F}) . Pour tout $x_1 \in E_1$, on note $f(x_1, \cdot)$ (resp. $f(\cdot, x_2)$) l'application $x_2 \in E_2 \mapsto f(x_1, x_2)$ (resp. $x_1 \in E_1 \mapsto f(x_1, x_2)$). Alors, pour tout $x_1 \in E_1$, l'application $f(x_1, \cdot)$ est mesurable de (E_2, \mathcal{E}_2) dans (F, \mathcal{F}) . De même, pour tout $x_2 \in E_2$, l'application $f(\cdot, x_2)$ est mesurable de (E_1, \mathcal{E}_1) dans (F, \mathcal{F}) .

2.2 Le produit de mesures

Les mesures μ_1 et μ_2 considérées dans cette partie peuvent être des mesures finies (et en particulier des probabilités) ou plus généralement des mesures σ -finies.

Définition 2.2. Une mesure μ sur une espace (E, \mathcal{E}) est dite σ -finie s'il existe une suite $(C_n)_{n \geq 1}$ d'éléments de \mathcal{E} telle que $E = \cup_{n \geq 1} C_n$ et $\mu(C_n) < \infty$ pour tout $n \geq 1$.

Autrement dit, une mesure est σ -finie lorsque l'espace entier peut être recouvert par des ensembles de mesure finie. Par exemple, la mesure de Lebesgue sur \mathbb{R} n'est pas finie mais elle est σ -finie.

Théorème 2.1 (Admis). Soient $(E_1, \mathcal{E}_1, \mu_1)$ et $(E_2, \mathcal{E}_2, \mu_2)$ deux espaces mesurés, μ_1 et μ_2 étant des mesures σ -finies. Pour tout $A \in \mathcal{E}_1 \otimes \mathcal{E}_2$ les applications définies par

$$\begin{aligned} f: E_1 &\rightarrow \mathbb{R}_+ & \text{et} & & g: E_2 &\rightarrow \mathbb{R}_+ \\ x &\mapsto \mu_2(A_2(x)) & & & y &\mapsto \mu_1(A_1(y)) \end{aligned}$$

où $A_2(x) = \{y \in E_2, (x, y) \in A\}$ et $A_1(y) = \{x \in E_1, (x, y) \in A\}$ sont mesurables et

$$\int_{E_1} f d\mu_1 = \int_{E_2} g d\mu_2. \quad (2.1)$$

En notant $\mu_1 \otimes \mu_2(A)$ la valeur commune de ces deux intégrales on définit une unique mesure σ -finie sur $\mathcal{E}_1 \otimes \mathcal{E}_2$ que l'on appellera mesure produit de μ_1 et μ_2 et que l'on notera $\mu_1 \otimes \mu_2$.

De plus, pour tout élément B de $\mathcal{E}_1 \otimes \mathcal{E}_2$ de la forme $B = B_1 \times B_2, B_1 \in \mathcal{E}_1, B_2 \in \mathcal{E}_2$,

$$(\mu_1 \otimes \mu_2)(B) = \mu_1(B_1)\mu_2(B_2). \quad \text{[Important!]} \quad (2.2)$$

Remarquons qu'étant donnés μ_1 et μ_2 il ne peut exister qu'une mesure vérifiant la propriété (2.2).

Exemple 2.1. On peut munir \mathbb{N}^2 , vu comme espace produit $\mathbb{N} \times \mathbb{N}$, de la tribu produit $\mathcal{P}(\mathbb{N}) \otimes \mathcal{P}(\mathbb{N})$ et de la mesure produit $\delta_{\mathbb{N}} \otimes \delta_{\mathbb{N}}$. Montrer que $\mathcal{P}(\mathbb{N}) \otimes \mathcal{P}(\mathbb{N})$ coïncide avec $\mathcal{P}(\mathbb{N}^2)$ et que $\delta_{\mathbb{N}} \otimes \delta_{\mathbb{N}}$ coïncide avec la mesure de comptage sur \mathbb{N}^2 , c'est-à-dire que pour tout $A \subset \mathbb{N}^2$,

$$(\delta_{\mathbb{N}} \otimes \delta_{\mathbb{N}})(A) = \text{card}(A). \quad (2.3)$$

Exemple 2.2. Si λ est la mesure de Lebesgue sur \mathbb{R} alors la mesure produit $\lambda_2 = \lambda \otimes \lambda$ est appelée mesure de Lebesgue sur \mathbb{R}^2 et pour tout ensemble $A \in \mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$, la quantité $\lambda_2(A)$ correspond à la notion d'aire de A . Pourquoi ?

Exemple 2.3. Munissons $\mathbb{R} \times \mathbb{N}$ de la tribu produit $\mathcal{B}(\mathbb{R}) \otimes \mathcal{P}(\mathbb{N})$. Caractériser les éléments de cette tribu et calculer leur mesure selon $\lambda \otimes \delta_{\mathbb{N}}$.

Exemple 2.4. On tire au hasard et de manière uniforme un point dans le rectangle $[0, 3] \times [1, 2]$. Quelle mesure de probabilité est naturellement associée à cette expérience ?

Exemple 2.5. Soient \mathbb{P}_1 et \mathbb{P}_2 deux mesures de probabilité sur (Ω, \mathcal{A}) , et A et B deux éléments de \mathcal{A} . Montrer que sous $\mathbb{P}_1 \otimes \mathbb{P}_2$, les événements $A \times \Omega$ et $\Omega \times B$ sont indépendants.

Exercice 2.1 (Invariance par translation). Montrer que la mesure λ_2 est invariante par translation.

Exercice 2.2 (Propriété d'échelle). Pour tout $\sigma \in \mathbb{R}$ et $A \subset \mathbb{R}^2$, on note σA l'ensemble $\{\sigma x : x \in A\}$. Montrer que la mesure λ_2 a la propriété d'échelle suivante : pour tout $A \in \mathcal{B}(\mathbb{R}^2)$ et $\sigma \in \mathbb{R}$,

$$\lambda_2(\sigma A) = \sigma^2 \lambda_2(A). \quad (2.4)$$

Exercice 2.3 (Mesure du graphe d'une fonction). Soit $f: \mathbb{R} \mapsto \mathbb{R}$ borélienne. On note $\Gamma = \{(x, f(x)) \in \mathbb{R}^2 : x \in \mathbb{R}\}$ le graphe de la fonction f . Montrer que Γ est un borélien de \mathbb{R}^2 et que $\lambda_2(\Gamma) = 0$.

Exercice 2.4. Montrer que le cercle $\mathcal{C} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ est un borélien de \mathbb{R}^2 de mesure nulle par rapport à la mesure de Lebesgue.

2.3 Théorème d'interversion de Fubini

2.3.1 Cas général

Théorème 2.2 (Admis). Soient $(E_1, \mathcal{E}_1, \mu_1)$ et $(E_2, \mathcal{E}_2, \mu_2)$ deux espaces mesurés, μ_1 et μ_2 étant des mesures σ -finies. Soit $f: (E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une application $\mathcal{E}_1 \otimes \mathcal{E}_2 - \mathcal{B}(\mathbb{R})$ mesurable et soit positive, soit intégrable par rapport à $\mu_1 \otimes \mu_2$. Alors les applications $x_1 \mapsto \int_{E_2} f(x_1, x_2) \mu_2(dx_2)$ et $x_2 \mapsto \int_{E_1} f(x_1, x_2) \mu_1(dx_1)$ sont respectivement $\mathcal{E}_1 - \mathcal{B}(\mathbb{R})$ et $\mathcal{E}_2 - \mathcal{B}(\mathbb{R})$ mesurables, et

$$\begin{aligned} \int_{E_1 \times E_2} f(x_1, x_2) (\mu_1 \otimes \mu_2)(dx_1, dx_2) &= \int_{E_1} \left(\int_{E_2} f(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1) \\ &= \int_{E_2} \left(\int_{E_1} f(x_1, x_2) \mu_1(dx_1) \right) \mu_2(dx_2). \end{aligned}$$

Cela signifie que l'on peut, sous de bonnes hypothèses, intégrer une fonction par rapport à une mesure produit en intégrant d'abord par rapport à une des deux variables puis par l'autre, l'ordre n'ayant pas d'importance.

2.3.2 Cas particuliers

Sommes doubles.

Dans le cas particulier où

$$(E_1, \mathcal{E}_1, \mu_1) = (E_2, \mathcal{E}_2, \mu_2) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \delta_{\mathbb{N}}), \quad (2.5)$$

le théorème de Fubini donne :

Corollaire 2.1. Soit $u: \mathbb{N} \times \mathbb{N} \mapsto \mathbb{R}$. Si $u(n, m) \geq 0$ pour tout $(n, m) \in \mathbb{N}^2$ alors

$$\sum_{n,m} u(n, m) = \sum_n \sum_m u(n, m) = \sum_m \sum_n u(n, m). \quad (2.6)$$

Si $u(n, m)$ est de signe quelconque mais que $\sum_{n,m} |u(n, m)| < \infty$ (c'est-à-dire si u est intégrable par rapport à la mesure produit $\delta_{\mathbb{N}} \otimes \delta_{\mathbb{N}}$) alors l'égalité ci-dessus est également vérifiée.

Remarque : dans le cas où u est positive, les séries peuvent diverger, auquel cas les trois membres de l'égalité valent $+\infty$. Si u est intégrable, les trois membres sont bien sûrs finis.

Interversion série-intégrale.

On s'intéresse désormais à la situation où

$$(E_1, \mathcal{E}_1, \mu_1) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda), \quad (E_2, \mathcal{E}_2, \mu_2) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \delta_{\mathbb{N}}). \quad (2.7)$$

Corollaire 2.2. Soit $(f_k)_{k \geq 1}$ une suite de fonctions mesurables sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si $f_k(x) \geq 0$ pour tout $x \in \mathbb{R}$ et pour tout $k \in \mathbb{N}$ alors

$$\int_{\mathbb{R}} \sum_{k \geq 1} f_k(x) \lambda(dx) = \sum_{k \geq 1} \int_{\mathbb{R}} f_k(x) \lambda(dx). \quad (2.8)$$

L'égalité ci-dessus est également vérifiée si

$$\int_{\mathbb{R}} \sum_{k \geq 1} |f_k(x)| \lambda(dx) < \infty, \quad (2.9)$$

c'est-à-dire si f est intégrable par rapport à la mesure produit $\delta_{\mathbb{N}} \otimes \lambda_{\mathbb{R}}$.

Intégrales doubles.

Passons maintenant au cas où

$$(E_1, \mathcal{E}_1, \mu_1) = (E_2, \mathcal{E}_2, \mu_2) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda). \quad (2.10)$$

Corollaire 2.3. Soit $f: (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction mesurable. Si f est positive alors

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx. \quad (2.11)$$

L'égalité ci-dessus est également vérifiée si

$$\int_{\mathbb{R}^2} |f(x, y)| dx dy < \infty, \quad (2.12)$$

c'est-à-dire si f est intégrable par rapport à la mesure produit $\lambda_2 = \lambda \otimes \lambda$.

Remarque 2.3. Dans l'équation (2.11), le $dx dy$ du membre de gauche est une notation simplifiée pour $\lambda_2(dx, dy)$, ou $d\lambda_2$.

Remarque 2.4. Dans les corollaires 2.2 et 2.3, l'intégrale sur \mathbb{R} peut être remplacée par une intégrale sur un intervalle $[a, b]$.

2.3.3 Exemples

Exemple 2.6. Soit $A = \{(x, y) \in \mathbb{R}^2 : y > x > 0\}$. L'intégrale de la fonction $(s, t) \in \mathbb{R}^2 \mapsto \frac{1}{t} e^{-t} \mathbf{1}_A(s, t)$ par rapport à la mesure de Lebesgue vaut 1 (on pourra s'autoriser l'abus de notation $\mathbf{1}_A(s, t) = \mathbf{1}_{\{t > s > 0\}}$).

Passons à un exemple où l'on ne peut pas intervertir :

Exemple 2.7. Calculons

$$\int_{[0,1]^2} \frac{x^2 - y^2}{(x^2 + y^2)^2} dx dy \quad (2.13)$$

de deux façons différentes, en intégrant d'abord par rapport à y puis par rapport à x et ensuite en changeant l'ordre d'intégration. Que peut-on en déduire ? Indice : calculer la dérivée de $\frac{y}{x^2 + y^2}$ par rapport à y .

Chapitre 3

Vecteurs aléatoires

3.1 Introduction

Dans ce chapitre on aborde le cas de l'observation simultanée de plusieurs quantités numériques résultant d'un seul tirage au sort. On commence par développer un cas extrêmement simple qui suffit toutefois à illustrer l'essentiel de ce qui sera présenté ici. Pour simplifier les notations, on se concentrera sur le cas de l'observation de deux quantités aléatoires. Le cas de l'observation d'un nombre fini $d \geq 3$ de quantités numériques résultant d'un seul tirage au sort est une généralisation naturelle du cas particulier $d = 2$.

Une question essentielle pour les applications est celle de l'existence d'un lien entre les composantes d'un vecteur aléatoire. Sont-elles indépendantes ? Si non comment quantifier leur « dépendance » ? Peut-on « prédire » le résultat du tirage au sort d'une variable aléatoire à l'aide du résultat du tirage au sort d'une autre variable aléatoire ?

3.2 Un exemple simple

Un garagiste dispose de 20 voitures à vendre qu'il classe suivant trois catégories de prix et trois catégories d'état. Ainsi chaque voiture est un élément de l'ensemble $\Omega = \{\omega_1, \dots, \omega_{20}\}$. A chacune de ces voitures, c'est à dire à chacun des éléments de Ω on associe une catégorie de prix par une variable aléatoire $X(\omega)$ et une catégorie d'état par une variable aléatoire $Y(\omega)$. Cela se résume par le tableau suivant :

(X, Y)	1	2	3	
1	2	2	3	7
2	1	1	4	6
3	1	1	5	7
	4	4	12	20

On suppose que l'on tire au hasard uniformément l'une des voitures en vente.

— À tout $(x, y) \in \{1, 2, 3\}^2$ on associe $\mathbb{P}((X, Y) = (x, y))$. Par exemple

$$\begin{aligned}\mathbb{P}((X, Y) = (2, 3)) &= \mathbb{P}(\{\omega : (X(\omega), Y(\omega)) = (2, 3)\}) \\ &= \mathbb{P}(\{\omega : X(\omega) = 2\} \cap \{\omega : Y(\omega) = 3\}) \\ &= 4/20.\end{aligned}$$

De la sorte on définit *la loi du couple* ou, pour insister sur le fait que l'on considère un couple, *la loi jointe* de (X, Y) . C'est une probabilité sur un ensemble produit. Il est naturel de se poser à son sujet les questions suivantes*

- est-ce une mesure produit ? Si oui qu'est-ce que cela nous dit sur X et Y ?
- quelle est la signification des intégrales ayant $\mathbb{P}^{(X, Y)}$ comme mesure de référence ?
- On peut aussi décider de ne tenir compte que du prix de la voiture obtenue lors du tirage au sort. Alors à tout $x \in \{1, 2, 3\}$ on associe $\mathbb{P}(X = x)$. Par exemple

$$\begin{aligned} \mathbb{P}(X = 2) &= \mathbb{P}((X, Y) = (2, 1)) + \mathbb{P}((X, Y) = (2, 2)) + \mathbb{P}((X, Y) = (2, 3)) \\ &= \sum_{y=1,2,3} \mathbb{P}((X, Y) = (2, y)) \\ &= 6/20. \end{aligned}$$

De la sorte on définit *la loi marginale de X* . C'est la loi de la variable aléatoire X « seule ».

- On peut aussi décider de ne tenir compte que de l'état de la voiture obtenue lors du tirage au sort. Alors à tout $y \in \{1, 2, 3\}$ on associe $\mathbb{P}(Y = y)$. Par exemple

$$\begin{aligned} \mathbb{P}(Y = 1) &= \mathbb{P}((X, Y) = (1, 1)) + \mathbb{P}((X, Y) = (2, 1)) + \mathbb{P}((X, Y) = (3, 1)) \\ &= \sum_{x=1,2,3} \mathbb{P}((X, Y) = (x, 1)) \\ &= 4/20. \end{aligned}$$

De la sorte on définit *la loi marginale de Y* . C'est la loi de la variable aléatoire Y « seule ».

- On peut aussi décider de ne tenir compte que du prix de la voiture *sachant la catégorie d'état à laquelle elle appartient*. Alors, étant donné $y \in \{1, 2, 3\}$, à tout $x \in \{1, 2, 3\}$ on associe $\mathbb{P}(X = x|Y = y)$. Par exemple

$$\mathbb{P}(X = 2|Y = 1) = \frac{\mathbb{P}(\{X = 2\} \cap \{Y = 1\})}{\mathbb{P}(Y = 1)} = \frac{\mathbb{P}((X, Y) = (2, 1))}{\mathbb{P}(Y = 1)} = 1/4.$$

De la sorte on définit *la loi de X conditionnellement à $Y = y$* . On a vu au chapitre 1 que pour toute valeur de y telle que $\mathbb{P}(Y = y) \neq 0$ l'application $\mathbb{P}(\cdot|Y = y)$ est une probabilité.

- On peut aussi décider de ne tenir compte que de l'état de la voiture *sachant la catégorie de prix à laquelle elle appartient*. Alors, étant donné $x \in \{1, 2, 3\}$, à tout $y \in \{1, 2, 3\}$ on associe $\mathbb{P}(Y = y|X = x)$. Par exemple

$$\mathbb{P}(Y = 2|X = 1) = \frac{\mathbb{P}((X, Y) = (1, 2))}{\mathbb{P}(X = 1)} = 2/7.$$

De la sorte on définit *la loi de Y conditionnellement à $X = x$* .

Ce qu'il est important de retenir de cet exemple est que les deux lois marginales et les lois conditionnelles se déduisent de la donnée de la loi jointe du couple (X, Y) . En revanche la connaissance des lois marginales seules ne suffit pas à « reconstruire » la loi du couple sauf dans certains cas particuliers. En effet, si la description des voitures en vente était donnée par le tableau

(X, Y)	1	2	3	
1	2	1	4	7
2	1	2	3	6
3	1	1	5	7
	4	4	12	20

*, ce n'est pas une liste exhaustive.

cela conduirait aux mêmes lois marginales que dans le premier cas alors que la loi du couple n'est pas la même. Ainsi l'objet le plus important dans cette famille de probabilités est la *loi du couple* et on va voir dans ce qui suit qu'elle obéit aux mêmes règles que les lois « simples » vues jusqu'à présent. Il faudra donc bien se garder de croire que la connaissance des lois marginales, qui sont à première vue plus proches des objets considérés jusqu'à présent que ne l'est la loi du couple, est suffisante pour mesurer le hasard associé à (X, Y) .

3.3 Loi jointe d'un couple

Dans tout ce qui suit $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace de probabilité. On note $\mathcal{B}(\mathbb{R}^2)$ la tribu des boréliens (c'est-à-dire la tribu engendrée par les ouverts) de \mathbb{R}^2 . On admettra que $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

Proposition 3.1. *Soient X et Y deux variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ [†]. Le couple $(X, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ est mesurable.*

Proposition 3.2. *Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$. L'application définie sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ par*

$$\mathbb{P}^{(X,Y)}(A) = \mathbb{P}((X, Y) \in A) = \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\})$$

est une mesure de probabilité. On l'appelle loi jointe du couple (X, Y) [‡].

Définition 3.1. *Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ de loi jointe $\mathbb{P}^{(X,Y)}$. On appelle loi marginale de X la loi \mathbb{P}^X de la variables aléatoire réelle X , qui se déduit de celle du couple (X, Y) par*

$$\begin{aligned} \mathbb{P}^X(A) &= \mathbb{P}(X \in A) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) \\ &= \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A \times \mathbb{R}\}) \\ &= \mathbb{P}^{(X,Y)}(A \times \mathbb{R}). \end{aligned}$$

On introduit à présent la notion d'espérance *d'une fonction mesurable* du couple (X, Y) . On peut aussi donner un sens à l'espérance du couple (X, Y) , ce qui sera fait au paragraphe 3.10.

Définition 3.2. *Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction mesurable.*

1. *Si φ est positive on pose*

$$\mathbb{E}[\varphi(X, Y)] = \int_{\Omega} \varphi(X(\omega), Y(\omega)) d\mathbb{P}.$$

2. *Si φ n'est pas de signe constant mais que $\mathbb{E}[|\varphi(X, Y)|] < \infty$ on pose*

$$\mathbb{E}[\varphi(X, Y)] = \int_{\Omega} \varphi(X(\omega), Y(\omega)) d\mathbb{P}.$$

Ce n'est rien d'autre que la définition 1.16 appliquée à la v.a.r. $\varphi(X, Y)$.

On possède pour les couples aléatoires un résultat analogue au théorème de transfert déjà rencontré pour les variables aléatoires, cf. section 1.4.3.

†. C'est-à-dire deux fonctions mesurables de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

‡. On dit aussi que $\mathbb{P}^{(X,Y)}$ est la mesure de probabilité induite par \mathbb{P} et (X, Y) .

Proposition 3.3 (Admis). Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction mesurable.

1. Si φ est positive on a

$$\begin{aligned}\mathbb{E}[\varphi(X, Y)] &= \int_{\Omega} \varphi(X(\omega), Y(\omega)) d\mathbb{P} \\ &= \int_{\mathbb{R}^2} \varphi(x, y) d\mathbb{P}^{(X, Y)}.\end{aligned}$$

2. Si φ n'est pas de signe constant mais que $\mathbb{E}[|\varphi(X, Y)|] < \infty$ l'égalité ci-dessus est encore vraie.

Comme pour les variables aléatoires cet énoncé admet une sorte de réciproque.

Proposition 3.4. Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$. S'il existe une mesure de probabilité Q sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ telle que pour toute application $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable bornée on ait

$$\mathbb{E}[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) dQ$$

alors nécessairement $Q = \mathbb{P}^{(X, Y)}$.

3.4 Variance et covariance

Définition 3.3. Soient X et Y deux variables aléatoires intégrables telles que XY est intégrable. On appelle covariance de X et Y le nombre réel défini par

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

Si X est de carré intégrable alors on définit

$$\mathbb{V}[X] = \text{Cov}(X, X). \quad (3.1)$$

Remarque 3.1. On verra que la covariance permet, en un certain sens, de mesurer le lien entre X et Y et de développer une géométrie des variables aléatoires.

Remarque 3.2. On verra plus loin que si X et Y sont de carré intégrable alors XY est nécessairement intégrable. C'est une condition suffisante mais ce n'est pas une condition nécessaire.

Proposition 3.5. Soient X, Y et Z des variables aléatoires telles que chacun des termes employés ci-dessous soient bien définis. On a

1. $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$;
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
3. Pour tout $a \in \mathbb{R}$ on a $\text{Cov}(X, a) = 0$;
4. Pour tout $a \in \mathbb{R}$ on a $\text{Cov}(X, aY) = a\text{Cov}(X, Y)$;
5. On a $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$;
6. On a $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}(X, Y)$.

3.5 Variables aléatoires indépendantes

Définition 3.4. Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que X et Y sont indépendantes si et seulement si pour tous $A, B \in \mathcal{B}(\mathbb{R})$ on a

$$\mathbb{P}((X, Y) \in A \times B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

ou, en d'autres termes,

$$\mathbb{P}^{(X, Y)}(A \times B) = \mathbb{P}^X(A)\mathbb{P}^Y(B).$$

Proposition 3.6. Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$. Les variables aléatoires X et Y sont indépendantes si et seulement si $\mathbb{P}^{(X, Y)} = \mathbb{P}^X \otimes \mathbb{P}^Y$.

On a une caractérisation de l'indépendance de deux variables aléatoires qui sera très utile dans les applications.

Proposition 3.7. Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$. Les variables aléatoires X et Y sont indépendantes si et seulement si pour tout couple d'applications mesurables bornées $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}$ on a

$$\mathbb{E}[\varphi(X)\psi(Y)] = \mathbb{E}[\varphi(X)]\mathbb{E}[\psi(Y)].$$

Remarque 3.3. Dans l'égalité de la proposition 3.7 l'espérance de gauche se calcule à l'aide de la loi jointe du couple (X, Y) tandis que chacune des espérances dans le terme de droite se calcule à l'aide des lois marginales.

Une première application de la proposition 3.7 :

Proposition 3.8. Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$. On suppose que X et Y sont indépendantes. Alors pour toutes applications $f, g : \mathbb{R} \rightarrow \mathbb{R}$ mesurables les variables aléatoires $f(X)$ et $g(Y)$ sont indépendantes.

Remarque 3.4. Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$. On suppose que X et Y sont indépendantes et on considère deux applications $f, g : \mathbb{R} \rightarrow \mathbb{R}$ mesurables et pas nécessairement bornées telles que $f(X)$ et $g(Y)$ sont intégrables. Alors la variable aléatoire $f(X)g(Y)$ est intégrable et

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

En effet, on a

$$\begin{aligned} \mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}^2} f(x)g(y) d\mathbb{P}^{(X, Y)} \\ &= \int_{\mathbb{R}^2} f(x)g(y) d(\mathbb{P}^X \otimes \mathbb{P}^Y) \\ &= \left(\int_{\mathbb{R}} f(x) d\mathbb{P}^X \right) \left(\int_{\mathbb{R}} g(y) d\mathbb{P}^Y \right) \\ &= \mathbb{E}[f(X)]\mathbb{E}[g(Y)] \end{aligned}$$

où l'avant-dernière égalité est une conséquence du théorème de Fubini. Par exemple, si X et Y sont indépendantes et intégrables on a $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. En particulier si X et Y sont indépendantes on a $\text{Cov}(X, Y) = 0$. La réciproque est fautive ! Il existe des couples (X, Y) de variables aléatoires dépendantes telles que $\text{Cov}(X, Y) = 0$ [§]. C'est pour cela que ce n'est qu'en un sens « faible » que la covariance mesure la dépendance de deux variables aléatoires.

Proposition 3.9. Soient X et Y deux variables aléatoires indépendantes, de carré intégrable. On a $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$.

Proposition 3.10. Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$. On suppose que X et Y sont indépendantes et de même loi. Alors pour toute application mesurable $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que $\mathbb{E}[|\varphi(X, Y)|] < \infty$ on a

$$\mathbb{E}[\varphi(X, Y)] = \mathbb{E}[\varphi(Y, X)].$$

Remarque 3.5. Remarquons que certaines propositions et définitions s'adaptent au cas où les variables X et Y du couple (X, Y) ne sont pas nécessairement à valeurs dans \mathbb{R} mais chacune dans un espace quelconque muni d'une tribu. Il s'agit des définitions 3.1, 3.2 et 3.4 ainsi que des propositions 3.2, 3.6, 3.7, 3.8 et 3.10.

§. On verra des exemples en TD.

3.6 Inégalités de Hölder et de Cauchy-Scharwz

Dans tout ce paragraphe $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de probabilité donné.

Définition 3.5. Soit p un nombre réel $p \geq 1$. On note \mathcal{L}^p l'ensemble des variables aléatoires définies sur Ω telles que $\mathbb{E}[|X|^p] < \infty$.

Proposition 3.11. « Inégalité de Hölder » Soient p, q deux nombres réels tels que $p, q \geq 1$ liés par $\frac{1}{p} + \frac{1}{q} = 1$. Soient X et Y deux variables aléatoires telles que $X \in \mathcal{L}^p$ et $Y \in \mathcal{L}^q$. On a

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} (\mathbb{E}[|Y|^q])^{\frac{1}{q}}.$$

Indice pour la démonstration. Montrer que pour tous nombres réels $a, b > 0$ et tout $\alpha \in [0, 1]$,

$$a^\alpha b^{1-\alpha} \leq \alpha a + (1 - \alpha)b. \quad (3.2)$$

□

Remarque 3.6. Lorsque $p = q = 2$, on retrouve l'inégalité de Cauchy-Schwarz.

Remarque 3.7. Il découle de l'inégalité de Hölder que si $X \in \mathcal{L}^p$ et $Y \in \mathcal{L}^q$ avec $\frac{1}{p} + \frac{1}{q} = 1$ alors XY est intégrable. En particulier, si X et Y sont de carré intégrable alors XY est intégrable.

Définition 3.6. Soit p un nombre réel $p \geq 1$. La relation binaire définie sur \mathcal{L}^p par $X \sim Y$ si et seulement si $X = Y$ presque sûrement (c'est-à-dire si $\mathbb{P}(X = Y) = 1$) est une relation d'équivalence et on appelle L^p l'ensemble des classes d'équivalence sur \mathcal{L}^p qui lui sont associées.

3.7 Couples de variables discrètes

Dans cette section on suppose données deux variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On suppose également que chacune de ces deux variables X et Y prend ses valeurs dans un ensemble au plus dénombrable : $X \in E = \{x_0, \dots, x_n, \dots\}$ et $Y \in F = \{y_0, \dots, y_n, \dots\}$. Comme $E \times F$ est un ensemble au plus dénombrable on peut voir le couple (X, Y) comme étant une variable aléatoire prenant ses valeurs dans un ensemble au plus dénombrable particulier et développer pour (X, Y) une théorie identique à celle développée pour l'étude d'une seule variable aléatoire à valeurs dans un ensemble au plus dénombrable.

Proposition 3.12. La donnée d'une probabilité \mathbb{P} sur $(E \times F, \mathcal{P}(E \times F))$ est équivalente à la donnée d'une suite $p = (p_{i,j})_{(i,j) \in \mathbb{N}^2}$ de nombres réels positifs ou nuls tels que $\sum_{(i,j) \in \mathbb{N}^2} p_{i,j} = 1$. La relation liant \mathbb{P} et p est la suivante : pour tout $A \in \mathcal{P}(E \times F)$

$$\mathbb{P}(A) = \sum_{(i,j):(x_i, y_j) \in A} p_{i,j}.$$

Ainsi, étant donné un couple aléatoire (X, Y) à valeurs dans $E \times F$ la donnée de sa loi jointe est équivalente à la donnée d'une suite $p^{(X,Y)} = (p_{i,j}^{(X,Y)})_{(i,j) \in \mathbb{N}^2}$ de nombres réels positifs ou nuls liés à $\mathbb{P}^{(X,Y)}$ par

$$p_{i,j}^{(X,Y)} = \mathbb{P}^{(X,Y)}(x_i, y_j) = \mathbb{P}((X, Y) = (x_i, y_j)).$$

En termes de mesures de probabilité on a

$$\mathbb{P}^{(X,Y)} = \sum_{(i,j) \in \mathbb{N}^2} p_{i,j}^{(X,Y)} \delta_{(x_i, y_j)}.$$

Il vient que pour tout $i \in \mathbb{N}$

$$\begin{aligned} \mathbb{P}^X(\{x_i\}) &= \mathbb{P}(X = x_i) \\ &= \mathbb{P}\left(\bigcup_{j \in \mathbb{N}} \{(X, Y) = (x_i, y_j)\}\right) \\ &= \sum_{j \in \mathbb{N}} \mathbb{P}((X, Y) = (x_i, y_j)) \\ &= \sum_{j \in \mathbb{N}} p_{i,j}^{(X,Y)} \end{aligned}$$

et de même pour tout $j \in \mathbb{N}$ on a $\mathbb{P}^Y(\{y_j\}) = \sum_{i \in \mathbb{N}} p_{i,j}^{(X,Y)}$. On peut alors exprimer les propriétés de la loi du couple (X, Y) à l'aide de la suite $p^{(X,Y)}$.

Proposition 3.13. *Soit (X, Y) un couple de variables aléatoires à valeurs dans $E \times F$ dont la loi est donnée par $p^{(X,Y)} = (p_{i,j}^{(X,Y)})_{(i,j) \in \mathbb{N}^2}$. Les variables aléatoires X et Y sont indépendantes si et seulement si pour tous $i, j \in \mathbb{N}$ on a $p_{i,j}^{(X,Y)} = p_i^X p_j^Y$.*

Proposition 3.14. *Soit $\varphi : E \times F \rightarrow \mathbb{R}$ et (X, Y) un couple de variables aléatoires à valeurs dans $E \times F$ de loi $p^{(X,Y)}$.*

1. Si φ est positive ou nulle on a

$$\mathbb{E}[\varphi(X, Y)] = \sum_{(i,j) \in \mathbb{N}^2} \varphi(x_i, y_j) p_{i,j}^{(X,Y)}.$$

2. Si $\mathbb{E}[|\varphi(X, Y)|] < \infty$ alors

$$\mathbb{E}[\varphi(X, Y)] = \sum_{(i,j) \in \mathbb{N}^2} \varphi(x_i, y_j) p_{i,j}^{(X,Y)}.$$

3.8 Couples à densité

Définition 3.7. *On appelle densité jointe toute fonction f définie sur \mathbb{R}^2 et à valeurs dans \mathbb{R} qui vérifie les conditions suivantes :*

1. f est mesurable ;
2. f est positive (sauf éventuellement sur une partie négligeable de \mathbb{R}^2) ;
3. $\int_{\mathbb{R}^2} f d(\lambda \otimes \lambda) = 1$ où λ est la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Remarque 3.8. *En général, on écrit simplement $\int_{\mathbb{R}^2} f(x, y) dx dy$ à la place de $\int_{\mathbb{R}^2} f d(\lambda \otimes \lambda)$.*

En pratique, les fonctions de densité jointe rencontrées dans ce cours seront continues (éventuellement par morceaux), ce qui garantit leur mesurabilité.

Proposition 3.15. *Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une densité jointe. L'application Q définie sur $\mathcal{B}(\mathbb{R}^2)$ par*

$$Q(A) = \int_{\mathbb{R}^2} \mathbf{1}_A f d(\lambda \otimes \lambda) = \int_A f(x, y) dx dy.$$

est une mesure de probabilité sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$.

Définition 3.8. *On dit qu'un couple aléatoire (X, Y) est un couple continu ou encore un couple à densité si et seulement s'il existe une densité jointe $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que pour tout $A \in \mathcal{B}(\mathbb{R}^2)$ on a*

$$\mathbb{P}((X, Y) \in A) = \int_{\mathbb{R}^2} \mathbf{1}_A(x, y) f(x, y) dx dy.$$

Si (X, Y) est un couple admettant f pour densité alors pour tout $A \in \mathcal{B}(\mathbb{R})$:

$$\begin{aligned} \mathbb{P}(X \in A) &= \mathbb{P}((X, Y) \in A \times \mathbb{R}) \\ &= \int_{\mathbb{R}^2} \mathbf{1}_{A \times \mathbb{R}}(x, y) f(x, y) d(\lambda \otimes \lambda) \\ &= \int_{\mathbb{R}^2} \mathbf{1}_A(x) \mathbf{1}_{\mathbb{R}}(y) f(x, y) d(\lambda \otimes \lambda) \\ &= \int_{\mathbb{R}^2} \mathbf{1}_A(x) f(x, y) d(\lambda \otimes \lambda) \\ &= \int_{\mathbb{R}} \mathbf{1}_A(x) \left(\int_{\mathbb{R}} f(x, y) \lambda(dy) \right) \lambda(dx). \end{aligned}$$

Ainsi, si f est une densité jointe pour (X, Y) alors $x \rightarrow f^X(x) = \int_{\mathbb{R}} f(x, y) \lambda(dy)$ est une densité pour X . On dira que f^X est la *densité marginale* de X . On définit de même la densité marginale de Y en prenant

$$f^Y(y) = \int_{\mathbb{R}} f(x, y) \lambda(dx).$$

Remarque 3.9. *On peut construire des exemples où X et Y sont des variables aléatoires à densité mais (X, Y) n'est pas un couple à densité. Cette remarque est à rapprocher de celle que l'on a faite au tout début de ce chapitre : la connaissance des marginales ne renseigne pas sur la loi du couple.*

Proposition 3.16. *Soit $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ une application mesurable et (X, Y) un couple de variables aléatoires à valeurs dans \mathbb{R}^2 admettant f comme densité jointe.*

1. Si φ est positive ou nulle on a

$$\mathbb{E}[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) f(x, y) d(\lambda \otimes \lambda).$$

2. Si $\int_{\mathbb{R}^2} |\varphi(x, y)| f(x, y) d(\lambda \otimes \lambda) < \infty$ alors

$$\mathbb{E}[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) f(x, y) d(\lambda \otimes \lambda).$$

On peut caractériser la densité des couples à densité à l'aide de la proposition suivante :

Proposition 3.17. *Soit (X, Y) un couple aléatoire et f une densité jointe. Le couple (X, Y) admet f pour densité si et seulement si pour toute application $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable bornée on a*

$$\mathbb{E}[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) f(x, y) d(\lambda \otimes \lambda).$$

Ce dernier résultat est très important car il permet en pratique de déterminer la loi de certains couples à densité comme on le verra à travers un exemple à la fin de ce paragraphe.

Proposition 3.18 (admise en partie). *Soit (X, Y) un couple de densité jointe $f^{(X, Y)}$. Les variables aléatoires X et Y sont indépendantes si et seulement si $f^X \otimes f^Y : (x, y) \mapsto f^X(x) f^Y(y)$ est une densité jointe pour le couple (X, Y) .*

3.9 Calcul de la densité d'un couple

Supposons que l'on se donne un couple densité (X, Y) et une fonction $\varphi : \mathbb{R}^2 \mapsto \mathbb{R}^2$. Le but de ce chapitre est de déterminer la densité du couple $\varphi(X, Y)$, sous de bonnes conditions sur φ garantissant l'existence d'une telle densité.

3.9.1 Rappels en une dimension

Commençons par quelques rappels dans le cas où on ne s'intéresse non pas à un couple mais à une seule variable aléatoire réelle.

Soit X une v.a.r à valeurs dans un ouvert U , de densité f . Soit $\varphi : U \rightarrow V$, avec V ouvert de \mathbb{R} . Supposons de plus que φ soit de classe \mathcal{C}_1 sur U et que φ' ne s'annule pas. On souhaite alors déterminer la loi de la v.a.r $Y = \varphi(X)$. Soit $\psi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable bornée. On a, d'après la proposition 1.1,

$$\mathbb{E}(\psi(Y)) = \mathbb{E}(\psi(\varphi(X))) = \int_U \psi(\varphi(x))f(x)dx. \quad (3.3)$$

D'après la formule du changement de variable vue en Analyse,

$$\int_U \psi(\varphi(x))f(x)dx = \int_V \psi(y)|\varphi'(\varphi^{-1}(y))|^{-1}f(\varphi^{-1}(y))dy. \quad (3.4)$$

Finalement, nous venons d'établir que

$$\mathbb{E}(\psi(Y)) = \int_V \psi(y)|\varphi'(\varphi^{-1}(y))|^{-1}f(\varphi^{-1}(y))dy, \quad (3.5)$$

ce qui signifie, en vertu de la proposition 1.2, que Y est une v.a.r de densité

$$y \in V \mapsto |\varphi'(\varphi^{-1}(y))|^{-1}f(\varphi^{-1}(y)). \quad (3.6)$$

Exemple 3.1. Soit U une variable aléatoire suivant la loi uniforme sur l'intervalle $[0, 1]$. Déterminer la loi de U^2 . Même question si U suit la loi uniforme sur $[-1, 1]$.

Ainsi, la formule du changement de variable permet de trouver la densité, et donc la loi, de $\varphi(X)$ à partir de celle de X . Nous allons donc introduire, dans la section suivante, une version bi-dimensionnelle de la formule de changement de variable.

3.9.2 Formule du changement de variable

Voyons premièrement comment « dériver » une fonction ayant \mathbb{R}^2 comme espace de départ.

On peut commencer par fixer une des deux variables et dériver par rapport à l'autre, ce qui donne la notion de dérivée partielle :

Définition 3.1 (Dérivée partielle). Soit U un ouvert de \mathbb{R} et $\varphi : (x, y) \in U \mapsto \varphi(x, y) \in \mathbb{R}$. La dérivée partielle de U par rapport à la première variable x (resp. la deuxième variable y), notée $\frac{\partial \varphi}{\partial x}$ (resp. $\frac{\partial \varphi}{\partial y}$) est la fonction, si elle existe,

$$\begin{aligned} (x, y) &\mapsto [x \mapsto \varphi(x, y)]'(x) = [\varphi(\cdot, y)]'(x) \\ (\text{resp. } (x, y) &\mapsto [y \mapsto \varphi(x, y)]'(y) = [\varphi(x, \cdot)]'(y) \end{aligned} \quad (3.7)$$

On peut également considérer les deux variables simultanément, ce qui donne la notion de différentielle :

Définition 3.2 (Différentielle, matrice jacobienne et jacobien). Soient U et V deux ouverts de \mathbb{R}^2 et $\varphi : (x, y) \in U \mapsto (\varphi_1(x, y), \varphi_2(x, y)) \in V$. La différentielle de φ en $(x, y) \in U$, notée $D\varphi(x, y)$, est l'application linéaire, si elle existe, vérifiant pour tout (u, v) dans un voisinage de 0,

$$\varphi(x + u, y + v) = \varphi(x, y) + D\varphi(x, y)(u, v) + o(\|(u, v)\|). \quad (3.8)$$

Si la différentielle en (x, y) existe alors on dit que φ est différentiable en (x, y) et la matrice canoniquement associée à $D\varphi(x, y)$ est alors appelée matrice jacobienne. L'application φ est dite différentiable en U si elle est différentiable en tout point de U . Le déterminant de $D\varphi(x, y)$, noté $\text{Jac}_\varphi(x, y)$, est appelé jacobien de φ en (x, y) .

Définition 3.3. Soient U et V deux ouverts de \mathbb{R}^2 et $\varphi: U \mapsto V$. On dit que φ est de classe \mathcal{C}_1 si φ est différentiable en U et si l'application $D\varphi: (x, y) \in U \mapsto D\varphi(x, y)$, appelée application différentielle, est continue.

Il y a une relation entre différentielle et dérivées partielles :

Proposition 3.1. Soient U et V deux ouverts de \mathbb{R}^2 et $\varphi: (x, y) \in U \mapsto (\varphi_1(x, y), \varphi_2(x, y)) \in V$. L'application φ est de classe \mathcal{C}_1 ssi les dérivées partielles $\frac{\partial \varphi_i}{\partial x}$ et $\frac{\partial \varphi_i}{\partial y}$ existent et sont continues sur U pour tout $i \in \{1, 2\}$. De plus,

$$D\varphi(x, y) = \begin{pmatrix} \frac{\partial \varphi_1}{\partial x}(x, y) & \frac{\partial \varphi_1}{\partial y}(x, y) \\ \frac{\partial \varphi_2}{\partial x}(x, y) & \frac{\partial \varphi_2}{\partial y}(x, y) \end{pmatrix}. \quad (3.9)$$

Voici enfin la formule du changement de variable, version « bi-dimensionnelle » :

Théorème 3.1 (admis en partie). Soient U et V deux ouverts de \mathbb{R}^2 et $\varphi: (x, y) \in U \mapsto (\varphi_1(x, y), \varphi_2(x, y)) \in V$ une bijection de classe \mathcal{C}_1 telle que Jac_φ ne s'annule pas sur U . Pour toute fonction $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable et intégrable sur U , on a

$$\int_U g(x, y) dx dy = \int_V g(\varphi^{-1}(s, t)) |\text{Jac}_\varphi(\varphi^{-1}(s, t))|^{-1} ds dt. \quad (3.10)$$

Voici d'autres formulations du théorème, en reprenant les mêmes hypothèses sur φ :

1. Pour toute fonction $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable et intégrable sur U , on a

$$\int_U g(x, y) dx dy = \int_V g(\varphi^{-1}(s, t)) |\text{Jac}_{\varphi^{-1}}(s, t)| ds dt. \quad (3.11)$$

2. Pour toute fonction $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable et telle que $g \circ \varphi$ est intégrable sur U ,

$$\int_U g(\varphi(x, y)) dx dy = \int_V g(s, t) |\text{Jac}_\varphi(\varphi^{-1}(s, t))|^{-1} ds dt. \quad (3.12)$$

3. Pour toute fonction $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable et intégrable sur V ,

$$\int_U g(\varphi(x, y)) |\text{Jac}_\varphi(x, y)| dx dy = \int_V g(s, t) ds dt. \quad (3.13)$$

3.9.3 Exemple

Voici un exemple de calcul de densité d'un couple. Soient X et Y deux variables aléatoires réelles indépendantes, X étant de loi exponentielle de paramètre $\alpha > 0$ et Y de loi exponentielle de paramètre $\beta > 0$. On suppose que $\alpha \neq \beta$. On rappelle qu'une variable aléatoire réelle est de loi exponentielle de paramètre $\alpha > 0$ si elle admet $f(x) = \alpha e^{-\alpha x} \mathbf{1}_{]0, +\infty[}(x)$ comme densité. On cherche à calculer la loi du couple $(X + Y, Y)$ **[Exemple traité en cours]**.

Application : formule de convolution pour la somme de deux v.a.r. à densité indépendantes.

3.10 Vecteurs aléatoires, $d \geq 3$

3.10.1 Produit de plus de deux mesures

Soit $d \geq 2$. Pour tout $1 \leq i \leq d$, on se donne un ensemble E_i muni d'une tribu \mathcal{E}_i . La tribu $\mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \dots \otimes \mathcal{E}_d$ est la tribu engendrée par les pavés de $E_1 \times \dots \times E_d$, c'est-à-dire les ensembles de la forme $A_1 \times \dots \times A_d$, où $A_i \in \mathcal{E}_i$ pour tout $1 \leq i \leq d$.

Le produit de tribus est *associatif*. Cela signifie que pour tout $1 \leq i < d$,

$$\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_d = (\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_i) \otimes (\mathcal{E}_{i+1} \otimes \dots \otimes \mathcal{E}_d). \quad (3.14)$$

Toutefois, il n'est pas commutatif.

Supposons désormais que chaque espace (E_i, \mathcal{E}_i) soit munie d'une mesure σ -finie, notée μ_i . La mesure produit $\mu_1 \otimes \dots \otimes \mu_d$ est définie comme l'unique mesure sur $(E_1 \times \dots \times E_d, \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_d)$ vérifiant

$$(\mu_1 \otimes \dots \otimes \mu_d)(A_1 \times \dots \times A_d) = \prod_{i=1}^d \mu_i(A_i), \quad (3.15)$$

pour tout $A_1 \in \mathcal{E}_1, \dots, A_d \in \mathcal{E}_d$. L'existence est admise : l'unicité découle du théorème de Dynkin. Comme pour le produit de tribus, le produit de mesures est associatif mais non commutatif.

3.10.2 Matrice de dispersion

Soit $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ un vecteur aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ c'est-à-dire que X_1, \dots, X_d sont des variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. Bien qu'on les écrive sous forme de vecteurs ligne (pour gagner de la place), les vecteurs aléatoires devront toujours être considérés comme des vecteurs colonne. Si chaque terme est défini, on introduit les notations

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$$

qui est lui aussi, malgré les apparences, un vecteur colonne et

$$D_X = \begin{pmatrix} \mathbb{V}[X_1] & & \text{Cov}(X_1, X_d) \\ & \ddots & \\ \text{Cov}(X_d, X_1) & & \mathbb{V}[X_d] \end{pmatrix}$$

la matrice de dispersion ou matrice de variance/covariance du vecteur X . Cet objet généralise la notion de variance. C'est une matrice symétrique (pourquoi?) de type (d, d) de terme général $\text{Cov}(X_i, X_j)$. De façon plus générale si

$$U = \begin{pmatrix} U_{1,1} & & U_{1,n} \\ & \ddots & \\ U_{m,1} & & U_{m,n} \end{pmatrix}$$

est une matrice aléatoire à m lignes et n colonnes c'est à dire une famille de $m \times n$ variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ on note

$$\mathbb{E}[U] = \begin{pmatrix} \mathbb{E}[U_{1,1}] & & \mathbb{E}[U_{1,n}] \\ & \ddots & \\ \mathbb{E}[U_{m,1}] & & \mathbb{E}[U_{m,n}] \end{pmatrix}.$$

Proposition 3.19. 1. Soit R une matrice déterministe quelconque de type (m, n) et X une matrice aléatoire quelconque de type (m, n) . On a

$$\mathbb{E}[R + X] = R + \mathbb{E}[X].$$

2. Soit A une matrice déterministe quelconque de type (k, m) et X une matrice aléatoire quelconque de type (m, n) . On a

$$\mathbb{E}[AX] = A\mathbb{E}[X].$$

3. Soit B une matrice déterministe quelconque de type (n, d) et X une matrice aléatoire quelconque de type (m, n) . On a

$$\mathbb{E}[XB] = \mathbb{E}[X]B.$$

4. Soit A une matrice déterministe quelconque de type (k, m) , X une matrice aléatoire quelconque de type (m, n) et B une matrice déterministe quelconque de type (n, d) . On a

$$\mathbb{E}[AXB] = A\mathbb{E}[X]B.$$

5. Soit A une matrice déterministe quelconque de type (k, m) , X un vecteur aléatoire quelconque de taille m et B un vecteur déterministe quelconque de taille k . On a

$$\mathbb{E}[AX + B] = A\mathbb{E}[X] + B.$$

Pour toute matrice A on notera A^* sa transposée.

Proposition 3.20. Soit X un vecteur aléatoire de taille d , de matrice de dispersion D_X .

1. On a

$$D_X = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^*].$$

2. Soit A une matrice déterministe à k lignes et d colonnes. On a $D_{AX} = AD_X A^*$.

3. Pour tout $u \in \mathbb{R}^d$ on a

$$\mathbb{V}[u^*X] = u^*D_X u.$$

3.10.3 Indépendance

Définition 3.9. On dit que les composantes X_1, \dots, X_d d'un vecteur $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ sont indépendantes (ou mutuellement indépendantes) si et seulement si pour tous $A_1 \in \mathcal{B}(\mathbb{R}), \dots, A_d \in \mathcal{B}(\mathbb{R})$ on a

$$\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_d \in A_d).$$

En d'autres termes, X_1, \dots, X_d sont indépendantes si et seulement si

$$\mathbb{P}^{(X_1, \dots, X_d)} = \mathbb{P}^{X_1} \otimes \dots \otimes \mathbb{P}^{X_d}.$$

Remarque 3.1. Toute sous-famille d'une famille de variables aléatoires indépendantes constitue elle-même une famille de variables aléatoires indépendantes.

Lemme 3.1 (« Lemme des paquets »). Soit $X = (X_1, \dots, X_d)$ un vecteur dont les composantes sont indépendantes. Soit $I = \{i_1, \dots, i_n\}$ et $J = \{j_1, \dots, j_m\}$ deux sous-ensembles non vides et disjoints de $\{1, \dots, d\}$. Les vecteurs $X^{(I)} = (X_{i_1}, \dots, X_{i_n})$ et $X^{(J)} = (X_{j_1}, \dots, X_{j_m})$ sont indépendants, c'est-à-dire

$$\mathbb{P}(X^{(I)} \in A, X^{(J)} \in B) = \mathbb{P}(X^{(I)} \in A)\mathbb{P}(X^{(J)} \in B) \quad (3.16)$$

pour tout $A \in \mathcal{B}(\mathbb{R}^n)$ et $B \in \mathcal{B}(\mathbb{R}^m)$.

Remarque 3.2. Nous admettrons que ce lemme se généralise à un nombre quelconque de paquets disjoints.

Enfin, nous rencontrerons à partir du prochain chapitre des *suites* de variables aléatoires. Cela motive la définition suivante :

Définition 3.10. Soit $(X_i)_{i \in I}$ une famille quelconque de variables aléatoires. On dit que ces variables sont indépendantes si pour tout sous-ensemble $J \subset I$ fini, les composantes du vecteur $(X_j)_{j \in J}$ sont indépendantes.

Notons que la remarque 3.1 et le lemme des paquets s'emploient également dans ce cadre plus général.

3.10.4 Indépendance affine

On dispose d'une autre notion d'indépendance qui ne repose que sur des liens entre variables aléatoires qui seraient linéaires et d'un critère simple pour vérifier que les composantes d'un vecteur sont indépendantes en ce sens.

Définition 3.11. On dit que les composantes X_1, \dots, X_d d'un vecteur $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ sont affinement dépendantes s'il existe des constantes réelles $\alpha_1, \dots, \alpha_d, \beta$ non toutes nulles telles que $\alpha_1 X_1 + \dots + \alpha_d X_d = \beta$ presque sûrement. Si X_1, \dots, X_d ne sont pas affinement dépendantes on dit qu'elles sont affinement indépendantes .

Proposition 3.21. Les composantes X_1, \dots, X_d d'un vecteur $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ de matrice de dispersion D_X sont affinement indépendantes si et seulement si D_X est inversible.

Remarque 3.10. Il est clair que des variables aléatoires peuvent être dépendantes sans être affinement dépendantes. Par exemple si U est une variable aléatoire de loi uniforme sur $[0, 1]$ alors U et U^2 sont dépendantes mais elles ne sont pas affinement dépendantes.

Chapitre 4

Théorèmes limites

4.1 Introduction

Une même expérience aléatoire répétée un grand nombre de fois avec des résultats indépendants les uns des autres présente une grande irrégularité au niveau des résultats individuels et, *dans le même temps*, donne lieu à des fréquences d'apparition qui sont elles extrêmement régulières. C'est un fait qui semble avoir été connu bien avant le développement d'une quelconque théorie des probabilités. « Réciproquement », que la théorie mathématique des probabilités permette de *prouver* l'existence de tels phénomènes donne du crédit à l'idée que cette théorie est un outil efficace de modélisation de phénomènes aléatoires. Dans ce chapitre on s'intéresse à de tels résultats.

On va donc considérer des suites de variables aléatoires indépendantes. On a vu que, formellement, les variables aléatoires sont des fonctions. Ainsi, les théorèmes limites exposés ici sont des résultats sur des limites de suites de fonction. Comme ces fonctions sont définies sur des structures mathématiques très riches, il y a plusieurs façons de définir leur convergence, ce qui nous occupera au paragraphe 4.4. On y verra aussi qu'il existe une sorte de hiérarchie entre ces modes de convergence. Au paragraphe 4.5 on verra les résultats de référence en la matière et au paragraphe 4.6 on verra comment combiner différents résultats de convergence pour en obtenir de nouveaux.

4.2 Suites de variables indépendantes

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles définies sur Ω .

Définition 4.1. *On dit d'une suite de variables aléatoires $(X_n)_{n \geq 1}$ qu'elle est une suite de variables aléatoires indépendantes si et seulement si pour tout $m \in \mathbb{N}^*$ et toute famille i_1, \dots, i_m d'entiers non-nuls distincts les variables aléatoires X_{i_1}, \dots, X_{i_m} sont indépendantes.*

Remarque 4.1. *Le simple fait qu'il existe des suites de variables aléatoires indépendantes n'est pas une évidence, mais nous l'admettons.*

4.3 Fonctions caractérisant la loi d'une variable ou d'un vecteur aléatoire

4.3.1 Fonction caractéristique d'une variable aléatoire réelle

Dans ce qui suit, X est une variable aléatoire réelle.

Définition 4.2. On appelle fonction caractéristique de X la fonction φ_X définie sur \mathbb{R} par $\varphi_X(t) = \mathbb{E}[e^{itX}]$.

Remarque 4.2. L'application $u \mapsto e^{itu}$ étant de module 1, pour toute variable aléatoire X la fonction caractéristique φ_X de X est bien définie et finie sur \mathbb{R} tout entier.

Proposition 4.1. Soient X et Y deux variables aléatoires réelles dont les fonctions caractéristiques coïncident. Alors X et Y ont même loi.

Proposition 4.2. La régularité de φ_X est liée à l'intégrabilité de X de la façon suivante :

1. Si X est intégrable alors φ_X est de classe \mathcal{C}^1 et $\mathbb{E}[X] = -i\varphi_X'(0)$.
2. Si X est de carré intégrable alors φ_X est de classe \mathcal{C}^2 et $\mathbb{E}[X^2] = -\varphi_X''(0)$.

Proposition 4.3. Pour tout $t \in \mathbb{R}$ et toutes constantes réelles a et b on a $\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at)$.

4.3.2 Fonction caractéristique d'un vecteur aléatoire

On peut étendre la notion de fonction caractéristique d'une variable aléatoire à un vecteur aléatoire.

Définition 4.3. Soit $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ un vecteur aléatoire de taille d . On appelle fonction caractéristique de X la fonction φ_X définie sur \mathbb{R}^d par

$$t = (t_1, \dots, t_d) \rightarrow \varphi_X(t) = \mathbb{E}[e^{it^*X}] = \mathbb{E}\left[\exp\left(i \sum_{k=1}^d t_k X_k\right)\right].$$

Comme dans le cas $d = 1$ la fonction caractéristique caractérise la loi d'un vecteur au sens où

Proposition 4.4. Soient $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ et $Y = (Y_1, \dots, Y_d) \in \mathbb{R}^d$ deux vecteurs aléatoires de même taille. Les vecteurs X et Y ont même loi si et seulement si pour tout $t \in \mathbb{R}^d$ on a $\varphi_X(t) = \varphi_Y(t)$.

Quand $d \geq 2$ la fonction caractéristique d'un vecteur nous renseigne sur la dépendance des composantes du vecteur.

Proposition 4.5. Les composantes X_1, \dots, X_d d'un vecteur $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ sont indépendantes si et seulement si pour tout $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ on a

$$\varphi_X(t) = \varphi_{X_1}(t_1) \cdots \varphi_{X_d}(t_d).$$

4.4 Quatre types de convergence

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles définies sur Ω et X une variable aléatoire réelle définie sur Ω elle aussi. On peut donner plusieurs sens à l'expression « la suite $(X_n)_{n \geq 1}$ converge vers X ».

Définition 4.4. On dit que la suite $(X_n)_{n \geq 1}$ converge presque sûrement vers X lorsque

$$\mathbb{P}(\{\omega \in \Omega \text{ tels que } X_n(\omega) \not\rightarrow X(\omega)\}) = 0$$

et on note $X_n \xrightarrow{p.s.} X$.

Remarque 4.3. La convergence presque sûre est donc la convergence simple sur tout Ω , sauf sur une partie négligeable, de la suite de fonctions X_n vers X .

Exemple 4.1. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires de même loi de Bernoulli de paramètre $p \in]0, 1[$ et $(S_n)_{n \geq 1}$ la suite de variables aléatoires définie par $S_n = \sum_{i=1}^n X_i$. On pose $Z_n = \frac{S_n}{n^2}$. Quelle est la limite presque-sûre de la suite (Z_n) ?

Définition 4.5. On dit que la suite $(X_n)_{n \geq 1}$ converge en probabilité vers X lorsque pour tout $\varepsilon > 0$ on a

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

et on note $X_n \xrightarrow{\mathbb{P}} X$.

Remarque 4.4. S'il existe un ε_0 tel que pour tout $0 < \varepsilon < \varepsilon_0$ on a

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

alors $X_n \xrightarrow{\mathbb{P}} X$ puisque pour tout $\varepsilon \geq \varepsilon_0$ on a

$$\{|X_n - X| > \varepsilon\} \subset \{|X_n - X| > \varepsilon_0/2\}$$

donc

$$0 \leq \mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon_0/2).$$

Exemple 4.2. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires telle que pour tout $n \geq 1$ on a $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$ et $\mathbb{P}(X_n = \frac{1}{n}) = \frac{1}{n}$. La suite (X_n) converge-t-elle en probabilité ?

Exemple 4.3. On considère une suite $(X_n)_{n \geq 1}$ de variables aléatoires indépendantes et identiquement distribuées qui ont pour loi la probabilité uniforme sur l'intervalle $[0, \theta]$, avec $\theta > 0$. La suite $(\max(X_1, \dots, X_n))_{n \geq 1}$ converge en probabilité vers θ .

Définition 4.6. On dit que la suite $(X_n)_{n \geq 1}$ converge en loi vers X lorsque pour tout réel x tel que $\mathbb{P}(X = x) = 0$ on a

$$\mathbb{P}(X_n \leq x) \rightarrow \mathbb{P}(X \leq x)$$

et on note $X_n \xrightarrow{L} X$.

On dispose de deux autres définitions équivalentes pour la convergence en loi

Proposition 4.6. La suite $(X_n)_{n \geq 1}$ converge en loi vers X si et seulement si pour toute application à valeurs réelles f définie sur \mathbb{R} , continue et bornée on a

$$\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)].$$

Proposition 4.7. La suite $(X_n)_{n \geq 1}$ converge en loi vers X si et seulement si la suite $(\varphi_n)_{n \geq 1}$ des fonctions caractéristiques des variables aléatoires X_n converge simplement vers la fonction caractéristique φ de X : pour tout $t \in \mathbb{R}$ on a

$$\varphi_n(t) = \mathbb{E}[e^{itX_n}] \rightarrow \varphi(t) = \mathbb{E}[e^{itX}].$$

Exemple 4.4. Pour tout entier $n \geq 1$ on considère X_n une variable aléatoire de loi uniforme sur

$$B_n = \left\{ x \in \mathbb{R} : \exists k \in \{1, \dots, n\} \text{ tel que } x = \frac{k}{n} \right\}.$$

Déterminer la limite en loi de la suite (X_n) .

Il y a une hiérarchie dans les différents types de convergence.

Proposition 4.8. Si la suite $(X_n)_{n \geq 1}$ converge presque sûrement vers X alors elle converge aussi en probabilité vers X .

Remarque 4.5. *La réciproque n'est pas vraie. Par exemple, si $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes telles que $P(X_n = 0) = 1 - \frac{1}{n}$ et $P(X_n = 1) = \frac{1}{n}$ alors $(X_n)_{n \geq 1}$ tend vers 0 en probabilité, mais pas presque-sûrement.*

Proposition 4.9. *Si la suite $(X_n)_{n \geq 1}$ converge en moyenne quadratique vers X alors elle converge aussi en probabilité vers X .*

Proposition 4.10. *Si la suite $(X_n)_{n \geq 1}$ converge en probabilité vers X alors elle converge aussi en loi vers X .*

On voit donc que la convergence en loi est le mode de convergence le plus faible.

4.5 Théorèmes limites fondamentaux

4.5.1 Lois des grands nombres

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $(X_n)_{n \geq 1}$ une suite de variables aléatoires définies sur Ω et à valeurs dans \mathbb{R} . **Dans tout ce qui suit on suppose ces variables aléatoires indépendantes et ayant toutes même loi.** Pour tout entier n on note

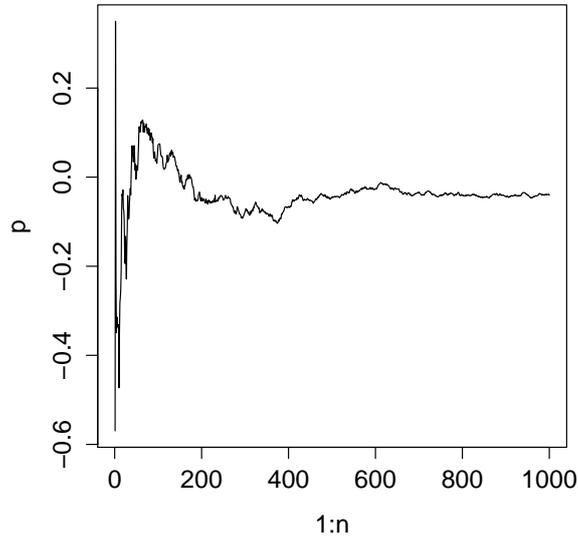
$$\bar{X}_n = \frac{1}{n} (X_1 + \cdots + X_n)$$

la moyenne empirique de X_1, \dots, X_n . La loi des grands nombres porte sur le comportement de \bar{X}_n quand $n \rightarrow \infty$. On trouve dans la littérature deux versions de la loi des grands nombres. C'est une anomalie puisque les conditions de la loi faible des grands nombres sont plus contraignantes que celles de la loi forte alors que sa conclusion est plus faible. Cette anomalie trouve son origine dans le fait qu'il est facile de prouver la loi faible mais assez difficile de prouver la loi forte.

Théorème 4.1. « **Loi faible des grands nombres** » *Si $\mathbb{E}[X_1^2] < \infty$ la suite de variables aléatoires $(\bar{X}_n)_{n \in \mathbb{N}}$ converge en probabilité vers $m = \mathbb{E}[X_1]$.*

Théorème 4.2. « **Loi forte des grands nombres** » *Si $\mathbb{E}[|X_1|] < \infty$ la suite de variables aléatoires $(\bar{X}_n)_{n \in \mathbb{N}}$ converge presque sûrement vers $m = \mathbb{E}[X_1]$.*

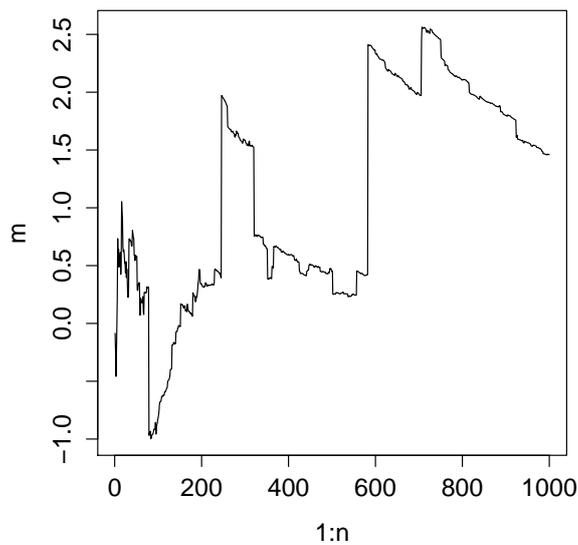
Ainsi, si on considère une suite $(X_n(\omega))_{n \geq 1}$ d'observations faite sur le résultat du tirage au sort d'une suite de variables aléatoires $(X_n)_{n \geq 1}$ indépendantes et de même loi intégrable, la suite $(\bar{X}_n(\omega))_{n \geq 1}$ converge.



Dans le graphe ci-dessus on a pris la loi $\mathcal{N}(0, 1)$ comme loi commune des X_i . Que se passe-t-il donc quand $\mathbb{E}[|X_1|] = \infty$? Prenons l'exemple de la loi de Cauchy, de densité $f(x) = \frac{1}{\pi(1+x^2)}$ sur \mathbb{R} tout entier. On a

$$\mathbb{E}[|X|] = \int_{\mathbb{R}} \frac{|x|}{\pi(1+x^2)} dx = \infty, \quad (4.1)$$

puisque en $+\infty$ et $-\infty$ on a $|x|f(x) \sim \frac{1}{\pi x}$ qui n'est pas intégrable. En d'autres mots, la fonction $x \rightarrow \frac{\pi}{1+x^2}$ ne décroît pas assez vite/assez fort en ∞ pour que $|x|f(x)$ soit intégrable, c'est-à-dire que la probabilité sous la loi de Cauchy d'observer de fortes valeurs est trop importante pour que la suite $(\bar{X}_n(\omega))_{n \geq 1}$ se stabilise. C'est ce que l'on observe sur le graphique suivant



4.5.2 Le théorème de la limite centrale

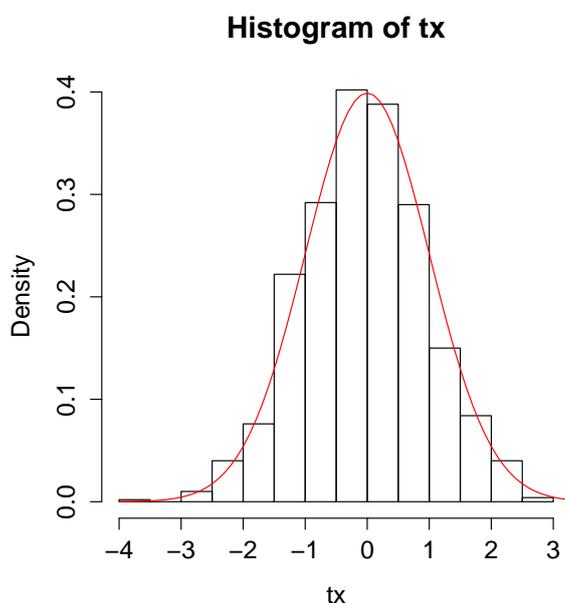
Le théorème de la limite centrale est une sorte de raffinement* de la loi des grands nombres. On suppose toujours que les variables aléatoires $(X_n)_{n \geq 1}$ sont indépendantes et de même loi.

Théorème 4.3. « Théorème de la limite centrale » *On suppose que X_1 est de carré intégrable et que $\sigma^2 = \mathbb{V}[X_1]$ est non nul. La suite de variables aléatoires*

$$\sqrt{n} \frac{(\bar{X}_n - m)}{\sigma}$$

converge en loi vers une variable aléatoire X de loi $\mathcal{N}(0, 1)$.

Dans le graphique ci-dessous on a superposé le tracé de la densité de la loi $\mathcal{N}(0, 1)$ et l'histogramme correspondant à des tirages indépendants de variables aléatoires distribuées comme $\sqrt{n} \frac{(\bar{X}_n - m)}{\sigma}$ où les X_i sont de loi de Poisson de paramètre 1.



Remarque 4.6. *Observons aussi la chose suivante : pour tout entier n on a*

$$\bar{X}_n = m + \frac{\sigma}{\sqrt{n}} \left(\sqrt{n} \frac{(\bar{X}_n - m)}{\sigma} \right)$$

donc, une interprétation combinée de la loi des grands nombres et du TCL est que quand « n est grand » \bar{X}_n est « distribué » comme $m + \frac{\sigma}{\sqrt{n}}U$ où U est de loi $\mathcal{N}(0, 1)$.

4.6 Combinaison de limites

On peut combiner des propriétés de convergence de variables aléatoires pour obtenir les résultats suivants

*. En quel sens ?

Théorème 4.4. Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue et $(X_n)_{n \geq 1}$ et X des variables aléatoires. On a

$$\begin{aligned} (i) \quad X_n &\xrightarrow{p.s.} X \Rightarrow g(X_n) \xrightarrow{p.s.} g(X) \\ (ii) \quad X_n &\xrightarrow{\mathbb{P}} X \Rightarrow g(X_n) \xrightarrow{\mathbb{P}} g(X) \\ (iii) \quad X_n &\xrightarrow{\mathcal{L}} X \Rightarrow g(X_n) \xrightarrow{\mathcal{L}} g(X). \end{aligned}$$

Théorème 4.5. « Lemme de Slutski » Soient $(X_n)_{n \geq 1}$, $(A_n)_{n \geq 1}$ et $(B_n)_{n \geq 1}$ trois suites de variables aléatoires pour lesquelles il existe une variable aléatoire X et deux constantes réelles a et b telles que $X_n \xrightarrow{\mathcal{L}} X$, $A_n \xrightarrow{\mathbb{P}} a$ et $B_n \xrightarrow{\mathbb{P}} b$. Alors $A_n X_n + B_n \xrightarrow{\mathcal{L}} aX + b$.

Exemple 4.5. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre $p \in]0, 1[$. Déterminer la limite en loi de la suite

$$\sqrt{n} \frac{(\bar{X}_n - p)}{\sqrt{\bar{X}_n(1 - \bar{X}_n)}}, \quad n \geq 1. \quad (4.2)$$

Exemple 4.6 (Un test statistique). Une pièce tombe 5 400 fois sur le côté pile en 10 000 lancers. Peut-on considérer la pièce comme mal équilibrée ?

Chapitre 5

Variables gaussiennes, vecteurs gaussiens

5.1 Introduction

On a vu au chapitre précédent que les variables aléatoires gaussiennes apparaissent naturellement quand on considère des suites de variables aléatoires indépendantes et de même loi de carré intégrable, ce *quelle que soit la loi commune de ces variables!* Une telle propriété est appelée une propriété d'universalité et confère à la loi normale un statut particulier.

Il est donc naturel de développer une théorie autour de cette loi, ce qui est l'objet du présent chapitre. Le résultat le plus remarquable que l'on trouve ici est celui affirmant que les composantes d'un vecteur gaussien sont indépendantes dès que leurs covariances sont nulles. Ainsi, à la propriété d'universalité évoquée plus haut vient s'ajouter une caractérisation simple de l'indépendance entre composantes d'un vecteur gaussien, qui est l'autre raison de la popularité de cet objet comme outil de modélisation.

5.2 Variables gaussiennes

On commence par rappeler la définition d'une variable aléatoire gaussienne.

Définition 5.1. *On dit qu'une variable aléatoire X suit une loi normale (ou encore loi gaussienne) de paramètres m et σ^2 (que l'on note $\mathcal{N}(m, \sigma^2)$) si et seulement si on est dans l'un des deux cas*

1. $\sigma^2 = 0$;
2. $\sigma^2 > 0$ et X admet

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

comme densité.

Remarque 5.1. *Il est important de garder en mémoire que les variables aléatoires constantes sont des variables aléatoires gaussiennes de variance nulle.*

Proposition 5.1. *On a $\mathbb{E}[X] = m, \mathbb{V}[X] = \sigma^2$, pour tout $t \in \mathbb{R}, M_X(t) = e^{mt + \frac{t^2\sigma^2}{2}}$ et $\varphi_X(t) = e^{imt - \frac{t^2\sigma^2}{2}}$.**

*. Dans les deux cas, il suffit de retenir le résultat pour $m = 0$ et $\sigma = 1$. Pourquoi ?

5.3 Vecteurs gaussiens

Définition 5.2. On dit que le vecteur $X = (X_1, \dots, X_d)$ est un vecteur gaussien si et seulement si pour tout d -uplet de nombres réels $(\lambda_1, \dots, \lambda_d)$ la variable aléatoire réelle $\sum_{i=1}^d \lambda_i X_i$ est une variable gaussienne.

Remarque 5.2. Si $X = (X_1, \dots, X_d)$ est un vecteur gaussien alors chacune de ses composantes est une variable aléatoire gaussienne. La réciproque est fautive comme l'illustre l'exercice 77 de la feuille de TD.

Remarque 5.3. Si $X = (X_1, \dots, X_d)$ est un vecteur gaussien alors tout vecteur extrait de X est lui-même un vecteur gaussien.

Proposition 5.2. Soient X_1, \dots, X_n des variables aléatoires gaussiennes mutuellement indépendantes. Alors le vecteur $X = (X_1, \dots, X_n)$ est un vecteur gaussien.

Remarque 5.4. Une conséquence immédiate de cette proposition est qu'une combinaison linéaire quelconque de variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne.

Proposition 5.3. Soient $X = (X_1, \dots, X_d)$ un vecteur gaussien, B une matrice de taille $r \times d$ et $m \in \mathbb{R}^r$. Le vecteur $Y = (Y_1, \dots, Y_r)$ défini par $Y = m + BX$ est un vecteur gaussien de moyenne $m + B\mathbb{E}[X]$ et de matrice de dispersion $BD_X B^*$.

Proposition 5.4. Soit $X = (X_1, \dots, X_d)$ un vecteur gaussien d'espérance m et de matrice de dispersion D . Pour tout $t \in \mathbb{R}^d$ on a

$$\varphi_X(t) = \exp\left(it^*m - \frac{1}{2}t^*Dt\right).$$

Ainsi la loi d'un vecteur gaussien est caractérisée par sa moyenne et sa matrice de dispersion. On note $\mathcal{N}_d(m, D)$ la loi d'un vecteur gaussien de taille d , d'espérance m et de matrice de dispersion D .

Proposition 5.5. À toute matrice D de taille $d \times d$ non-nulle symétrique et positive et à tout $m \in \mathbb{R}^d$ on peut associer un vecteur aléatoire gaussien de loi $\mathcal{N}_d(m, D)$.

On a vu au chapitre précédent que les constantes sont des cas particuliers de variables aléatoires gaussiennes et donc que certaines variables aléatoires gaussiennes n'admettent pas de densité. Il en va de même pour les vecteurs gaussiens : certains n'admettent pas de densité. En fait, pour qu'un vecteur aléatoire gaussien admette une densité, il faut et il suffit que sa matrice de dispersion soit inversible.

Proposition 5.6. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire gaussien de loi $\mathcal{N}_d(m, D)$. Si D est inversible la densité jointe de X_1, \dots, X_d est donnée par

$$f(x_1, \dots, x_d) = \frac{1}{(2\pi)^{d/2} \sqrt{|\det D|}} \exp\left(-\frac{1}{2}(x - m)^* D^{-1} (x - m)\right).$$

Exemple 5.1. Soit le vecteur gaussien (X, Y) d'espérance $(0, 1)$ et de matrice de covariance $\begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 \end{pmatrix}$. Calculer sa densité par rapport à la mesure de Lebesgue

C'est la propriété suivante qui fait la singularité des vecteurs gaussiens et un outil de modélisation très facile à manier

Proposition 5.7. Les composantes d'un vecteur gaussien sont mutuellement indépendantes si et seulement si la matrice de dispersion associée est une matrice diagonale.