

Licence MIASHS Universités de Rennes 1 et Rennes 2
Deuxième année

Introduction à la théorie des probabilités

Magalie Fromont

Table des matières

1	Espaces probabilisés - Généralités	7
1.1	Introduction - Hasard, aléatoire et probabilités	7
1.2	Univers des possibles, évènements élémentaires	7
1.3	Évènements, tribu	8
1.4	Mesure de probabilité ou probabilité	9
2	Probabilités conditionnelles et indépendance	11
2.1	Probabilités conditionnelles	11
2.1.1	Exemple et définition	11
2.1.2	Propriétés	12
2.2	Indépendance	13
3	Variables aléatoires réelles	15
3.1	Définitions - premiers exemples	15
3.2	Loi d'une variable aléatoire réelle	15
3.3	Fonction de répartition - fonction de masse - fonction de densité	16
3.3.1	La fonction de répartition d'une v.a.r.	16
3.3.2	Fonction de masse et classification des v.a.r.	17
3.4	Variables aléatoires de loi discrète	17
3.5	Variables aléatoires de loi absolument continue - densité	18
3.6	Espérance - Moments - Quantiles d'une v.a.r.	19
3.6.1	Espérance	19
3.6.2	Moments (centrés ou non)	20
3.6.3	Inégalités faisant intervenir les moments	21
3.6.4	Médiane - Quantiles d'une v.a.r.	22
3.7	Calcul de lois	22
3.8	Corrélation et indépendance	24
3.8.1	Définitions et premières propriétés	24
3.8.2	Somme de v.a.r., covariance et corrélation	24
3.9	Transformées	26
3.9.1	Fonction génératrice des v.a.r. discrètes à valeurs dans \mathbb{N}	26
3.9.2	Fonction génératrice des moments et transformée de Laplace	27
3.9.3	Fonction caractéristique	27
4	Vecteurs aléatoires	29
4.1	Définitions - premiers exemples	29
4.2	Loi conjointe, lois marginales	29
4.2.1	Fonctions de répartition - densités	30
4.2.2	Espérance mathématique - Moments d'un vecteur aléatoire	32
4.3	Variables aléatoires marginales indépendantes	33

4.4	Calcul de lois	34
5	Lois usuelles dans \mathbb{R} et dans \mathbb{R}^n	37

Références :

- ***Probabilités via l'intégrale de Riemann***, C. Suquet,
- *Calcul des probabilités*, D. Foata et A. Fuchs,
- *Probability and random processes*, G. Grimmett and D. Stirzaker,
- *Cours et exercices de probabilités appliquées*, M. Lefebvre,
- *Probabilités, analyse des données et statistique*, G. Saporta,
- *Théorie des probabilités en vue des applications statistiques*, P. Tassi et S. Legait,
- *Cours de probabilités*, A. Monfort.

Chapitre 1

Espaces probabilisés - Généralités

1.1 Introduction - Hasard, aléatoire et probabilités

La théorie des probabilités est en lien étroit avec le réel (dans lequel chacun est d'ailleurs familier avec les termes : hasard, aléatoire, évènement, probabilité), en ce sens qu'elle permet de poser un modèle mathématique sur des expériences aléatoires (c'est-à-dire où le hasard ou l'aléa intervient, dont le résultat ne peut être prévu avec une totale certitude), que l'on pourra ainsi mieux comprendre et étudier.

Un peu d'histoire...

Le mot *hasard* est un mot d'origine arabe : *az-zahr*, le dé. Il est apparu en français pour désigner tout d'abord un jeu de dés, puis plus généralement un évènement non prévisible, et par extension le mode d'apparition de ce type d'évènement.

Bien avant la définition d'un cadre très général permettant de modéliser le plus grand nombre d'expériences aléatoires possible, de nombreux principes de calcul de probabilités d'évènements ont été avancés. Les travaux de Pierre de Fermat, Blaise Pascal et Christian Huygens (17ème siècle), puis Pierre-Simon de Laplace, Abraham de Moivre, Jacques Bernoulli, et Denis Siméon Poisson (18ème) et Carl Friedrich Gauss et Henri Poincaré (19ème) sont des avancées majeures dans la conception de ces principes.

La théorie générale des probabilités n'a été introduite rigoureusement qu'en 1933 par Andreï Kolmogorov, nécessitant préalablement le développement des théories de la mesure et de l'intégration.

Les probabilités dans la filière MIAHS, pourquoi? Parce qu'elles sont à la base de la statistique inférentielle...

1.2 Univers des possibles, évènements élémentaires

Définition 1. *L'ensemble des résultats possibles d'une expérience aléatoire est représenté par un ensemble noté Ω , appelé univers (des possibles) ou ensemble fondamental. Les sous-ensembles de Ω contenant un seul élément de Ω sont appelés les évènements élémentaires.*

Remarque : les évènements élémentaires contiennent en général l'information maximale dont on dispose sur une expérience aléatoire.

Exemple : pour modéliser l'expérience aléatoire correspondant à deux lancers successifs d'un dé à six faces, on prendra $\Omega = \{\omega = (i, j), i \in \{1, \dots, 6\}, j \in \{1, \dots, 6\}\} = \{1, \dots, 6\}^2$ et non l'ensemble des sommes possibles des points apparaissant sur les dés $\Omega = \{2, \dots, 12\}$. Chaque singleton $\{\omega = (i, j)\} \subset$

Ω est un évènement élémentaire, et on a $\Omega = \{(1, 1)\} \cup \dots \cup \{(1, 6)\} \cup \dots \cup \{(6, 1)\} \cup \dots \cup \{(6, 6)\}$.

Remarque : il est fréquent de confondre les sous-ensembles $\{\omega\}$ de Ω que sont les évènements élémentaires avec les éléments ω de Ω . On dira souvent (par abus de langage ou par souci de simplification des notations) que ω est un évènement élémentaire.

1.3 Évènements, tribu

Dans la suite, on voudra définir ou calculer la probabilité de certains sous-ensembles de Ω , dont la caractérisation peut se faire simplement par des mots (en utilisant le vocabulaire logique, avec les opérations logiques : "et", "ou", "non") ou à l'aide des évènements élémentaires et du formalisme de la théorie des ensembles (avec les opérations ensemblistes : \cup , \cap , \setminus , et le passage au complémentaire $\bar{E} = \Omega \setminus E$).

En reprenant l'exemple des deux lancers de dé successifs, on peut s'intéresser à la probabilité de E : "la somme des points apparaissant sur les dés est égale à 10", que l'on peut aussi caractériser par $E = \{(4, 6), (5, 5), (6, 4)\}$ ou encore $E = \{(4, 6)\} \cup \{(5, 5)\} \cup \{(6, 4)\}$.

Lorsque Ω est fini ou dénombrable, on pourra le faire pour tous les sous-ensembles possibles de Ω .

Définition 2. L'ensemble de tous les sous-ensembles possibles de Ω est appelé l'**ensemble des parties de Ω** et est noté $\mathcal{P}(\Omega)$.

En revanche, lorsque Ω n'est ni fini, ni dénombrable, par exemple $\Omega = [0, 1]$ ou $\Omega = \mathbb{R}$, certains sous-ensembles de Ω seront trop complexes pour qu'on puisse en définir ou calculer une probabilité.

On se restreindra donc à une famille $\mathcal{F} \subset \mathcal{P}(\Omega)$ contenant les sous-ensembles dont on pourra définir ou calculer une probabilité, appelés **évènements**.

Par souci de cohérence, cette famille devra vérifier certaines contraintes de stabilité : elle devra être ce que l'on appelle une tribu.

Définition 3. Une famille \mathcal{F} de sous-ensembles de Ω , i.e. $\mathcal{F} \subset \mathcal{P}(\Omega)$, est une **tribu** sur Ω si elle :

- possède l'ensemble vide ou Ω : $\emptyset \in \mathcal{F}$ ou $\Omega \in \mathcal{F}$,
- est stable par passage au complémentaire : $\forall E \in \mathcal{F}, \bar{E} \in \mathcal{F}$,
- est stable par union dénombrable : pour toute suite $(E_i)_{i \in \mathbb{N}}$ d'éléments de \mathcal{F} , $\cup_{i \in \mathbb{N}} E_i \in \mathcal{F}$.

On peut remarquer qu'une tribu est par définition stable par unions finies, intersections finies ou dénombrables.

On peut désormais établir les correspondances suivantes entre le vocabulaire ensembliste et le vocabulaire probabiliste.

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste
\emptyset	ensemble vide	évènement impossible
Ω	ensemble plein	univers des possibles
ω	élément de Ω	confondu avec $\{\omega\}$: évènement élémentaire
$E \in \mathcal{F}$	sous-ensemble de Ω	évènement
$A \subset B$	A est inclus dans B	A implique B
$A \cup B$	union de A et B	A ou B
$A \cap B$	intersection de A et B	A et B
$A \setminus B$	A privé de B	A et non B
\bar{A}	complémentaire de A dans Ω	évènement contraire de A
$A \cap B = \emptyset$	A et B sont disjoints	A et B sont incompatibles
$(A_i)_{i \in \mathbb{N}^*},$ $A_i \cap A_j = \emptyset \forall (i, j), i \neq j$	les A_i sont deux à deux disjoints	les A_i sont deux à deux incompatibles

Quelques exemples de tribus :

- Tribu triviale : $\mathcal{F} = \{\emptyset, \Omega\}$;
- Tribu pleine : $\mathcal{F} = \mathcal{P}(\Omega)$;
- Tribu engendrée par un sous-ensemble Ω' de Ω : $\mathcal{F} = \{\emptyset, \Omega, \Omega', \overline{\Omega'}\}$.

Expérience aléatoire 1 : Pile ou Face. On lance une pièce. On considère les événements élémentaires suivants : P : "Obtenir pile", F : "Obtenir face". La tribu est composée de $\Omega = P \cup F$, son complémentaire \emptyset , F et P . On a aussi $F \cup F = F$, $\overline{F} = P$, $F \cap P = \emptyset$, $F \cup F \cup P = \Omega$ etc. (on tourne en rond). Finalement $\mathcal{F} = \{\emptyset, F, P, \Omega\}$.

Expérience aléatoire 2 : Pile, Face, Oiseau. On lance une pièce. On considère les événements élémentaires suivants : P : "Obtenir pile", F : "Obtenir face", O : "Un oiseau tape à la fenêtre, rentre, prend la pièce et s'envole avec". La tribu est composée de $\Omega = P \cup F \cup O$, son complémentaire \emptyset , F , P et O . On a aussi toutes les unions et intersections : $F \cup O$, $F \cup P$, $P \cup O$, $P \cap O$, $\overline{P} = F \cup O$, $\overline{P \cup F} = O$...
 $\mathcal{F} = \{\emptyset, F, P, O, F \cup P, F \cup O, P \cup O, \Omega\}$.

Expérience aléatoire 3 : Pile, Face, Oiseau, Chat. On lance une pièce. On considère les événements élémentaires suivants : P : "Obtenir pile", F : "Obtenir face", O : "Un oiseau tape à la fenêtre, rentre, prend la pièce et s'envole avec", C : "Un chat rentre en même temps que l'oiseau, l'attrape, le mange, et repose la pièce côté Face."

$\mathcal{F} = \{\emptyset, F, P, O, C, F \cup P, F \cup O, F \cup C, P \cup O, P \cup C, O \cup C, F \cup P \cup C, F \cup P \cup O, F \cup O \cup C, P \cup O \cup C, \Omega\}$.

Définition 4. Si \mathcal{O} est une famille de sous-ensembles de Ω , on appelle **tribu engendrée par \mathcal{O}** la plus petite tribu contenant \mathcal{O} . C'est l'intersection de toutes les tribus sur Ω contenant \mathcal{O} .

Lorsque $\Omega = \mathbb{R}$, on considère en général la tribu des boréliens ou tribu borélienne.

Définition 5. On appelle **tribu des boréliens** ou **tribu borélienne** sur \mathbb{R} la tribu notée $\mathcal{B}(\mathbb{R})$ engendrée par la famille \mathcal{O} des intervalles ouverts. Ses éléments sont appelés les **boréliens de \mathbb{R}** .

Et lorsque $\Omega = \mathbb{R}^d$, on généralise cette définition.

Définition 6. On appelle **tribu des boréliens** ou **tribu borélienne** sur \mathbb{R}^d la tribu notée $\mathcal{B}(\mathbb{R}^d)$ engendrée par la famille \mathcal{O} des pavés ouverts $\prod_{k=1}^d]a_k, b_k[$. Ses éléments sont appelés les **boréliens de \mathbb{R}^d** .

1.4 Mesure de probabilité ou probabilité

La probabilité telle que nous allons la définir est une fonction qui à un événement associe un nombre compris entre 0 et 1 censé mesurer les chances de réalisation de cet événement.

Définition 7. Soit Ω l'univers des possibles et \mathcal{F} une tribu sur Ω . On appelle **mesure de probabilité** ou **probabilité** sur (Ω, \mathcal{F}) toute application P de \mathcal{F} dans $[0, 1]$ telle que :

- $P(\Omega) = 1$;
- (σ -additivité) Pour toute suite $(A_i)_{i \in \mathbb{N}^*}$ d'éléments de \mathcal{F} deux à deux disjoints,

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

Le triplet (Ω, \mathcal{F}, P) est appelé **espace probabilisé** ou **espace de probabilité**.

À partir de là, l'expérience est complètement modélisée : on donne les évènements considérés et leurs probabilités.

Exemple des lancers de dé. On a vu que $\Omega = \{1, \dots, 6\}^2$ et puisque Ω est un ensemble fini, on considère la tribu pleine $\mathcal{F} = \mathcal{P}(\Omega)$. L'application $P : \mathcal{F} \rightarrow [0, 1]$ qui à $A \in \mathcal{F}$ associe $P(A) = \frac{\#A}{\#\Omega}$ est une probabilité sur (Ω, \mathcal{F}) .

Généralisation : équiprobabilité et probabilité uniforme sur un univers fini. Pour un univers Ω de cardinal fini, dès lors qu'on considère que tous les évènements élémentaires ont la même chance de réalisation (i.e. qu'ils sont **équiprobables**), on le munira de la tribu pleine $\mathcal{F} = \mathcal{P}(\Omega)$ et de la probabilité définie par :

$$P(A) = \frac{\#A}{\#\Omega} \quad \forall A \in \mathcal{F}.$$

Cette probabilité est appelée **probabilité uniforme** sur Ω .

Autres exemples : mesure ou loi de Dirac en un évènement élémentaire $\{\omega_0\}$ d'un univers fini ou dénombrable, lois de Mendel, loi de répartition théorique des couleurs de M&M's (15% jaune, 12% rouge, 23% orange, 23% bleu, 12% marron, 15% vert).

Proposition 1. Soit $\Omega = \{\omega_i, i \in I\}$ un univers au plus dénombrable.

Une probabilité P sur $(\Omega, \mathcal{F} = \mathcal{P}(\Omega))$ est entièrement caractérisée par la donnée de $(P(\{\omega_i\}))_{i \in I}$.

Si $(p_i)_{i \in I}$ est une famille telle que $p_i \in [0, 1] \forall i \in I$ et $\sum_{i \in I} p_i = 1$, $P : \mathcal{F} \rightarrow [0, 1]$, $\{\omega_i\} \mapsto P(\{\omega_i\}) = p_i$ définit une probabilité sur $(\Omega, \mathcal{F} = \mathcal{P}(\Omega))$.

Autre exemple avec Ω non dénombrable, par exemple $\Omega = \mathbb{R}$ muni de la tribu borélienne : la loi uniforme sur $[0, 1]$. Remarquer que pour tout $x \in [0, 1]$, $P(\{x\}) = 0$, donc que $P([0, 1]) = P(\cup_{x \in [0, 1]} \{x\}) \neq \sum_{x \in [0, 1]} P(\{x\})$.

On verra que pour caractériser une probabilité sur $(\Omega = \mathbb{R}, \mathcal{F} = \mathcal{B}(\mathbb{R}))$, il suffit de donner $F(x) = P(]-\infty, x])$ pour tout $x \in \mathbb{R}$ (fonction de répartition). Il en est de même pour Ω au plus dénombrable d'ailleurs.

Proposition 2 (Propriétés d'une probabilité). Pour tout espace de probabilité (Ω, \mathcal{F}, P) , P vérifie les propriétés suivantes.

- *Monotonie* : si $A \subset B$, $P(A) \leq P(B)$. Plus précisément : $P(B) = P(A) + P(B \setminus A)$ (donc $P(B \setminus A) = P(B) - P(A)$, $P(\bar{A}) = 1 - P(A)$ et $P(\emptyset) = 0$).
- *Additivité forte* : $P(A) + P(B) = P(A \cup B) + P(A \cap B)$ (donc $P(A \cup B) = P(A) + P(B) - P(A \cap B)$).
- *Sous σ -additivité* :

$$P\left(\cup_{i=1}^{+\infty} A_i\right) \leq \sum_{i=1}^{+\infty} P(A_i).$$

- *Formule de Poincaré* : si $n \geq 2$,

$$P\left(\cup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) + \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}).$$

- *Continuité monotone croissante* : si $(A_i)_{i \in \mathbb{N}}$ est une suite d'évènements croissante pour l'inclusion, alors

$$P\left(\cup_{i=1}^{+\infty} A_i\right) = \lim_{i \rightarrow +\infty} P(A_i).$$

- *Continuité monotone décroissante* : si $(A_i)_{i \in \mathbb{N}}$ est une suite d'évènements décroissante pour l'inclusion, alors

$$P\left(\cap_{i=1}^{+\infty} A_i\right) = \lim_{i \rightarrow +\infty} P(A_i).$$

Chapitre 2

Probabilités conditionnelles et indépendance

On considère dans toute la suite une expérience aléatoire modélisée par un espace probabilisé (Ω, \mathcal{F}, P) .

2.1 Probabilités conditionnelles

2.1.1 Exemple et définition

Exemple : on prend un étudiant au hasard en Licence 2 MIASHS et on s'intéresse d'abord à la probabilité que cet étudiant soit dans le parcours Économie.

On considère l'espace probabilisé (Ω, \mathcal{F}, P) , où Ω est l'ensemble des étudiants de L2 MIASHS, $\mathcal{F} = \mathcal{P}(\Omega)$ et $P(\{\omega\}) = \frac{1}{\#\Omega}$ (équiprobabilité). Soit E l'ensemble des étudiants du parcours Économie, et S l'ensemble des étudiants du parcours SHS. Alors la probabilité cherchée est

$$P(E) = \frac{\#E}{\#\Omega}.$$

Si on dispose de l'information supplémentaire que l'étudiant est de sexe masculin, la probabilité qu'il soit dans le parcours Économie se trouve-t-elle modifiée ?

Si M désigne l'ensemble des étudiants de sexe masculin, et F l'ensemble des étudiants de sexe féminin, cela revient à ne considérer comme univers des possibles que M .

La probabilité cherchée est la probabilité que E se réalise sachant que M s'est réalisé, appelée probabilité (conditionnelle) de E sachant M et se note $P(E|M)$. Elle est égale à :

$$P(E|M) = \frac{\#E \cap M}{\#M} = \frac{\frac{\#E \cap M}{\#\Omega}}{\frac{\#M}{\#\Omega}} = \frac{P(E \cap M)}{P(M)}.$$

↔ On introduit pour répondre à cette question le concept général (défini même lorsqu'il n'y a pas équiprobabilité) de probabilité conditionnelle.

Définition 8. Soit C un évènement tel que $P(C) > 0$. Pour tout évènement A , la **probabilité (conditionnelle) de A sachant C** est définie par :

$$P(A|C) = \frac{P(A \cap C)}{P(C)}.$$

Par convention, on peut aussi définir $P(A|C)$ lorsque $P(C) = 0$ en posant $P(A|C) = 0$ dans ce cas.

Remarque : Il est important de voir que ce qui a été modifié ici n'est pas l'évènement A mais la probabilité de cet évènement. La probabilité $P(\cdot)$ est devenue $P(\cdot|C)$.

2.1.2 Propriétés

Proposition 3. Soit C un évènement tel que $P(C) > 0$.

$P(\cdot|C) : \mathcal{F} \rightarrow [0, 1]$, $A \mapsto P(A|C)$ est une nouvelle probabilité sur (Ω, \mathcal{F}) . Elle vérifie donc toutes les propriétés d'une probabilité.

- $P(\emptyset|C) = 0$;
- $P(\Omega|C) = 1$;
- Si $C \subset A$, $P(A|C) = 1$;
- Si $A \in \mathcal{F}$, $P(\bar{A}|C) = 1 - P(A|C)$;
- Monotonie : Si $A, B \in \mathcal{F}$, $A \subset B \Rightarrow P(A|C) \leq P(B|C)$;
- Additivité forte : Si $A, B \in \mathcal{F}$, $P(A|C) + P(B|C) = P(A \cup B|C) + P(A \cap B|C)$ (donc $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$).
- Sous σ -additivité :

$$P\left(\bigcup_{i=1}^{+\infty} A_i \middle| C\right) \leq \sum_{i=1}^{+\infty} P(A_i|C).$$

- Continuité monotone croissante : si $(A_i)_{i \in \mathbb{N}}$ est une suite d'évènements croissante pour l'inclusion, alors

$$P\left(\bigcup_{i=1}^{+\infty} A_i \middle| C\right) = \lim_{i \rightarrow +\infty} P(A_i|C).$$

- Continuité monotone décroissante : si $(A_i)_{i \in \mathbb{N}}$ est une suite d'évènements décroissante pour l'inclusion, alors

$$P\left(\bigcap_{i=1}^{+\infty} A_i \middle| C\right) = \lim_{i \rightarrow +\infty} P(A_i|C).$$

Proposition 4 (Formule des probabilités composées). Soit A_1 et A_2 deux évènements de probabilité non nulle. Alors

$$P(A_1 \cap A_2) = P(A_1|A_2)P(A_2) = P(A_2|A_1)P(A_1),$$

donc

$$P(A_1|A_2) = \frac{P(A_2|A_1)P(A_1)}{P(A_2)}.$$

Soit A_1, \dots, A_n n évènements tels que $P(A_1 \cap \dots \cap A_n) \neq 0$. Alors

$$P(A_1 \cap \dots \cap A_n) = P(A_n|A_1 \cap \dots \cap A_{n-1})P(A_{n-1}|A_1 \cap \dots \cap A_{n-2})P(A_{n-2}|A_1 \cap \dots \cap A_{n-3}) \dots P(A_3|A_2 \cap A_1)P(A_2|A_1)P(A_1).$$

Preuve. Pour tout $j \in \{1, \dots, n\}$, $A_1 \cap \dots \cap A_n \subset A_1 \cap \dots \cap A_j$, donc $0 < P(A_1 \cap \dots \cap A_n) \leq P(A_1 \cap \dots \cap A_j)$. Aucun évènement $A_1 \cap \dots \cap A_j$ n'est de probabilité nulle. On a alors

$$\begin{aligned} & P(A_n|A_1 \cap \dots \cap A_{n-1})P(A_{n-1}|A_1 \cap \dots \cap A_{n-2})P(A_{n-2}|A_1 \cap \dots \cap A_{n-3}) \dots P(A_2|A_1)P(A_1) \\ &= \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})} \times \frac{P(A_1 \cap \dots \cap A_{n-1})}{P(A_1 \cap \dots \cap A_{n-2})} \times \frac{P(A_1 \cap \dots \cap A_{n-2})}{P(A_1 \cap \dots \cap A_{n-3})} \times \dots \times \frac{P(A_1 \cap A_2)}{P(A_1)} \times P(A_1) \\ &= P(A_1 \cap \dots \cap A_n). \end{aligned}$$

Dans une expérience aléatoire réelle, il est souvent plus simple de calculer des probabilités conditionnelles que des probabilités.

On a déjà vu que $P(A_1|A_2)$ et $P(A_2)$ permettent de calculer $P(A_1 \cap A_2)$.

En fait, en considérant une partition au plus dénombrable d'évènements de Ω c'est-à-dire une famille au plus dénombrable $(C_i)_{i \in I}$ d'évènements de Ω tels que :

- $\forall i \in I$, $C_i \neq \emptyset$,
- $\forall i \neq j \in I$, $C_i \cap C_j, C_i \cap C_j = \emptyset$ (évènements deux à deux incompatibles),

— $\cup_{i \in I} C_i = \Omega$,

on peut calculer des probabilités d'évènements par décomposition.

Proposition 5 (Formule des probabilités totales). *Soit $(C_i)_{i \in I}$ une partition au plus dénombrable d'évènements de Ω de probabilité non nulle. Alors pour tout évènement $A \in \mathcal{F}$,*

$$P(A) = \sum_{i \in I} P(A \cap C_i) = \sum_{i \in I} P(A|C_i)P(C_i).$$

Cas particulier : si C est un évènement tel que $P(C) \neq 0$ et $P(C) \neq 1$, alors

$$P(A) = P(A|C)P(C) + P(A|\bar{C})P(\bar{C}).$$

↔ Exemple des étudiants de L2 MIASHS.

Et le sens inverse ?

Proposition 6 (Formule de Bayes). *Soit A un évènement de probabilité non nulle et si $(C_i)_{i \in I}$ est une partition au plus dénombrable d'évènements de Ω de probabilité non nulle, alors*

$$P(C_i|A) = \frac{P(A|C_i)P(C_i)}{P(A)} = \frac{P(A|C_i)P(C_i)}{\sum_{i \in I} P(A|C_i)P(C_i)}.$$

Cas particulier : si C est un évènement tel que $P(C) \neq 0$ et $P(C) \neq 1$, alors

$$P(C|A) = \frac{P(A|C)P(C)}{P(A)} = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|\bar{C})P(\bar{C})}.$$

Exemples. Exercices de TD, cas du test de dépistage (classique).

2.2 Indépendance

La notion d'indépendance est fondamentale en probabilités (et en statistique plus tard). Intuitivement, deux évènements sont indépendants si la réalisation de l'un n'a aucune influence sur la réalisation ou non de l'autre.

Définition 9. *Deux évènements A et B sont indépendants si $P(A \cap B) = P(A)P(B)$.*

Remarques :

- Si $P(B) > 0$, l'indépendance de A et B s'écrit encore $P(A|B) = P(A)$, et on retrouve la notion intuitive d'indépendance.
- Si $P(A) = 0$ ou 1 , A est indépendant de tout évènement, y compris de lui-même.
- Deux évènements A et B incompatibles ne sont indépendants que si l'un des deux est de probabilité nulle. Attention à ne pas confondre l'indépendance et l'incompatibilité!

↔ On reprend l'exemple des étudiants de L2 MIASHS. Les évènements E et M sont-ils indépendants ?

Exemple du tirage d'un nombre au hasard dans $\{1, \dots, 12\}$. On prend les évènements A : "le nombre tiré est pair", et B : "le nombre tiré est un multiple de 3". Les deux évènements sont indépendants. Si on reprend la même expérience mais en remplaçant $\{1, \dots, 12\}$ par $\{1, \dots, 13\}$, les évènements A : "le nombre tiré est pair", et B : "le nombre tiré est un multiple de 3" ne sont pas indépendants.

On le verra en calculant $P(A \cap B)$ et $P(A)P(B)$. Mais les probabilités conditionnelles aident ici clairement à mieux comprendre pourquoi...

Proposition 7. Soit A et B deux évènements indépendants. Alors

- A et \bar{B} sont indépendants,
- \bar{A} et B sont indépendants,
- \bar{A} et \bar{B} sont indépendants.

Preuve. $P(A \cap \bar{B}) = P(A \setminus A \cap B) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\bar{B})$. Les autres points se démontrent avec le même type d'arguments.

Définition 10. Soit $(A_i)_{i \in \mathbb{N}}$ une suite d'évènements. Alors on dit qu'ils sont :

- **deux à deux indépendants** si pour tout $(i, j) \in \mathbb{N}^2, i \neq j, A_i$ et A_j sont indépendants,
- **mutuellement indépendants** si pour tout sous-ensemble fini $\{i_1, \dots, i_k\}$ de \mathbb{N} , $P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \times \dots \times P(A_{i_k})$.

Cas particulier. A, B, C sont deux à deux indépendants si $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$ et $P(B \cap C) = P(B)P(C)$. Ils sont mutuellement indépendants si $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$ et aussi $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Exercice : simplification de la formule de Poincaré en cas d'indépendance mutuelle.

Exercice : si A est indépendant de B et C , A est indépendant de tout évènement construit sur B et C ($B \cup C, B \cap C, B \cup \bar{C}, B \cap \bar{C} \dots$).

Chapitre 3

Variables aléatoires réelles

On considère une expérience aléatoire modélisée par un espace probabilisé (Ω, \mathcal{F}, P) .

3.1 Définitions - premiers exemples

Définition 11. On appelle **variable aléatoire réelle (v.a.r.)** sur (Ω, \mathcal{F}) toute application $X : \Omega \rightarrow \mathbb{R}$ telle que : pour tout $x \in \mathbb{R}$, l'ensemble $X^{-1}(]-\infty, x]) = \{\omega \in \Omega, X(\omega) \leq x\}$ appartient à \mathcal{F} , ou de façon équivalente : pour tout intervalle I de \mathbb{R} , $X^{-1}(I) = \{\omega \in \Omega, X(\omega) \in I\} \in \mathcal{F}$, ou encore : pour tout borélien B de \mathbb{R} ($B \in \mathcal{B}(\mathbb{R})$), $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\} \in \mathcal{F}$.

Propriétés immédiates.

Les propriétés algébriques usuelles ainsi que la composition par une fonction réelle **mesurable** φ (i.e. telle que pour tout borélien B de \mathbb{R} , $\varphi^{-1}(B) \in \mathcal{B}(\mathbb{R})$) conservent la notion de variable aléatoire (ainsi, si X et Y sont deux variables aléatoires réelles, λ un nombre réel, λX , $X + Y$, XY , $|X|$, \sqrt{X} , $1/X$, X^n , e^X , $\ln X$ sont des variables aléatoires réelles).

Notation. Si $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire réelle sur (Ω, \mathcal{F}) , l'ensemble $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ est un évènement généralement noté $\{X \in B\}$, dont la probabilité est notée $P(X \in B)$ et se lit "probabilité que X soit dans B ". L'ensemble B peut prendre diverses formes et on écrit par exemple, $P(X = b)$ au lieu de $P(X \in \{b\})$, $P(X \leq b)$ au lieu de $P(X \in]-\infty, b])$...

Ces probabilités sont lues "probabilité que X soit égal à b ", "probabilité que X soit inférieur ou égal à b "...

Si $X_1, X_2 : \Omega \rightarrow \mathbb{R}$ sont deux variables aléatoires réelles sur (Ω, \mathcal{F}) , la probabilité $P(X_1^{-1}(]-\infty, b_1]) \cap X_2^{-1}(]-\infty, b_2])$) est notée $P(X_1 \leq b_1, X_2 \leq b_2)$ et lue "probabilité que X_1 soit inférieur ou égal à b_1 et que X_2 soit inférieur ou égal à b_2 ".

Exemples liés aux expériences aléatoires du chapitre précédent. Exemple de l'indicatrice d'un évènement.

3.2 Loi d'une variable aléatoire réelle

La notion de variable aléatoire réelle permet de probabiliser l'espace mesurable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On a en effet la proposition suivante.

Proposition 8. Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle sur (Ω, \mathcal{F}) . Alors l'application $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ définie pour tout borélien B de \mathbb{R} par

$$P_X(B) = P(X^{-1}(B)) = P(X \in B),$$

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

Définition 12 (Loi d'une v.a.r.). *La probabilité P_X ainsi définie est appelée loi de probabilité de X , ou simplement loi de X . On dit encore que X suit la loi de probabilité P_X , et on peut noter $X \sim P_X$.*

Remarque importante. L'espace $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ est un nouvel espace de probabilité. De plus, pour toute mesure de probabilité Q sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, il existe toujours un espace de probabilité (Ω, \mathcal{F}, P) et une variable aléatoire réelle $X : \Omega \rightarrow \mathbb{R}$ sur (Ω, \mathcal{F}) telle que $P_X = Q$ (il suffit de prendre $(\Omega, \mathcal{F}, P) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), Q)$ et pour X l'application identique de Ω).

On peut ainsi parler de variable aléatoire X ayant une loi de probabilité P_X sans spécifier l'espace de probabilité (Ω, \mathcal{F}, P) défini à la base. En pratique, en général, on ne se préoccupera plus de cet espace de probabilité (Ω, \mathcal{F}, P) .

3.3 Fonction de répartition - fonction de masse - fonction de densité

3.3.1 La fonction de répartition d'une v.a.r.

Définition 13 (Fonction de répartition d'une v.a.r.). *Si $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire réelle sur (Ω, \mathcal{F}) , on appelle **fonction de répartition de X** la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie pour tout $x \in \mathbb{R}$ par $F_X(x) = P_X(]-\infty, x]) = P(X \leq x)$.*

Proposition 9. *La fonction de répartition d'une v.a.r. X satisfait les propriétés suivantes :*

1. Pour tout $x \in \mathbb{R}$, $0 \leq F_X(x) \leq 1$.
2. $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow +\infty} F_X(x) = 1$.
3. F_X est une fonction croissante, continue à droite en tout point x de \mathbb{R} .
4. Pour tout $x \in \mathbb{R}$, $P(X = x) = F_X(x) - F_X(x^-)$ et l'ensemble des points de discontinuité de F_X , égal à $D_X = \{x, P(X = x) > 0\}$, est au plus dénombrable.

Proposition 10. *Pour toute fonction réelle F d'une variable réelle vérifiant les propriétés 1., 2., 3. de la Proposition 9, il existe une unique loi de probabilité Q sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $F(x) = Q(]-\infty, x])$ pour tout réel x .*

Conséquence : La loi d'une v.a.r. est entièrement caractérisée par sa fonction de répartition.

Une preuve de l'existence (Importante pour la simulation de certaines v.a.r.). On va montrer que si F est une fonction réelle d'une variable réelle vérifiant les propriétés 1., 2., 3. de la Proposition 9, il existe une variable aléatoire X telle que $F = F_X$.

On considère U une variable aléatoire de loi uniforme sur $[0, 1]$, c'est-à-dire telle que pour tout $[a, b] \subset [0, 1]$, $P(U \in [a, b]) = b - a$.

Cas particulier où F est continue et strictement croissante. Dans ce cas, la fonction F est une bijection de \mathbb{R} dans $[0, 1]$, et on peut considérer sa réciproque F^{-1} . Si l'on pose $X = F^{-1}(U)$, alors pour tout réel x ,

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(F^{-1}(U) \leq x) \\ &= P(U \leq F(x)) \\ &= F(x). \end{aligned}$$

Cas général. On définit l'inverse généralisée de F . On pose ainsi pour tout $u \in [0, 1]$, $F^{-1}(u) = \text{Inf}\{x \in \mathbb{R}, F(x) \geq u\}$, et on vérifie que pour $u \in [0, 1]$,

$$F^{-1}(u) \leq x \Leftrightarrow F(x) \geq u.$$

Soit $A = \{x \in \mathbb{R}, F(x) \geq u\}$. Si $x_0 \in A$, et $x_1 > x_0$, alors $x_1 \in A$. Par conséquent, A est de la forme $]a, +\infty[$ ou $[a, +\infty[$. Mais si on prend une suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de A qui tend en décroissant vers a , comme $F(x_n) \geq u$ et F est continue à droite, on a $F(a) \geq u$. Donc $a \in A$, et A est de la forme $[a, +\infty[$. $F^{-1}(u)$ est un minimum et $A = [F^{-1}(u), +\infty[$. On reprend la fin de la preuve précédente pour conclure.

3.3.2 Fonction de masse et classification des v.a.r.

Définition 14 (Fonction de masse d'une v.a.r.). Si $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire réelle sur (Ω, \mathcal{F}) de fonction de répartition F_X , la fonction $x \in \mathbb{R} \mapsto P(X = x) = F_X(x) - F_X(x^-)$ est appelée la **fonction de masse** de X .

On rappelle que l'ensemble des points de discontinuité de F_X , égal à $D_X = \{x, P(X = x) > 0\}$, est au plus dénombrable.

Définition 15 (loi continue, loi discrète). Si $D_X = \emptyset$, autrement dit si F_X est continue, on dit que la loi de X est **continue** ou **diffuse**.

Si $P(X \in D_X) = \sum_{x \in D_X} P(X = x) = 1$, on dit que la loi de X est **discrète** de support D_X , ou que X est une v.a.r. discrète à valeurs dans D_X .

Remarques :

- La loi de probabilité d'une v.a.r. discrète peut être vue comme une probabilité sur $(D_X, \mathcal{P}(D_X))$ ou sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- La loi de X est continue si et seulement si pour tout $x \in \mathbb{R}$, $P(X = x) = 0$.
- Il est facile de montrer que si $a, b \in \mathbb{R}$,
 - $P(X \in]a, b]) = F_X(b) - F_X(a)$,
 - $P(X \in [a, b]) = F_X(b) - F_X(a^-) = F_X(b) - F_X(a) + P(X = a)$,
 - $P(X \in]a, b[) = F_X(b^-) - F_X(a) = F_X(b) - P(X = b) - F_X(a)$,
 - $P(X \in [a, b[) = F_X(b^-) - F_X(a^-) = F_X(b) - P(X = b) - F_X(a) + P(X = a)$.
 Donc si la loi de X est continue, $P(X \in]a, b]) = P(X \in [a, b]) = P(X \in [a, b]) = P(X \in [a, b]) = F_X(b) - F_X(a)$.

Proposition 11. Il existe des v.a.r. dont la loi de probabilité Q n'est ni continue ni discrète. On peut alors montrer qu'il existe $\alpha \in]0, 1[$, Q_1 une loi discrète, Q_2 une loi continue tels que $Q = \alpha Q_1 + (1 - \alpha)Q_2$. La loi Q est dite **mixte**.

3.4 Variables aléatoires de loi discrète

Proposition 12 (Définition alternative). Toute application $X : \Omega \rightarrow \mathbb{R}$ telle que

- $X(\Omega)$ est un ensemble au plus dénombrable d'éléments notés $\{x_k, k \in \mathcal{K}\}$ ($\mathcal{K} \subset \mathbb{N}$);
- pour tout $k \in \mathcal{K}$, $X^{-1}(\{x_k\}) \in \mathcal{F}$;

est une v.a.r. discrète sur (Ω, \mathcal{F}) de support $D_X \subset X(\Omega)$.

Proposition 13 (Caractérisation). La loi d'une v.a.r. discrète à valeurs dans $D_X = \{x_k, k \in \mathcal{K}\}$ est entièrement caractérisée par sa fonction de masse calculée sur cet ensemble c'est-à-dire par $P(X = x_k)$ pour tout $k \in \mathcal{K}$ (avec $\sum_{k \in \mathcal{K}} P(X = x_k) = 1$). En effet, pour tout borélien (par exemple un intervalle) B de \mathbb{R} ,

$$P(X \in B) = \sum_{k \in \mathcal{K}, x_k \in B} P(X = x_k).$$

Exemples : $B =]a, b]$, $B = [a, b]$, $B = [a, b[$.

Proposition 14 (Fonction de répartition). *Si X est une v.a.r. discrète à valeurs dans un ensemble $\{x_k, k \in \mathcal{K}\}$,*

$$F_X(x) = \sum_{k \in \mathcal{K}, x_k \leq x} P(X = x_k).$$

F_X est une fonction constante par morceaux ou en escalier.

↔ Représentation graphique.

Les v.a.r. discrètes ont déjà été traitées en détail dans le cours de Probabilités du premier semestre.

On s'intéresse ici un peu plus particulièrement à des v.a.r. de lois continues particulières : les v.a.r. de lois absolument continues (les autres ont des lois dites **singulières**.)

3.5 Variables aléatoires de loi absolument continue - densité

Définition 16 (v.a.r. de loi absolument continue). *La loi de probabilité P_X d'une v.a.r. est dite **absolument continue** s'il existe une fonction f_X positive et intégrable telle que*

$$\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x f_X(t) dt. \quad (3.1)$$

Définition 17. *On appelle **densité** de la loi P_X (ou par abus densité de X) toute fonction f_X qui vérifie (3.1).*

Remarque importante : la fonction f_X n'est unique : on peut avoir deux fonctions f_X^1 et f_X^2 vérifiant (3.1) avec $f_X^1 \neq f_X^2$ sur un nombre fini de points de \mathbb{R} .

Proposition 15. *On a les propriétés suivantes.*

1. *Si X est une v.a.r. de loi absolument continue de densité f_X , il est évident que $\int_{-\infty}^{+\infty} f_X(t) dt = 1$.*
2. *Si X est une v.a.r. de loi absolument continue de densité f_X , pour tous $a, b \in \mathbb{R}$,*

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = \int_a^b f_X(t) dt,$$

$$P(X \leq b) = P(X < b) = \int_{-\infty}^b f_X(t) dt \text{ et } P(X > a) = P(X \geq a) = \int_a^{+\infty} f_X(t) dt.$$

3. *Si X est une v.a.r. de loi absolument continue de densité f_X , X est de loi continue i.e. F_X est continue.*
4. *Si X est une v.a.r. de loi absolument continue de densité f_X continue en x_0 , alors F_X est dérivable en x_0 et $F'_X(x_0) = f_X(x_0)$.*
5. *(Important en pratique) Si X est une v.a.r. dont la fonction de répartition F_X est continue, dérivable sauf éventuellement en un nombre fini de points, alors X est de loi absolument continue de densité F'_X .*

Proposition 16. *Toute fonction f réelle positive, d'une variable réelle, intégrable et telle que $\int_{-\infty}^{+\infty} f(t) dt = 1$, est la densité d'une loi de probabilité.*

↔ Idée intuitive de ce qu'est une densité en regardant des intervalles $[x - \delta, x + \delta]$ (δ petit) par exemple.

↔ Graphes. Comparaison avec les v.a.r. de loi discrète.

3.6 Espérance - Moments - Quantiles d'une v.a.r.

On considère ici un espace probabilisé (Ω, \mathcal{F}, P) , et une v.a.r. X sur (Ω, \mathcal{F}) , de loi de probabilité P_X .

3.6.1 Espérance

L'espérance d'une v.a.r. est lorsqu'elle existe, la moyenne des valeurs de cette variable, pondérées par leurs probabilités de réalisation.

↔ Historiquement, gain espéré dans un jeu d'argent.

Comment traduire cette définition pour les v.a.r. de lois discrètes, puis absolument continues ?

Définition 18 (Espérance).

1. Si X est une v.a.r. discrète à valeurs dans $X(\Omega) = \{x_k, k \in \mathcal{K}\}$, positive ou telle que $\sum_{k \in \mathcal{K}} |x_k| P(X = x_k) < +\infty$, alors l'**espérance** de X est définie par :

$$E[X] = \sum_{k \in \mathcal{K}} x_k P(X = x_k).$$

2. Si X est une v.a.r. de loi absolument continue (de densité f_X), positive ou telle que $\int_{-\infty}^{+\infty} |x| f_X(x) dx < +\infty$, alors l'**espérance** de X est définie par :

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx.$$

En fait, on a une écriture globale pour l'espérance de X : $E[X] = \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbb{R}} x dP_X(x)$ (théorie générale de l'intégration par rapport à une mesure de probabilité), qui est valable de façon plus générale, lorsque X n'est pas forcément de loi discrète ou absolument continue, dès que $\int_{\Omega} |X(\omega)| dP(\omega) = \int_{\mathbb{R}} |x| dP_X(x) < +\infty$.

Remarques.

$E[|X|]$ est toujours définie et si $E[|X|] < \infty$, $E[X]$ est bien définie avec $|E[X]| \leq E[|X|]$. On a en réalité une inégalité plus générale (inégalité de Jensen).

Le calcul de $E[X]$ lorsqu'elle existe se fait sans avoir besoin de connaître (Ω, \mathcal{F}, P) .

Interprétation de $E[X]$ comme indicateur de centralité. On parle souvent de valeur "moyenne" de X .

Définition 19. Si $E[X]$ existe et $|E[X]| < +\infty$, on dit que X a une **espérance finie**.

Si $E[X]$ existe et $E[X] = 0$, on dit que X est **centrée**.

Si $E[X]$ existe, on dit qu'on **centre** ou **recentre** X lorsqu'on lui retranche $E[X]$.

Exemples :

1. Somme des points après deux lancers de dés.
2. Considérons X la v.a.r. discrète pouvant prendre les valeurs $x_k = k$ pour tout $k \in \mathbb{N}^*$, dont la loi est définie par $P(X = k) = 1/(k(k+1))$. Commençons par remarquer que $\sum_{k=1}^{+\infty} \frac{1}{k(k+1)} = \sum_{k \in \mathbb{N}^*} \left(\frac{1}{k} - \frac{1}{k+1} \right) = 1$ (dominos ou somme télescopique). $E[X]$ est définie car X est positive mais elle est infinie : $\sum_{k=1}^{+\infty} \frac{k}{k(k+1)} = \sum_{k=1}^{+\infty} \frac{1}{k+1} = +\infty$.
3. Loi uniforme sur un intervalle.

Contre-exemples :

1. Considérons X la v.a.r. discrète pouvant prendre les valeurs $x_k = (-1)^{k+1}(k+1)$ pour tout $k \in \mathbb{N}^*$, dont la loi est définie par $P(X = x_k) = 1/(k(k+1))$. L'espérance de X n'est pas définie puisque la série $\sum_{k=1}^{+\infty} \frac{|x_k|}{k(k+1)} = \sum_{k \in \mathbb{N}^*} \frac{1}{k}$ est divergente.

2. Loi de Cauchy.

Théorème 1 (Théorème de transfert). *Soit φ une fonction d'une variable réelle, à valeurs dans \mathbb{R} .*

1. *Si X est une v.a.r. discrète à valeurs dans $X(\Omega) = \{x_k, k \in \mathcal{K}\}$, positive ou telle que $\sum_{k \in \mathcal{K}} |\varphi(x_k)| P(X = x_k) < +\infty$, alors $\varphi(X)$ est une v.a.r. discrète, dont l'espérance est donnée par :*

$$E[\varphi(X)] = \sum_{k \in \mathcal{K}} \varphi(x_k) P(X = x_k).$$

2. *Si φ est une fonction mesurable i.e. telle que pour tout borélien B de \mathbb{R} , $\varphi^{-1}(B)$ est un borélien de \mathbb{R} , et si X est une v.a.r. de loi absolument continue (de densité f_X), positive ou telle que $\int_{-\infty}^{+\infty} |\varphi(x)| f_X(x) dx < +\infty$, alors $\varphi(X)$ est une v.a.r. dont l'espérance est donnée par :*

$$E[\varphi(X)] = \int_{-\infty}^{+\infty} \varphi(x) f_X(x) dx.$$

Le théorème de transfert a une très grande utilité en pratique : grâce à lui, pour calculer l'espérance de $Y = \varphi(X)$, on a seulement besoin de la loi de X , et non de la loi de Y .

Exemples.

Proposition 17 (Propriétés de l'espérance). *On suppose que X et Y sont deux v.a.r. dont l'espérance est bien définie. Alors, on a les propriétés suivantes.*

- *Linéarité : si $a, b \in \mathbb{R}$, l'espérance de $aX + bY$ existe et $E[aX + bY] = aE[X] + bE[Y]$. En particulier, $E[aX + b] = aE[X] + b$.*
- *Monotonie : si $X \leq Y$, alors $E[X] \leq E[Y]$. En particulier, si $X \geq 0$, $E[X] \geq 0$.*

Proposition 18 (Inégalité de Jensen). *Si X est une v.a.r. à valeurs dans un intervalle $]a, b[$, d'espérance finie, si φ est une fonction réelle convexe sur $]a, b[$, alors*

$$\varphi(E[X]) \leq E[\varphi(X)].$$

3.6.2 Moments (centrés ou non)

On introduit ici d'autres caractéristiques de la loi de X , qui rendent compte de sa dispersion, par exemple les moments, ou les moments centrés.

Proposition 19. *Soit $m, m' \in \mathbb{N}^*$, $m < m'$. Si $E[|X|^{m'}] < +\infty$, alors $E[|X|^m] < +\infty$.*

Preuve. Pour $x \geq 1$, on a $x^m = x^{m'} x^{m-m'} \leq x^{m'}$, et pour $0 < x < 1$, $x^m < 1$. Par conséquent, pour tout $x > 0$, on a $x^m \leq x^{m'} + 1$. On a donc en particulier $|X|^m \leq 1 + |X|^{m'}$, et comme $E[1 + |X|^{m'}] = 1 + E[|X|^{m'}]$, le résultat suit par monotonie de l'espérance.

Définition 20 (Moments). *Soit $a \in \mathbb{R}$, $m \in \mathbb{N}^*$. Si X est positive ou si $E[|X|^m] < +\infty$, on appelle **moment d'ordre m de X** le nombre $E[X^m]$, et **moment absolu d'ordre m de X** le nombre $E[|X|^m]$. Si X est positive ou $E[|X|] < +\infty$, et si $(X - E[X])^m$ est positive ou $E[|X - E[X]|^m] < +\infty$, on appelle **moment centré d'ordre m de X** le nombre $E[(X - E[X])^m]$.*

Remarque : si X admet un moment d'ordre r' fini, alors elle admet un moment d'ordre r fini pour tout r tel que $0 < r < r'$.

On s'intéresse en particulier aux moments d'ordre 2.

Définition 21 (Variance, écart-type). *Le moment centré d'ordre 2 de X est appelé la variance de X , et noté $Var(X)$:*

$$Var(X) = E[(X - E[X])^2].$$

Sa racine carrée est appelée l'écart-type de X , noté $\sigma(X) = \sqrt{Var(X)} = \sqrt{E[(X - E[X])^2]}$.

Propriétés immédiates.

1. Si $a \in \mathbb{R}$, $Var(aX) = a^2 Var(X)$. En particulier, $Var(-X) = Var(X)$.
2. Si $a \in \mathbb{R}$, $Var(a + X) = Var(X)$.

Définition 22. *Si $E[X] = 0$ et $Var(X) = 1$, on dit que X est **centrée réduite** ou **standard**. Si $E[|X|] < +\infty$ et $Var(X) < +\infty$, on dit qu'on **centre et on réduit** ou qu'on **standardise** X si l'on calcule $\frac{X - E[X]}{\sigma(X)}$.*

Il résulte de la Proposition 19 que si X admet un moment d'ordre 2 fini, alors elle a une espérance finie.

Proposition 20 (Formule de Koenig). *X a un moment d'ordre 2 fini si et seulement si son espérance $E[X]$ et sa variance $Var(X)$ existent et sont finies, et on a :*

$$Var(X) = E[X^2] - (E[X])^2.$$

Preuve. Si X a un moment d'ordre 2 fini, $E[X]$ est finie et on a $(X - E[X])^2 = X^2 - 2XE[X] + (E[X])^2$, d'où $Var(X) = E[X^2] - (E[X])^2$. Pour la réciproque, on remarque que si $E[|X|] < +\infty$, $Var(X) < +\infty$, comme $E[X^2] = E[(X - E[X] + E[X])^2] = E[(X - E[X])^2] + (E[X])^2 + E[(X - E[X])E[X]] = E[(X - E[X])^2] + (E[X])^2 < +\infty$.

Remarque : cette proposition nous donne une nouvelle méthode de calcul de la variance, souvent plus aisée à utiliser.

Proposition 21. *Si $E[X^2] < +\infty$, pour tout réel a , $E[(X - a)^2] \geq Var(X)$.*

Preuve. Il suffit d'écrire $E[(X - a)^2] = E[((X - E[X]) + (E[X] - a))^2]$ et de développer.

Interprétation en termes de dispersion.

Définition 23 (Skewness, Kurtosis). *Le **coefficient d'asymétrie** (ou **skewness**) d'une v.a.r. X est le moment d'ordre 3 de X que l'on a standardisée (centrée réduite) : $\mathbb{E} \left[\left(\frac{X - E[X]}{\sqrt{Var(X)}} \right)^3 \right]$.*

*Le **coefficient d'aplatissement** (ou **kurtosis**) de X est le moment d'ordre 4 de X que l'on a standardisée : $\mathbb{E} \left[\left(\frac{X - E[X]}{\sqrt{Var(X)}} \right)^4 \right]$.*

3.6.3 Inégalités faisant intervenir les moments

Les moments permettent, comme on l'a dit ci-dessus, de donner une indication sur la dispersion d'une variable. Cette dispersion peut être précisée par exemple à l'aide des inégalités suivantes.

Théorème 2 (Inégalité de Markov). *Pour tout réel $a > 0$,*

$$P(|X| \geq a) \leq \frac{E[|X|]}{a}.$$

Conséquence : pour tout réel $a > 0$, et tout $m \in \mathbb{N}^*$,

$$P(|X| \geq a) \leq \frac{E[|X|^m]}{a^m}.$$

\Leftrightarrow Si X est positive, l'existence d'un moment d'ordre élevé assure une décroissance d'autant plus rapide de la queue de distribution de X .

Théorème 3 (Inégalité de Bienaymé-Chebychev). *Si $E[X^2] < +\infty$, alors pour tout réel $a > 0$,*

$$P(|X - E[X]| \geq a) \leq \frac{Var(X)}{a^2}.$$

Preuve. Il suffit d'appliquer l'inégalité de Markov à la variable aléatoire $(X - E[X])^2$.

Remarque : Ces deux inégalités ont un très grand intérêt théorique. Cependant, l'inégalité de Bienaymé-Chebychev ne fournit pas en général une bonne majoration de $P(|X - E[X]| \geq a)$. On verra en TD une autre inégalité découlant de l'inégalité de Markov, l'inégalité de Chernov, qui en donne une meilleure approximation.

3.6.4 Médiane - Quantiles d'une v.a.r.

L'espérance est-elle toujours le meilleur indicateur de centralité d'une v.a.r. ?

Définition 24 (Inverse généralisée). *Si X a pour fonction de répartition F_X , on définit l'inverse généralisée de F_X :*

$$\forall u \in [0, 1], F_X^{-1}(u) = \inf\{x, F_X(x) \geq u\}.$$

Proposition 22. F_X^{-1} vérifie les propriétés suivantes.

- Pour tout $x \in \mathbb{R}$, $F_X^{-1}(F_X(x)) \leq x$;
- Pour tout $u \in [0, 1]$, $F_X(F_X^{-1}(u)) \geq u$;
- Pour tout $u \in [0, 1]$, $F_X^{-1}(u) \leq x \Leftrightarrow u \leq F_X(x)$;
- Si F_X est continue, strictement croissante, alors F_X^{-1} est la fonction réciproque de F , et pour tout $x \in \mathbb{R}$, $F_X^{-1}(F_X(x)) = x$, pour tout $u \in [0, 1]$, $F_X(F_X^{-1}(u)) = u$.

Définition 25 (Quantiles). *Pour $\alpha \in]0, 1[$, le quantile d'ordre α ou α -quantile de X est défini par :*

$$q_\alpha = F_X^{-1}(\alpha).$$

Le quantile d'ordre $1/2$ est appelé la **médiane** de X .

Proposition 23. *Le quantile q_α d'ordre α de X vérifie :*

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

En particulier, la médiane $M = q_{0,5}$ de X vérifie :

$$P(X \leq M) \geq 1/2 \text{ et } P(X \geq M) \geq 1/2.$$

Interprétation graphique. Exemples.

3.7 Calcul de lois

Dans ce paragraphe, on considère toujours l'espace probabilisé (Ω, \mathcal{F}, P) , et une v.a.r. X définie sur (Ω, \mathcal{F}) , dont on connaît la loi de probabilité P_X . On cherche ici à répertorier les méthodes de détermination de la loi d'une variable aléatoire définie comme $Y = \varphi(X)$, φ étant une fonction réelle mesurable.

Première méthode. Utilisation de la fonction de répartition.

On a vu que la fonction de répartition caractérise entièrement la loi d'une variable aléatoire, donc on peut chercher à exprimer par exemple $F_Y = F_{\varphi(X)}$ en fonction de F_X . Cette méthode est valable pour tout type de v.a.r.

Cependant, lorsqu'on connaît le type de la variable $\varphi(X)$, on préfère caractériser la loi par des fonctions plus explicites (fonction de masse pour une loi discrète, fonction de densité pour une loi absolument continue).

Pour une v.a.r. de loi discrète, la fonction de masse s'obtient à partir de la fonction de répartition très facilement ($P(Y = y) = F_Y(y) - F_Y(y^-)$).

Pour une v.a.r. de loi absolument continue, dont la densité est dérivable sauf éventuellement en un nombre fini de points, la densité s'obtient également à partir de la fonction de répartition par dérivation $f_Y = (F_Y)'$.

Il est en fait souvent plus facile de déterminer directement les fonctions de masse et de densité sans passer par la détermination de la fonction de répartition.

Deuxième méthode. Détermination directe de la fonction de masse dans le cas d'une loi discrète.

On cherche ici à exprimer directement $P(Y = y) = P(\varphi(X) = y)$ pour tout $y \in \varphi(D_X)$, et on utilise pour cela les propriétés de base des probabilités.

Troisième méthode. Détermination directe de la fonction de densité dans le cas d'une loi absolument continue.

Cette méthode est basée sur le théorème suivant.

Théorème 4. Soit X une v.a.r. dont la loi est absolument continue, de densité f . Alors pour toute fonction réelle mesurable (ou continue) bornée ψ , on a

$$E[\psi(X)] = \int_{-\infty}^{+\infty} \psi(x)f(x)dx. \quad (3.2)$$

Réciproquement, si pour toute fonction ψ réelle mesurable (ou continue) bornée, l'équation (3.2) est vérifiée, alors f est la densité de la loi de X .

Il s'agit ici de trouver une fonction g telle que pour toute fonction mesurable (ou continue) bornée ψ ,

$$E[\psi(Y)] = E[\psi(\varphi(X))] = \int_{-\infty}^{+\infty} \psi(y)g(y)dy.$$

Par identification, on pourra alors affirmer que g est la densité de $Y = \varphi(X)$.

Comment trouver cette fonction ?

Un cas simple est celui où X admet pour densité $f_X \mathbf{1}_U$, où U est un ouvert de \mathbb{R} , et où φ est un difféomorphisme de U sur V .

On peut en effet écrire, puisque $\psi \circ \varphi$ est une fonction réelle mesurable,

$$\begin{aligned} E[\psi(\varphi(X))] &= \int_U \psi(\varphi(x))f_X(x)dx \\ &= \int_V \psi(y)f_X(\varphi^{-1}(y))|(\varphi^{-1})'(y)|dy, \end{aligned}$$

la dernière ligne étant obtenue par le théorème de changement de variables. Ainsi, la densité de $\varphi(X)$ est donnée par $y \mapsto f_X(\varphi^{-1}(y))|(\varphi^{-1})'(y)|\mathbb{1}_V(y)$.

Dans les autres cas, on peut toujours découper le support de X en morceaux sur lesquels φ est un difféomorphisme, ou revenir à la première méthode...

Exemples.

3.8 Corrélation et indépendance

3.8.1 Définitions et premières propriétés

On considère toujours le même espace de probabilité (Ω, \mathcal{F}, P) .

Définition 26. Deux v.a.r. X et Y définies sur (Ω, \mathcal{F}) sont dites **indépendantes** si pour tous boréliens A et B de \mathbb{R} , les évènements $(X \in A)$ et $(Y \in B)$ sont indépendants c'est-à-dire :

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Proposition 24.

- Deux v.a.r. X et Y discrètes définies sur (Ω, \mathcal{F}) , à valeurs dans $\{x_k, k \in \mathcal{K}\}$ et $\{y_j, j \in \mathcal{J}\}$ respectivement sont indépendantes si pour tous $k \in \mathcal{K}$, $j \in \mathcal{J}$,

$$P(X = x_k, Y = y_j) = P(X = x_k)P(Y = y_j).$$

- Deux v.a.r. X et Y définies sur (Ω, \mathcal{F}) sont indépendantes si pour tous $x, y \in \mathbb{R}$,

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y).$$

Définition 27. Si X_1, \dots, X_m sont des v.a.r. définies sur (Ω, \mathcal{F}) , elles sont dites **indépendantes** si pour tous sous-ensembles A_1, \dots, A_m de \mathbb{R} , les évènements $(X_1 \in A_1), \dots, (X_m \in A_m)$ sont mutuellement indépendants.

Une suite $(X_n)_{n \in \mathbb{N}}$ de v.a.r. définies sur (Ω, \mathcal{F}) est dite **indépendante** si toute sous-suite finie est indépendante au sens de la définition ci-dessus.

Proposition 25. Soit X et Y deux v.a.r. indépendantes, f et g deux fonctions continues (au moins mesurables) dont les domaines de définition contiennent respectivement $X(\Omega)$ et $Y(\Omega)$. Alors les v.a.r. $f(X)$ et $g(Y)$ sont indépendantes.

Remarque : on verra d'autres propriétés des v.a.r. indépendantes dans le chapitre sur les couples aléatoires, en particulier pour les v.a.r. absolument continues.

3.8.2 Somme de v.a.r., covariance et corrélation

Proposition 26. Si X et Y sont deux v.a.r. indépendantes avec $X(\Omega) \subset \mathbb{N}$ et $Y(\Omega) \subset \mathbb{N}$, la loi de la variable aléatoire $X + Y$ est donnée par :

$$\forall n \in \mathbb{N}, P(X + Y = n) = \sum_{i=0}^n P(X = i)P(Y = n - i).$$

Exemple : somme de v.a.r. indépendantes de lois de Poisson.

On sait que si X et Y sont deux v.a.r. définies sur (Ω, \mathcal{F}) , d'espérance finie, $E[X + Y] = E[X] + E[Y]$ (linéarité de l'espérance). Que peut-on dire de l'espérance du produit de X et Y ?

Théorème 5 (Inégalité de Cauchy Schwarz). *Soit X et Y deux v.a.r. définies sur (Ω, \mathcal{F}) . Si X et Y ont des moments d'ordre 2 finis :*

$$|E[XY]| \leq (E[X^2])^{1/2}(E[Y^2])^{1/2}.$$

Preuve. L'existence de $|E[XY]|$ est assurée par celle de $E[X^2]$ et de $E[Y^2]$. De même, pour tout $t \in \mathbb{R}$, le polynôme

$$t^2 E[Y^2] + 2tE[XY] + E[X^2] = E[(X + tY)^2],$$

est défini et positif. Ceci n'est possible que si son discriminant est négatif. Ceci se traduit par : $\Delta = 4(E[XY])^2 - 4E[X^2]E[Y^2] \leq 0$, ce qui fournit exactement l'inégalité cherchée.

Remarque : il y a égalité dans l'inégalité de Cauchy Schwarz si et seulement si $\Delta = 0$ i.e. le trinôme considéré admet une unique racine réelle t_0 ou encore $E[(X + t_0Y)^2] = 0$, ce qui équivaut à $X = -t_0Y$ (presque sûrement i.e. avec probabilité égale à 1) : les deux variables X et Y sont (presque sûrement) colinéaires ou liées par une relation affine.

Proposition 27. *Si X et Y sont deux v.a.r. indépendantes définies sur (Ω, \mathcal{F}) , d'espérance finie, alors XY a une espérance finie et*

$$E[XY] = E[X]E[Y].$$

Attention : la réciproque est fautive !

On peut généraliser l'inégalité de Cauchy Schwarz avec l'inégalité de Hölder (mais la preuve est plus difficile).

Théorème 6 (Inégalité de Hölder). *Soit X et Y deux v.a.r. définies sur (Ω, \mathcal{F}) , et p, q deux réels tels que $p, q > 1$ et $\frac{1}{p} + \frac{1}{q} = 1$. Si X et Y ont des moments d'ordre p et q finis respectivement, alors*

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

Regardons maintenant la variance d'une somme de v.a.r.

Proposition 28. *Si X et Y sont deux v.a.r. définies sur (Ω, \mathcal{F}) ayant des moments d'ordre 2 finis, alors $X + Y$ a une variance finie et*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2E[(X - E[X])(Y - E[Y])].$$

Définition 28. *Si X et Y sont deux v.a.r. définies sur (Ω, \mathcal{F}) ayant des moments d'ordre 2 finis, on appelle **covariance** de X et Y la quantité :*

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

Remarques :

- $\text{Cov}(X, X) = \text{Var}(X)$.
- La proposition précédente se traduit de la façon suivante : si X et Y sont deux v.a.r. définies sur (Ω, \mathcal{F}) ayant des moments d'ordre 2 finis, alors $X + Y$ a une variance finie et

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

Proposition 29. *Soit X et Y deux v.a.r. définies sur (Ω, \mathcal{F}) ayant des moments d'ordre 2 finis.*

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
2. Pour tous réels a, b, c, d , $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$.
3. $|\text{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$.

4. Si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$, mais la réciproque est fautive. On peut regarder ce contre-exemple : soit X la variable aléatoire discrète de loi $P_X = \frac{1}{3}(\delta_{-1} + \delta_0 + \delta_1)$, et $Y = X^2$. On a $E[X] = 0$ et $E[XY] = 0$, d'où $\text{Cov}(X, Y) = 0$ alors que $P(Y = 1, X = 0) = 0 \neq P(Y = 1)P(X = 0)$ (Y et X^2 ne sont clairement pas indépendantes).

Définition 29. Si X et Y sont deux v.a.r. définies sur (Ω, \mathcal{F}) ayant des moments d'ordre 2 finis, et telles que $\sigma(X)\sigma(Y) > 0$. On appelle **coefficient de corrélation (linéaire)** de X et Y la quantité :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

On dit que X et Y sont **corrélées** lorsque $\rho(X, Y) \neq 0 \Leftrightarrow \text{Cov}(X, Y) \neq 0$, **non corrélées** lorsque $\rho(X, Y) = 0 \Leftrightarrow \text{Cov}(X, Y) = 0$.

Proposition 30. On a $-1 \leq \rho(X, Y) \leq 1$, avec $|\rho| = 1$ si et seulement si X et Y sont liées par une relation affine (i.e. $Y = aX + b$).

Preuve. Voir la remarque suivant l'inégalité de Cauchy Schwarz.

Proposition 31 (Formule de Koenig). Soit X et Y sont deux v.a.r. définies sur (Ω, \mathcal{F}) ayant une covariance. Alors

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

Proposition 32. Soit X_1, \dots, X_n des v.a.r. définies sur (Ω, \mathcal{F}) ayant un moment d'ordre 2 fini. Alors

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i,j=1}^n \text{Cov}(X_i, Y_j) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i,j=1, i \neq j}^n \text{Cov}(X_i, Y_j).$$

Proposition 33. Soit X_1, \dots, X_n des v.a.r. définies sur (Ω, \mathcal{F}) ayant un moment d'ordre 2 fini et indépendantes. Alors

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i).$$

Conséquence importante en statistique inférentielle : soit X_1, \dots, X_n des v.a.r. définies sur (Ω, \mathcal{F}) ayant un moment d'ordre 2 fini, indépendantes et identiquement distribuées (i.i.d.). Alors

$$E \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = E[X_1] \text{ et } \text{Var} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n} \text{Var}(X_1).$$

3.9 Transformées

3.9.1 Fonction génératrice des v.a.r. discrètes à valeurs dans \mathbb{N}

On considère ici une v.a.r. discrète X à valeurs dans \mathbb{N} ou dans un sous-ensemble de \mathbb{N} .

Définition 30. La fonction génératrice de X est définie par

$$G_X(t) = E[t^X], \text{ pour tout } t \text{ pour lequel cette espérance existe et est finie.}$$

Proposition 34. La fonction G_X vérifie les propriétés suivantes.

1. G_X caractérise la loi de X .
2. Si G_X est dérivable en 1, alors X admet une espérance finie et $E[X] = G'_X(1)$.

3. Si G_X est deux fois dérivable en 1, alors X admet un moment d'ordre deux fini et $E[X(X-1)] = G_X''(1)$, d'où $\text{Var}(X) = G_X''(1) + G_X'(1) - (G_X'(1))^2$.
4. Si G_X est k fois dérivable en 1, alors X admet un moment d'ordre k fini et $E[X(X-1)\dots(X-k+1)] = G_X^{(k)}(1)$.
5. Si Y est une v.a.r. discrète à valeurs dans \mathbb{N} ou dans un sous-ensemble de \mathbb{N} , indépendante de X , alors $G_{X+Y}(t) = G_X(t)G_Y(t)$. Attention : la réciproque est fautive.

Exemple : fonction génératrice d'une somme de deux v.a.r. de loi de Bernoulli indépendantes.

3.9.2 Fonction génératrice des moments et transformée de Laplace

On considère ici une v.a.r. X discrète ou absolument continue.

Définition 31. La fonction génératrice des moments de X est définie par

$$M_X(t) = E[e^{tX}], \text{ pour tout } t \text{ pour lequel cette espérance existe et est finie.}$$

La transformée de Laplace de X est définie par

$$L_X(t) = M_X(-t) = E[e^{-tX}], \text{ pour tout } t \text{ pour lequel cette espérance existe et est finie.}$$

Proposition 35. La fonction génératrice des moments M_X vérifie les propriétés suivantes.

1. M_X caractérise la loi de X .
2. Si M_X existe dans un intervalle ouvert autour de 0, alors M_X est n fois dérivable en 0 et $M_X^{(n)}(0) = E[X^n]$.

Exemple : espérance et variance de la loi exponentielle.

3.9.3 Fonction caractéristique

Définition 32. La fonction caractéristique de X est la fonction à valeurs complexes définie par

$$\forall t \in \mathbb{R}, \phi_X(t) = E[e^{itX}] = E[\cos(tX)] + iE[\sin(tX)].$$

Remarque : si X est une v.a.r. absolument continue de densité f_X ,

$$\forall t \in \mathbb{R}, \phi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx.$$

La fonction caractéristique de X est la transformée de Fourier inverse (à un facteur 2π) près dans l'exponentielle de sa densité.

La densité f_X s'obtient alors à l'aide de la formule d'inversion :

$$\forall t \in \mathbb{R}, f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_X(t) dt.$$

Proposition 36. La fonction caractéristique ϕ_X vérifie les propriétés suivantes.

1. ϕ_X caractérise la loi de X .
2. $\phi_X(0) = 1$.
3. $|\phi_X(t)| \leq 1$ pour tout réel t .
4. $\phi_X(-t) = \overline{\phi_X(t)}$. En particulier, si la loi de X est symétrique, alors ϕ_X est une fonction réelle paire.

5. Pour tout $(a, b) \in \mathbb{R}^2$, $\phi_{aX+b}(t) = e^{ibt} \phi_X(at)$.
6. Si la variable X admet un moment d'ordre r fini, alors ϕ_X est dérivable r fois au point $t = 0$, et : $E[X^r] = (-i)^r \phi_X^{(r)}(0)$. Attention : la réciproque est fautive.
7. Si Y est une v.a.r. indépendante de X , alors $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$. Attention : la réciproque est fautive.

Exemple : somme de deux v.a.r. de lois normales. La fonction caractéristique d'une v.a.r. X de loi normale $\mathcal{N}(m, \sigma^2)$ est égale à

$$\phi_X(t) = e^{itm} e^{-\frac{\sigma^2}{2}t^2}.$$

Ainsi, on peut montrer que la fonction caractéristique de la somme de deux v.a.r. indépendantes de lois normales $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$ respectivement est égale à

$$\phi_X(t) = e^{it(m_1+m_2)} e^{-\frac{\sigma_1^2+\sigma_2^2}{2}t^2}.$$

On reconnaît ainsi la fonction caractéristique d'une v.a.r. X de loi normale $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Chapitre 4

Vecteurs aléatoires

On généralise dans ce chapitre les notions introduites dans le chapitre précédent.

On considère une expérience aléatoire modélisée par un espace probabilisé (Ω, \mathcal{F}, P) .

Remarque importante : Par commodité d'écriture, un vecteur $(x_1, \dots, x_n)'$ sera noté ici (x_1, \dots, x_n) (on ne précisera pas la transposée) sauf dans quelques cas où cela sera précisé.

4.1 Définitions - premiers exemples

Pour tout $i \in \{1, \dots, n\}$ ($n \in \mathbb{N}^*$), on note π_i la projection canonique $\mathbb{R}^n \rightarrow \mathbb{R}$, $x = (x_1, \dots, x_n) \mapsto x_i$.

Définition 33 (Vecteur aléatoire, variables aléatoires marginales). *On appelle **variable aléatoire à valeurs dans \mathbb{R}^n** ou **vecteur aléatoire de dimension n** défini sur (Ω, \mathcal{F}) toute application $X : \Omega \rightarrow \mathbb{R}^n$ telle que pour tout borélien B de \mathbb{R}^n ($B \in \mathcal{B}(\mathbb{R}^n)$), l'ensemble $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ appartient à \mathcal{F} .*

*Si pour tout $i = 1, \dots, n$, $X_i = \pi_i(X)$ est la i ème coordonnée de X dans \mathbb{R}^n (de telle sorte que $X = (X_1, \dots, X_n)$), X_1, \dots, X_n sont des variables aléatoires réelles, appelées **variables aléatoires marginales**.*

*Lorsque $n = 2$, $X = (X_1, X_2)$ est aussi appelé **couple aléatoire**.*

Notation. Si $X : \Omega \rightarrow \mathbb{R}^n$ est un vecteur aléatoire défini sur (Ω, \mathcal{F}) , si B est un borélien de \mathbb{R}^n , l'ensemble $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ est un évènement généralement noté $\{X \in B\}$ et dont la probabilité notée $P(X \in B)$ se lit "probabilité que X soit dans B ". Suivant les différentes formes de B , on adopte une terminologie similaire à celle des variables aléatoires réelles.

4.2 Loi conjointe, lois marginales

La notion de vecteur aléatoire permet de probabiliser l'espace mesurable $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. On a en effet la proposition suivante.

Proposition 37. *Soit $X : \Omega \rightarrow \mathbb{R}^n$ un vecteur aléatoire défini sur (Ω, \mathcal{F}) . Alors l'application $P_X : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$ définie pour tout borélien B de \mathbb{R}^n par*

$$P_X(B) = P(X^{-1}(B)),$$

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

Définition 34 (Loi conjointe d'un vecteur aléatoire). *La mesure de probabilité P_X ainsi définie est appelée **loi conjointe** ou **loi jointe** du vecteur aléatoire X .*

On dit encore que X suit la loi conjointe P_X ou a pour loi conjointe P_X , et on peut noter $X \sim P_X$.

Remarques. L'espace $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_X)$ est un nouvel espace de probabilité. De plus, pour toute mesure de probabilité Q sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, il existe toujours un espace de probabilité (Ω, \mathcal{F}, P) et un vecteur aléatoire $X : \Omega \rightarrow \mathbb{R}^n$ défini sur (Ω, \mathcal{F}) tels que $P_X = Q$.

Comme pour les variables aléatoires réelles, on peut donc aussi parler de vecteur aléatoire X ayant une loi conjointe P_X sans spécifier l'espace de probabilité (Ω, \mathcal{F}, P) sur lequel X est défini à la base.

La loi conjointe d'un vecteur aléatoire $X = (X_1, \dots, X_n)$ détermine les lois des variables marginales X_1, \dots, X_n .

Proposition 38. *La loi conjointe P_X du vecteur aléatoire $X = (X_1, \dots, X_n)$ de dimension n défini sur (Ω, \mathcal{F}) permet de déterminer les lois (dites **lois marginales**) P_{X_1}, \dots, P_{X_n} des variables aléatoires marginales X_1, \dots, X_n . On a en effet pour tout borélien B de \mathbb{R} , $P_{X_i}(B) = P_X(\pi_i^{-1}(B))$ ($i \in \{1, \dots, n\}$).*

Remarque importante : la connaissance des lois marginales ne détermine pas la loi conjointe P_X . Tout va dépendre des "relations" pouvant exister entre les variables marginales X_1, \dots, X_n .

4.2.1 Fonctions de répartition - densités

Définition 35 (Fonction de répartition conjointe). *Si $X : \Omega \rightarrow \mathbb{R}^n$ est un vecteur aléatoire défini sur (Ω, \mathcal{F}) , on appelle **fonction de répartition conjointe/jointe** de la loi de probabilité de X ou (par abus) de X la fonction $F_X : \mathbb{R}^n \rightarrow [0, 1]$ définie par $F_X(x_1, \dots, x_n) = P_X(\prod_{i=1}^n]-\infty, x_i]) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$.*

Les fonctions de répartition conjointes ne sont pas très souvent utilisées. On cite donc ici sans les démontrer les trois propositions suivantes.

Proposition 39. *La fonction de répartition conjointe F_X d'un vecteur aléatoire X satisfait les propriétés suivantes.*

1. Pour tout $x \in \mathbb{R}^n$, $0 \leq F_X(x) \leq 1$.
2. F_X est une fonction croissante, continue à droite en chacun de ses arguments x_1, \dots, x_n .
3. $F_X(x)$ tend vers 0 lorsque l'un des arguments x_i tend vers $-\infty$, et tend vers 1 lorsque tous les x_i tendent vers $+\infty$.

Proposition 40. *La fonction de répartition conjointe caractérise la loi conjointe.*

Proposition 41. *La fonction de répartition conjointe d'un vecteur aléatoire $X = (X_1, \dots, X_n)$ détermine les fonctions de répartition F_{X_1}, \dots, F_{X_n} des variables aléatoires marginales X_1, \dots, X_n (appelées **fonctions de répartition marginales**). En effet, pour tout $i \in \{1, \dots, n\}$,*

$$F_{X_i}(x_i) = P(X_i \leq x_i) = \lim_{x_1, \dots, x_{i-1}, x_{i+1}, x_n \rightarrow +\infty} F_X(x_1, \dots, x_n) = F_X(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty).$$

Par commodité d'écriture, on ne considère dans toute la suite de ce paragraphe que des couples aléatoires ($n = 2$). Toutes les notions vues ici s'étendent néanmoins aux vecteurs aléatoires de dimension n (n quelconque).

Définition 36 (Loi discrète). *La loi conjointe P_X d'un couple aléatoire $X = (X_1, X_2)$ est dite **discrète** s'il existe un sous-ensemble S_X de \mathbb{R}^2 fini ou dénombrable tel que pour tout $(x_1, x_2) \in S_X$, $P_X(\{(x_1, x_2)\}) > 0$ et $P_X(S_X) = \sum_{(x_1, x_2) \in S_X} P_X(\{(x_1, x_2)\}) = 1$. L'ensemble S_X est alors appelé **support** de la loi conjointe de X .*

Remarque : un couple aléatoire a une loi conjointe discrète si et seulement si ses lois marginales sont discrètes.

Proposition 42. Si $X = (X_1, X_2)$ est un couple aléatoire de loi conjointe P_X discrète, si S_{X_1} désigne le support de X_1 , S_{X_2} celui de X_2 , alors $S_X \subset S_{X_1} \times S_{X_2}$ et les lois marginales de X_1 et X_2 sont données par :

$$P_{X_1}(\{(x_1)\}) = \sum_{x_2 \in S_{X_2}} P_X(\{(x_1, x_2)\}) \quad \forall x_1 \in S_{X_1}, \quad P_{X_2}(\{(x_2)\}) = \sum_{x_1 \in S_{X_1}} P_X(\{(x_1, x_2)\}) \quad \forall x_2 \in S_{X_2}.$$

Exemple. On lance successivement trois pièces de monnaie (équilibrées) et on désigne respectivement par X et Y les variables aléatoires représentant le nombre de *faces* apparues sur les deux premières pièces et le nombre de *piles* sur les deux dernières.

La loi conjointe du couple (X, Y) et les lois marginales de X et Y sont données dans le tableau suivant :

$Y \backslash X$	0	1	2	P_X
0	0	1/8	1/8	1/4
1	1/8	2/8	1/8	1/2
2	1/8	1/8	0	1/4
P_Y	1/4	1/2	1/4	1

Définition 37 (Loi absolument continue). La loi conjointe P_X d'un couple aléatoire $X = (X_1, X_2)$ est dite **absolument continue** s'il existe une fonction f_X positive et intégrable telle que pour tout borélien B de \mathbb{R}^2 ,

$$P_X(B) = \int \int_B f_X(x_1, x_2) dx_1 dx_2. \quad (4.1)$$

Définition 38 (Densité conjointe). On appelle **densité conjointe** de la loi P_X ou (par abus) de X toute fonction f_X qui vérifie (4.1).

Proposition 43. Une densité conjointe f_X vérifie les propriétés suivantes.

1. $\int \int_{\mathbb{R}^2} f_X(x_1, x_2) dx_1 dx_2 = 1$.
2. Si F_X désigne la fonction de répartition conjointe de X , $F_X(x_1, x_2) = \int \int_{]-\infty, x_1] \times]-\infty, x_2]} f_X(x_1, x_2) dx_1 dx_2$.
3. Si f_X est continue en (x_1, x_2) , alors $f_X(x_1, x_2) = \frac{\partial^2 F_X}{\partial x_1 \partial x_2}(x_1, x_2)$.
4. Si la fonction de répartition F_X est de classe C^2 sur \mathbb{R}^2 , alors la loi de X est absolument continue et $f_X(x_1, x_2) = \frac{\partial^2 F_X}{\partial x_1 \partial x_2}(x_1, x_2)$. Attention : la réciproque est fautive !
5. La fonction f_X caractérise la loi conjointe P_X .

Proposition 44. Toute fonction f_X de \mathbb{R}^2 dans \mathbb{R} , positive, intégrable, et telle que

$$\int \int_{\mathbb{R}^2} f_X(x_1, x_2) d\lambda(x_1) d\lambda(x_2) = 1,$$

est la densité conjointe d'une loi absolument continue sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$.

Proposition 45. Si le couple $X = (X_1, X_2)$ admet une loi conjointe absolument continue, alors ses lois marginales sont absolument continues et sa densité conjointe détermine les densités des lois marginales (appelées **densités marginales**) f_{X_1} et f_{X_2} de la façon suivante :

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{(X_1, X_2)}(x_1, x_2) dx_2, \quad f_{X_2}(x_2) = \int_{\mathbb{R}} f_{(X_1, X_2)}(x_1, x_2) dx_1.$$

Preuve : Théorème de Fubini-Tonnelli.

Exemple. Soit (X, Y) un couple aléatoire dont la densité conjointe est définie par :

$$\begin{cases} xy/2 & \text{si } 0 \leq y \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$$

1. La fonction de répartition du couple (X, Y) est donnée par les équations suivantes. Si $x \leq 0$ ou si $y \leq 0$, $F(x, y) = 0$. Si $0 \leq y \leq x \leq 2$, $F(x, y) = \frac{1}{2} \int_0^y v (\int_v^x u du) dv = \frac{1}{4} \int_0^y v(x^2 - v^2) dv = \frac{y^2}{16}(2x^2 - y^2)$. Si $0 \leq y \leq 2 \leq x$, $F(x, y) = F(2, y) = \frac{y^2}{16}(8 - y^2)$. Si $0 < x \leq y$ et $x \leq 2$, $F(x, y) = F(x, x) = \frac{x^4}{16}$. Enfin, si $x \geq 2$, $y \geq 2$, $F(x, y) = 1$.

2. Les densités marginales s'obtiennent par exemple par intégration de la densité conjointe : $f_X(x) = \frac{x^3}{4} \mathbb{1}_{[0,2]}(x)$, et $f_Y(y) = \frac{y}{4}(4 - y^2) \mathbb{1}_{[0,2]}(y)$ ou à partir de la fonction de répartition conjointe (plus difficile).

3. On peut calculer par exemple $P(Y \geq X^2)$.

$$P(Y \geq X^2) = \int \int \frac{xy}{2} \mathbb{1}_{0 \leq y \leq x \leq 2} \mathbb{1}_{y \geq x^2} dx dy = \int \int \frac{xy}{2} \mathbb{1}_{0 \leq x \leq 1} \mathbb{1}_{x^2 \leq y \leq x} dx dy \text{ (faire un graphe), d'où}$$

$$P(Y \geq X^2) = \frac{1}{2} \int_0^1 x (\int_{x^2}^x y dy) dx = 1/48.$$

4.2.2 Espérance mathématique - Moments d'un vecteur aléatoire

On considère un vecteur aléatoire $X = (X_1, \dots, X_n)$ défini sur (Ω, \mathcal{F}) de loi conjointe notée P_X .

Définition 39. On dit que $X = (X_1, \dots, X_n)$ a une espérance si ses variables marginales X_1, \dots, X_n possèdent une espérance, et on note

$$E[X] = (E[X_1], \dots, E[X_n]).$$

Les propriétés de l'espérance pour les vecteurs aléatoires découlent donc directement de celles de l'espérance des variables aléatoires réelles. En particulier, on a le résultat de linéarité suivant.

Proposition 46. Si A est une matrice réelle de taille $p \times n$, alors

$$E[AX] = AE[X].$$

On utilise les termes de centrage et recentrage comme pour les variables aléatoires réelles.

Définition 40. Un vecteur aléatoire $X = (X_1, \dots, X_n)$ dit **centré** si $E[X] = 0$.

On dit que l'on centre un vecteur aléatoire $X = (X_1, \dots, X_n)$ lorsqu'on lui retire son espérance.

Théorème 7 (Théorème de transfert). Soit φ une fonction de $\mathbb{R}^n \rightarrow \mathbb{R}$.

1. Si X est un vecteur aléatoire discret à valeurs dans $\{x_k, k \in \mathcal{K}\}$ tel que $\sum_{k \in \mathcal{K}} |\varphi(x_k)| P_X(\{(x_k)\}) < +\infty$, alors $\varphi(X)$ est une v.a.r. discrète, dont l'espérance est donnée par :

$$E[\varphi(X)] = \sum_{k \in \mathcal{K}} \varphi(x_k) P_X(\{(x_k)\}).$$

2. Si φ est une fonction mesurable i.e. telle que pour tout borélien B de \mathbb{R}^n , $\varphi^{-1}(B)$ est un borélien de \mathbb{R}^n , et si X est un vecteur aléatoire de loi absolument continue (de densité f_X) tels que $\int_{\mathbb{R}^n} |\varphi(x)| f_X(x) dx < +\infty$, alors $\varphi(X)$ est une v.a.r. dont l'espérance est donnée par :

$$E[\varphi(X)] = \int_{\mathbb{R}^n} \varphi(x) f_X(x_1, \dots, x_n) dx.$$

Définition 41. Un vecteur aléatoire $X = (X_1, \dots, X_n)$ dont toutes les variables marginales possèdent un moment d'ordre 2 fini (i.e. $E[X_i^2] < +\infty \forall i = 1, \dots, n$) est dit **du second ordre**.

On appelle **matrice des moments du second ordre** de X la matrice carrée M_X de taille $n \times n$ de terme général $m_{i,j} = E[X_i X_j]$.

On appelle **matrice des variances-covariances** de X la matrice carrée Σ_X de taille $n \times n$ de terme général $\sigma_{i,j}^2 = Cov(X_i, X_j) = E[(X_i - E[X_i])(X_j - E[X_j])] = E[X_i X_j] - E[X_i]E[X_j]$.

Remarques :

1. L'existence des $m_{i,j}$ et $\sigma_{i,j}^2$ est assurée par l'inégalité de Cauchy-Schwarz ;
2. Les matrices M_X et Σ_X sont des matrices symétriques ;
3. On rappelle que pour $i = 1, \dots, n$, $Cov(X_i, X_i) = Var(X_i)$ donc les éléments diagonaux de la matrice de variances-covariances sont les variances des variables marginales.

Définition 42. Un vecteur aléatoire du second ordre est dit **réduit** si $\Sigma_X = I$.

Pour les deux propositions suivantes, on précise à nouveau la transposée.

En admettant que l'on définit une matrice aléatoire de la même façon qu'un vecteur aléatoire, on a le résultat suivant.

Proposition 47. Si $X = (X_1, \dots, X_n)'$ est un vecteur aléatoire du second ordre,

$$M_X = E[XX'],$$

et

$$\Sigma_X = E[(X - E[X])(X - E[X])'].$$

Des propriétés de linéarité de l'espérance, on déduit la proposition suivante.

Proposition 48. Si $X = (X_1, \dots, X_n)'$ est un vecteur aléatoire du second ordre et A une matrice réelle de taille $p \times n$, on a

$$M_{AX} = AM_X A',$$

et

$$\Sigma_{AX} = A\Sigma_X A'.$$

Proposition 49. Une matrice symétrique Σ est la matrice de variances-covariances d'un vecteur aléatoire si et seulement si elle est positive.

Preuve. On considère un vecteur aléatoire $U = (U_1, \dots, U_n)$ de matrice de variances-covariances $\Sigma_U = I$ (on verra après le paragraphe sur l'indépendance qu'un tel vecteur existe). On sait que toute matrice symétrique positive Σ s'écrit sous la forme $\Sigma = P\Delta P'$, P étant une matrice orthogonale (matrice de changement de base), Δ une matrice diagonale positive (composée des valeurs propres de Σ). On peut donc introduire $\Sigma^{1/2} = P\Delta^{1/2}P'$, où $\Delta^{1/2}$ est la matrice diagonale, dont les éléments diagonaux sont les racines carrées positives des éléments diagonaux de Δ . Le vecteur $X = \Sigma^{1/2}U$ a pour matrice de variances-covariances Σ .

Proposition 50. Soit X un vecteur aléatoire dont la matrice de variances-covariances Σ_X est inversible.

Le vecteur $Y = \Sigma_X^{-1/2}X$ est réduit, et le vecteur $\Sigma_X^{-1/2}(X - E[X])$ est centré réduit.

4.3 Variables aléatoires marginales indépendantes

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire défini sur (Ω, \mathcal{F}) .

On rappelle que les variables marginales X_1, \dots, X_n sont indépendantes si et seulement si pour tous réels x_1, \dots, x_n ,

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \dots P(X_n \leq x_n),$$

ce qui se traduit par :

$$F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n).$$

Dans le cas des vecteurs aléatoires de loi absolument continue, on peut montrer à l'aide du théorème de Fubini-Tonelli le résultat suivant sur les densités.

Théorème 8. Si les variables marginales X_1, \dots, X_n sont indépendantes, de lois marginales absolument continues, si l'on note f_{X_1}, \dots, f_{X_n} les densités marginales respectives, alors la densité conjointe de X est de la forme :

$$f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n).$$

Réciproquement, si le vecteur aléatoire $X = (X_1, \dots, X_n)$ possède une densité conjointe de la forme

$$f_X(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n),$$

alors les variables marginales X_1, \dots, X_n sont indépendantes, et leurs densités respectives sont égales à une constante multiplicative près à f_1, \dots, f_n .

4.4 Calcul de lois

On considère dans ce paragraphe un espace de probabilité (Ω, \mathcal{F}, P) et un vecteur aléatoire $X = (X_1, \dots, X_n)$ de dimension n défini sur (Ω, \mathcal{F}) , de loi conjointe P_X , de fonction de répartition conjointe F_X .

Il s'agit ici d'apprendre à déterminer pour une fonction mesurable $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^p$, la loi du vecteur aléatoire (ou si $p = 1$, de la variable aléatoire réelle) $Y = \varphi(X)$.

Première méthode : utilisation de la fonction de répartition conjointe

On cherche ici à exprimer la fonction de répartition F_Y de Y en fonction de F_X , en utilisant par exemple le théorème de transfert appliqué à une indicatrice :

$$P(Y_1 \leq y_1, \dots, Y_p \leq y_p) = P(\varphi(X) \in]-\infty, y_1] \times \dots \times]-\infty, y_p]).$$

Si la loi de $Y = \varphi(X)$ est discrète, on peut en déduire la fonction de masse. Si la loi de $Y = \varphi(X)$ est absolument continue, on peut dériver F_Y par rapport à ses p variables et en déduire la densité de Y .

Deuxième méthode : détermination directe de la fonction de masse dans le cas d'une loi discrète

Troisième méthode : détermination directe de la densité dans le cas d'une loi absolument continue

Supposons que X est à valeurs dans un ouvert $U \subset \mathbb{R}^n$.

Supposons tout d'abord également que φ est une bijection de U dans un ouvert V de \mathbb{R}^n , ayant des dérivées partielles continues. On note J_ψ le jacobien de la bijection inverse $\psi = \varphi^{-1}$, i.e. si

$$\begin{aligned} \varphi : (x_1, \dots, x_n) &\mapsto (y_1, \dots, y_n) = (\varphi_1(x_1, \dots, x_n), \dots, \varphi_n(x_1, \dots, x_n)), \\ \psi = \varphi^{-1} : (y_1, \dots, y_n) &\mapsto (x_1, \dots, x_n) = (\psi_1(y_1, \dots, y_n), \dots, \psi_n(y_1, \dots, y_n)), \end{aligned}$$

$$J_\psi = \det \begin{bmatrix} \frac{\partial \psi_1(y_1, \dots, y_n)}{\partial y_1} & \dots & \frac{\partial \psi_1(y_1, \dots, y_n)}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial \psi_n(y_1, \dots, y_n)}{\partial y_1} & \dots & \frac{\partial \psi_n(y_1, \dots, y_n)}{\partial y_n} \end{bmatrix}.$$

Si $J_\psi \neq 0$ (sauf éventuellement sur un ensemble négligeable I tel que $\psi(I)$ est aussi négligeable), alors $Y = (Y_1, \dots, Y_n) = (\varphi_1(X_1, \dots, X_n), \dots, \varphi_n(X_1, \dots, X_n))$ est un vecteur aléatoire de loi absolument continue, dont la densité conjointe est donnée par :

$$f_Y(y_1, \dots, y_n) = f_X(\psi_1(y_1, \dots, y_n), \dots, \psi_n(y_1, \dots, y_n)) |J_\psi| \mathbb{1}_V(y_1, \dots, y_n).$$

Ce résultat est basé sur un théorème d'identification, couplé à un changement de variables sur des intégrales multiples, comme pour les variables aléatoires réelles.

Supposons maintenant que $\varphi : (x_1, \dots, x_n) \in U \mapsto (\varphi_1(x_1, \dots, x_n), \dots, \varphi_p(x_1, \dots, x_n))$ est à valeurs dans \mathbb{R}^p , avec $p < n$.

On introduit alors une fonction $\tilde{\varphi} : (x_1, \dots, x_n) \in U \mapsto (\tilde{\varphi}_1(x_1, \dots, x_n), \dots, \tilde{\varphi}_n(x_1, \dots, x_n))$ qui est une bijection de U dans un ouvert V de \mathbb{R}^n et telle que $\tilde{\varphi}_1 = \varphi_1, \dots, \tilde{\varphi}_p = \varphi_p$. On obtient par le résultat précédent, sous les conditions de régularité adéquates, l'expression de la densité conjointe de $\tilde{\varphi}(X) = (\varphi_1(X_1, \dots, X_n), \dots, \varphi_p(X_1, \dots, X_n), \tilde{\varphi}_{p+1}(X_1, \dots, X_n), \dots, \tilde{\varphi}_n(X_1, \dots, X_n))$, et en intégrant cette densité conjointe par rapport à ses $(n - p)$ dernières variables, on en déduit la densité conjointe de $(\varphi_1(X_1, \dots, X_n), \dots, \varphi_p(X_1, \dots, X_n))$.

Dans tous les autres cas, on peut revenir à la première méthode.

Exemple : cas d'un changement en polaires.

Autre exemple fondamental : la densité d'une somme de deux variables aléatoires réelles **indépendantes** X_1 et X_2 , de densités respectives f_1 et f_2 , est donnée par le **produit de convolution** $f_1 * f_2(u) = \int f_1(u - v)f_2(v)dv$.

Chapitre 5

Lois usuelles dans \mathbb{R} et dans \mathbb{R}^n

